

Пытьев Ю.П., Шишмарев И.А. **Теория вероятностей, математическая статистика и элементы теории возможностей для физиков.** — М.:, 2010. — 408с. — ISBN

В первой главе представлен достаточно полный курс теории вероятностей, включающий важную для физических приложений теорию случайных процессов.

Во второй главе представлены основные разделы математической статистики и ее приложений в теории измерительно-вычислительных преобразователей как средств измерений и в теории статистических решений.

В третьей главе рассмотрены элементы теории возможностей как альтернативной теории вероятностей модели случайности, позволяющей эмпирически восстанавливать математические модели объектов, в том числе стохастических, вероятностные модели которых не могут быть построены эмпирически. Рассмотрены приложения теории возможностей в задачах оптимизации решений, анализа и интерпретации измерительного эксперимента и др.

ОГЛАВЛЕНИЕ

Предисловие	7
Глава 1. Теория вероятностей	9
1.0. Введение	9
1. Эксперимент с конечным числом равновероятных исходов (9).	
2. Геометрические вероятности (10).	
1.1. Пространство элементарных событий. Алгебра событий	13
1.1.1. Элементарные события. События. Пространство элементарных событий (13).	
1.1.2. Операции над событиями (14).	
1.1.3. Свойства операций над событиями. Принцип двойственности (16).	
1.1.4. Алгебра событий (18).	
1.2. Классическая теоретико-вероятностная модель	19
1.2.1. Определение классической вероятности. Примеры (19).	
1.2.2. Свойства классической вероятности (24).	
1.3. Аксиоматическое построение теории вероятностей.	28
1.3.1. Аксиомы теории вероятностей (28).	
1.3.2. Дискретное вероятностное пространство (31).	
1.3.3. Свойства вероятности. Непрерывность (34).	
1.4. Условная вероятность. Независимость	37
1.4.1. Формулы полной вероятности и Байеса (42).	
1.5. Последовательность независимых испытаний	44
1.6. Распределение Пуассона	48
1.7. Локальная и интегральная предельные теоремы Муавра–Лапласа	53
1.8. Случайные величины, функции распределения	58
1.8.1. Случайные величины (58).	
1.8.2. Функции распределения и их свойства (60).	
1.8.3. Дискретные и абсолютно непрерывные случайные величины (64).	
1.8.4. Векторные (многомерные) случайные величины (68).	
1.8.5. Независимость случайных величин (73).	
1.8.6. Функции случайных величин (75).	
1.9. Числовые характеристики случайных величин	84
1.9.1. Моменты случайных величин (84).	
1.9.2. Центральные моменты (89).	
1.9.3. Свойства математического ожидания и диспер-	

сии (90). 1.9.4. Условное математическое ожидание. Наилучший в среднеквадратичном прогноз (94). 1.9.5. Моменты векторных случайных величин (99).	
1.10. Законы больших чисел	103
1.11. Характеристические функции. Центральные предельные теоремы	114
1.11.1. Характеристические функции (114). 1.11.2. Теорема о непрерывности (121). 1.11.3. Центральные предельные теоремы (126). 1.11.4. Применения центральных предельных теорем. Примеры (134).	
1.12. Конечные однородные цепи Маркова	137
1.12.1. Определение цепи Маркова (137). 1.12.2. Вероятность перехода за n шагов (142). 1.12.3. Эргодичность (143).	
1.13. Случайные процессы	148
1.13.1. Процессы Пуассона и Винера (149). 1.13.2. Процессы второго порядка (Корреляционная теория) (156). 1.13.3. Марковские процессы (160). 1.13.4. Стационарные процессы (174).	
Глава 2. Математическая статистика	188
2.1. Распределения ортогональных проекций нормального случайного вектора и функций от них	188
2.1.1. Введение (188). 2.1.2. Ортогональные преобразования, ортогональные проекции $N(0, \sigma^2 I)$ -случайного вектора, функции ортогональных проекций (191).	
2.2. Интервальное оценивание параметров нормального распределения	195
2.2.1. Интервальная оценка математического ожидания при известной дисперсии (195). 2.2.2. Интервальная оценка дисперсии при известном математическом ожидании (196). 2.2.3. Интервальная оценка математического ожидания при неизвестной дисперсии (197). 2.2.4. Интервальная оценка дисперсии при неизвестном математическом ожидании (198).	
2.3. Общая задача интервального оценивания	198
2.4. Точечные оценки	201
2.4.1. Несмещенные оценки минимальной дисперсии (205). 2.4.2. Эффективные оценки (206). 2.4.3. Достаточные статистики (210). 2.4.4. Оценки максимального правдоподобия (213).	
2.5. Линейная модель измерений	216
2.5.1. Несмещенные оценки минимальной дисперсии значений $\alpha_1, \dots, \alpha_k$ (217). 2.5.2. Метод наименьших квадратов оценивания значений $\alpha_1, \dots, \alpha_k$ (219). 2.5.3. Несмещенная оценка дисперсии σ^2 , $n > k$ (221). 2.5.4. Доверительные множества для значений $\alpha_1, \dots, \alpha_k$ в случае нормального распределения ошибок измерений (222).	
2.6. Методы анализа и интерпретации данных измерительного эксперимента	226
2.6.1. Измерительно-вычислительный преобразователь как средство измерения (226). 2.6.2. ИВП $[A, \Sigma, U]$ как максимально	

точная версия идеального ИП $[U, 0]$ (231). 2.6.3. О методе «наименьших квадратов» (234). 2.6.4. Качество ИВП $[A, \Sigma_\nu, U]$ и ИП $[A, \Sigma_\nu]$ как компоненты ИВП (235). 2.6.5. Влияние уточнения модели ИП на его качество (239). 2.6.6. Эффект «предварительной обработки» данных измерения (240). 2.6.7. Комбинирование данных независимых измерений (241). 2.6.8. Собственные базисы ИП $[A, \Sigma_\nu]$ и ИВП $[A, \Sigma_\nu, U]$ (243). 2.6.9. Эффективные ранги ИП $[A, \Sigma_\nu]$ и ИВП $[A, \Sigma_\nu, U]$ (247). 2.6.10. Методы инструментального синтеза линейного ИП на ИВП (249). 2.6.11. Гаусовский ИП $[A, \Sigma_\nu]$ (251). 2.6.12. Гауссовский ИП $[A, \Sigma_\nu^2 I]$, Σ_ν^2 неизвестно (252).	
2.7. Задачи проверки статистических гипотез	254
2.7.1. Локально наиболее мощные критерии (256). 2.7.2. Случай простой гипотезы и простой альтернативы. Наиболее мощный критерий (261). 2.7.3. Доверительные множества и задачи проверки гипотез (263).	
2.8. Элементы теории статистических решений	264
2.8.1. Средний (ожидаемый) риск. Рандомизация решения (267). 2.8.2. Минимаксное правило решения (269). 2.8.3. Байесовское правило решения (270). 2.8.4. Байесовское правило в случае невозможности наблюдений над природой (271). 2.8.5. Байесово действие (272). 2.8.6. Байесовская классификация (275). 2.8.7. Правило решения, минимизирующее ожидаемое число ошибок (275).	
2.9. Приложение	276
Глава 3. Элементы теории возможностей и её применения	281
3.0. Предисловие	281
1. Вероятность. Проблемы эмпирического построения и интерпретации (282). 2. Возможность как мера предопределенности исходов стохастического эксперимента (284). 3. Классы эквивалентных возможностей (285). 4. Шкала значений возможности. Возможность события (286). 5. Необходимость. Шкала значений необходимости (288). 6. Принцип относительности (289). 7. Максимальная согласованность возможности с вероятностью. Стохастическая модель возможности (290). 8. Эмпирическое построение и эмпирическая интерпретация возможности (291).	
3.1. Элементы теории возможностей	297
3.1.0. Предисловие (297). 3.1.1. Шкала значений возможности. Интеграл. Определение, свойства (297). 3.1.2. Мера возможности. Определение, свойства (301). 3.1.3. Нечеткие множества, элементы, события (304). 3.1.4. Мера необходимости. Определение, свойства (311). 3.1.5. Интегрирование по возможности и по необходимости (315). 3.1.6. Независимость. Условные возможность и необходимость (317).	
3.2. Стохастические модели возможности.	329
3.2.0. Введение (329). 3.2.1. Возможность, максимально согласованная с вероятностью (333). 3.2.2. Метод построения возможно-	

сти, максимально согласованной с вероятностью (334). 3.2.3. Возможность, согласованная с вероятностью на σ -алгебре. Гранулирование Ω (337). 3.2.4. Когерентные разбиения $\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}$ и

$$\mathbb{Pr} = \bigcup_{e \in (0,1)} \mathbb{Pr}_{(e)} \quad (339).$$

3.3. Эмпирическое построение возможности	341
3.3.0. Введение (341). 3.3.1. Эмпирическое построение $\mathbb{Pr}$ -измеримой возможности, $\mathbb{Pr} \in \mathbb{Pr}$ (342). 3.3.2. Эмпирическое построение $\mathbb{Pr}^1 - \dots, \mathbb{Pr}^k$ -измеримой возможности, $\mathbb{Pr}^s \in \mathbb{Pr}_{(e)}$, $s = 1, \dots, k$ (345). 3.3.3. [(349). 3.3.4. Алгоритм эмпирического восстановления возможности, максимально согласованной с вероятностью, изменяющейся в процессе испытаний (352).	
3.4. Экспертное восстановление возможности	354
3.4.1. Восстановление распределения нечеткого элемента путем попарных сравнений возможностей его значений (354). 3.4.2. Восстановление распределения нечеткого элемента путем упорядочения возможностей его значений (357).	
3.5. Нечеткие оптимальные решения	359
3.5.0. Введение (359). 3.5.1. Нечеткая модель. Идентификация (361). 3.5.2. Критерий минимума возможности (риска) потерь (363). 3.5.3. Правило решения, минимизирующее риск потерь (364). 3.5.4. Четкие правила идентификации (368). 3.5.5. Минимаксное правило. Фазификация (369). 3.5.6. Критерий минимума необходимости потерь (370). 3.5.7. Правило решения, минимизирующее необходимость потерь, дуальную возможности потерь (371). 3.5.8. Нечеткая модель. Оценивание (374). 3.5.9. Правило оценивания, минимизирующее возможность потерь (375). 3.5.10. Четкое правило оценивания (377). 3.5.11. Правило, минимизирующее необходимость потерь, дуальную возможности (378).	
3.6. Методы анализа и интерпретации данных измерений	381
3.6.1. Возможностные модели измерения и его интерпретации (381). 3.6.2. Редукция измерения при априори произвольном измеряемом сигнале (384). 3.6.3. Редукция измерения при нечеткой априорной информации об измеряемом сигнале (387).	
3.7. Приложение	392
3.7.0. Введение (392). 3.7.1. Алгоритм эмпирического упорядочения вероятностей элементарных событий, не изменяющихся в процессе испытаний (394). 3.7.2. Алгоритм эмпирического интервального оценивания вероятностей элементарных событий (399). 3.7.3. Алгоритм эмпирического упорядочения вероятностей элементарных событий, изменяющихся в процессе испытаний (399).	
3.8. Список обозначений Главы 3	402
Список литературы	404

ПРЕДИСЛОВИЕ

С момента издания учебного пособия "Курс теории вероятностей и математической статистики для физиков" Ю. П. Пытьева и И. А. Шишмарева прошло 27 лет. Всё это время книга активно используется как основное учебное пособие по семестровому курсу теории вероятностей и математической статистики для студентов 3-го курса.

Последние годы ощущалась острая необходимость переиздания книги, износившейся физически и не отражавшей важных для физиков последних достижений: в теории измерительно-вычислительных систем как средств измерений [18], в теории и методах эмпирического построения, исследования и проверки адекватности математических моделей [14, 19] и в других областях приложений теории вероятностей и математической статистики [40]. Наконец, в книге не отражены результаты фундаментальных исследований альтернативных вероятностной характеристик феномена случайности, повышенный интерес к которым обусловлен неэффективностью вероятностных методов моделирования сложных физических, технических, социальных и других объектов, субъективных суждений и т. п. [6, 12, 13, 17, 31, 32].

В ряде случаев неэффективность вероятностного моделирования объясняется его неадекватностью, когда нечеткость и неточность формулировок, свойственная моделям названных объектов, не может быть охарактеризована в терминах теории вероятностей. В других, более распространенных случаях, принципиальные трудности возникают при эмпирическом построении заведомо стохастического объекта ¹⁾, когда в процессе наблюдений за объектом его вероятностные характеристики непредсказуемо эволюционируют. Наконец, нередки случаи, в которых вероятностная модель оказывается неприемлемо сложной и громоздкой, и на практике используется неадекватное ее приближение.

В связи с отмеченными трудностями во второе издание добавлена глава "Элементы теории возможностей", в которой теория возможностей представлена как альтернативная теории вероятностей модель феномена случайности, позволяющая при определенных условиях эти трудности преодолеть и, более того, — решать все прикладные задачи, типичные для приложений теории вероятностей и математической статистики [13].

¹⁾ Стохастическим называется объект, математической моделью которого в каждый момент является некоторое вероятностное пространство.

Второе издание заново отредактировано в целом: устранены замеченные неточности, ошибки и опечатки, уточнены формулировки и доказательства некоторых теорем, переработаны § 9.4 гл. 1, § 8 гл. 2, заново написан § 6 гл. 2, и т.д.

Авторы глубоко благодарны О. С. Князевой и Ю. М. Нагорному, взявшим на себя труд подготовки компьютерного варианта рукописи и оригинал-макета, А. И. Чуличкову и М. Л. Сердобольской, прочитавшим рукопись и устранившим замеченные погрешности.

Глава 1

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1.0. Введение

Начнем курс теории вероятностей с описания простых, называемых стохастическими, экспериментов, в связи с которыми возникает интуитивное понятие вероятности. Анализ этих экспериментов позволит лучше ориентироваться в дальнейшем формальном построении теории вероятностей.

1. Эксперимент с конечным числом равновероятных исходов.

Рассмотрим эксперимент, который выполняется при соблюдении некоторого комплекса условий $\mathcal{S}$. Предположим, что при фиксированном $\mathcal{S}$ эксперимент может быть повторен неограниченное число раз, но при повторении результаты его могут быть различными. Речь идет об эксперименте со случайным исходом. Теория вероятностей изучает математические модели таких экспериментов, называемых далее стохастическими.

В качестве примера стохастического эксперимента рассмотрим бросание игральной кости¹⁾. Исход эксперимента, такой, например, как выпадение одной из шести граней, можно считать непредсказуемым (случайным); всякий такой исход назовем элементарным. Исходом эксперимента в данном случае необязательно считать выпадение одной из граней. Можно, например, условиться, что эксперимент имеет не шесть элементарных, а лишь три исхода: A_1 — «выпадение одной из граней 1, 2 или 3», A_2 — «выпадение одной из граней 4 или 5» и, наконец, A_3 — «выпадение грани 6». Дело в том, что в рассматриваемом эксперименте, в силу *симметрии игральной кости*, ни один из элементарных исходов нельзя считать более вероятным, чем любой другой, и если n — число всех различных элементарных исходов (в данном случае 6), то естественно считать все элементарные исходы равновероятными и каждому приписать одинаковую вероятность, равную $1/n$. Так определенная вероятность на практике призвана оценивать частоту каждого элементарного исхода в серии большого числа взаимно независимых повторений эксперимента, называемых *испытаниями*, если под часто-

¹⁾ Игральная кость — куб из однородного материала, шесть граней которого перенумерованы.

той элементарного исхода понимать отношение числа его появлений к числу испытаний.

Вслед за этим можно «вычислить» вероятность любого результата эксперимента. Именно, если $n(A)$ — число элементарных исходов, приводящих к результату эксперимента A , то естественно вероятность $P(A)$ исхода A определить как отношение числа элементарных исходов, приводящих к A , к числу n всех элементарных исходов¹⁾, то есть равенством $P(A) = n(A)/n$.

Итак, вероятность выпадения каждой грани при бросании кости равна $1/6$, вероятность исхода A — «выпадение либо 2, либо 3» — равна $(1 + 1)/6 = 1/6 + 1/6 = 1/3$ и т. д. При этом в понятие «вероятность» вкладывается следующий интуитивный смысл: при многократном повторении эксперимента мы ожидаем, что отношение фактического числа $N(A)$ исходов A к общему числу N испытаний с увеличением N приближается и остается близким к значению вероятности $P(A) = n(A)/n$. В данном случае следует ожидать, что каждая грань при большом числе бросаний будет выпадать примерно в одной шестой всех исходов, а результат эксперимента A_1 будет наблюдаться втрое чаще, т. е. примерно в половине испытаний. Это свойство стохастического эксперимента принято называть устойчивостью частот.

Так это или не так, в каждом конкретном случае может свидетельствовать лишь реальный эксперимент. Накопленные на практике многочисленные наблюдения действительно подтверждают факт устойчивости частот в рассмотренном эксперименте. При большом числе бросаний частота выпадения каждой грани и в самом деле близка к $1/6$. Однако не следует думать, что всякий эксперимент со случайным исходом при фиксированном «комплексе условий S » обладает устойчивостью частот, или, как еще говорят, — статистической устойчивостью. В теории вероятностей, как правило, речь идет об экспериментах со свойством статистической устойчивости их результатов.

2. Геометрические вероятности. Интуитивное представление о вероятности может быть составлено также в связи со следующим мысленным стохастическим экспериментом. Пусть на отрезок $[a, b]$ длины $l = b - a$ «наугад» бросается точка. Какова вероятность того, что точка попадет на отрезок $[\alpha, \beta]$, содержащийся в $[a, b]$?

Ответ в данном случае очевиден. Поскольку вероятность попасть в $[\alpha, \beta]$ не зависит от того, где именно на $[a, b]$ расположен отрезок $[\alpha, \beta]$, то искомая вероятность $P(A)$ равна $(\beta - \alpha)/l$, т. е. отношению длин отрезков $[\alpha, \beta]$ и $[a, b]$. Здесь A обозначает факт попадания точки на $[\alpha, \beta]$. Ответ будет таким же, если вместо отрезка $[\alpha, \beta]$, выбрать любое подмножество отрезка $[a, b]$, лишь бы для него можно было определить длину и последняя равнялась бы $\beta - \alpha$. Понятно, что такой вывод

¹⁾ Такая конструкция вероятности лежит в основе так называемой классической теоретико-вероятностной модели, которая будет рассмотрена в §1.2.

целиком обусловлен интерпретацией условий эксперимента, согласно которым точка бросается на $[a, b]$ «наугад».

Для рассмотренного эксперимента характерно, что возможно бесконечное множество (даже континуум) элементарных исходов — попаданий точки на отрезок $[a, b]$. В таких случаях вероятность удобно задавать с помощью так называемой плотности вероятности. В примере с бросанием точки плотность вероятности $p(\cdot)$ определяется равенством $p(x) = 1/l$, $x \in [a, b]$, при этом $P(A) = \int_a^b dx/l = \int_a^b p(x) dx$, $\int_a^b p(x) dx = 1$.

Возвращаясь к эксперименту с игральной костью, заметим, что теперь суммированию вероятностей элементарных исходов, приводящих к исходу A , отвечает интегрирование плотности вероятности по множеству, определяющему исход A , а вероятность каждого элементарного исхода равна нулю.

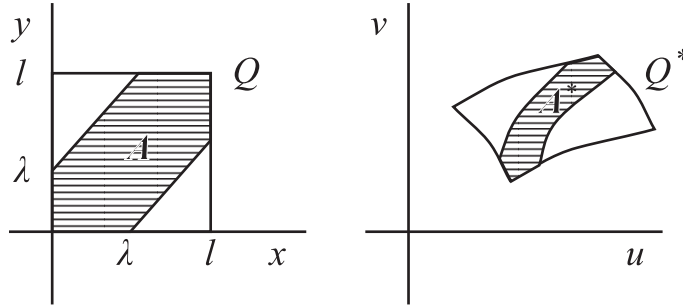


Рис. 1.1. Квадрат Q , область A на плоскости $\{(x, y)\}$ и их образы Q^* , A^* на плоскости $\{(u, v)\}$.

Пусть на отрезок $[a, b]$ наугад бросаются две точки. Какова вероятность того, что расстояние между ними окажется не больше λ , $0 \leq \lambda \leq l$? Задача, очевидно, эквивалентна следующей: в квадрат $Q = \{(x, y) : 0 \leq x \leq l, 0 \leq y \leq l\}$ наугад бросается точка (x, y) , какова вероятность того, что $|x - y| \leq \lambda$? Иначе говоря, какова вероятность того, что точка попадет в заштрихованную область квадрата на рис. 1.1? Искомая вероятность, очевидно, равна отношению площади заштрихованной области к площади квадрата Q :

$$P(A) = \int_A \frac{dxdy}{l^2} = (l^2 - (l - \lambda)^2)/l^2.$$

Заметим, что если $u = u(x, y)$, $v = v(x, y)$, $(x, y) \in \{(x, y)\}$, — криволинейные координаты на плоскости $\{(x, y)\}$ и на плоскости $\{(u, v)\}$ квадрат Q представлен фигурой Q^* , то вероятность попасть в указанную на рис. 1.1 область A квадрата Q может быть получена по формуле $P(A) = \int_A \frac{dxdy}{l^2} = \int_{A^*} \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \frac{dudv}{l^2}$. В свою очередь, послед-

нее равенство можно интерпретировать следующим образом: $P(A)$ — вероятность точки (u, v) попасть в область A^* , если характер «бросаний» точки в область Q^* «контролируется» плотностью вероятности $p(u, v) = \frac{1}{I^2} \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$, $(u, v) \in Q^*$, заданной на плоскости $\{(u, v)\}$. В область Q^* точка бросается не «наугад».

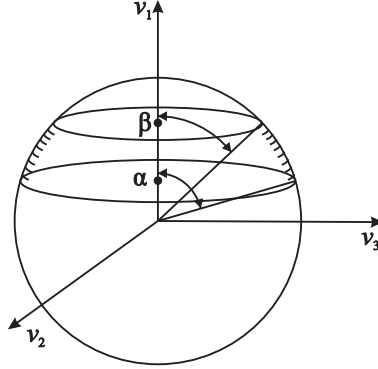


Рис. 1.2. Сфера $v_1^2 + v_2^2 + v_3^2 = 2E/m$ и полоска на ней концов векторов (v_1, v_2, v_3) , удовлетворяющих условию $\alpha \leq v_1 \leq \beta$.

Рассмотрим частицу с энергией $E = mv^2/2$, движущуюся в случайном направлении. Пусть (v_1, v_2, v_3) — вектор скорости частицы в некоторой декартовой системе координат. Какова вероятность того, что $\alpha \leq v_1 \leq \beta$? Поскольку по условию задачи «все направления равновероятны», искомая вероятность равна отношению площади заштрихованной полоски к площади сферы радиуса $v = \sqrt{v_1^2 + v_2^2 + v_3^2}$, изображенных на рис. 1.2. Последняя равна $4\pi v^2 = 8\pi E/m$, а площадь полоски дается интегралом

$$v^2 \int_0^{2\pi} d\varphi \int_{\vartheta_\beta}^{\vartheta_\alpha} \sin\vartheta d\vartheta = v^2 \cdot 2\pi \frac{\beta - \alpha}{v} = 2\pi \sqrt{\frac{2E}{m}} (\beta - \alpha), \quad -v \leq \alpha \leq \beta \leq v,$$

поэтому $P(\alpha, \beta) = (\beta - \alpha) / \left(2\sqrt{\frac{2E}{m}} \right)$.

Для системы n одинаковых частиц с фиксированной полной энергией $E = \sum_{i=1}^n mv_{(i)}^2/2$ вероятность того, что $\alpha \leq v_{(i)1} \leq \beta$, может быть подсчитана вполне аналогично, если рассмотреть полоску $\alpha \leq v_{(i)1} \leq \beta$ на $(3n - 1)$ -мерной сфере $\sum_{i=1}^n (v_{(i)1}^2 + v_{(i)2}^2 + v_{(i)3}^2) = 2E/m$.

Наконец, если n — число наблюдаемых частиц, движущихся в случайных направлениях, энергия каждой из которых равна $mv^2/2$ и

$n_1(\alpha, \beta)/n$ — относительная доля тех из них, для которых $\alpha \leq v_1 \leq \beta$, то при $n \rightarrow \infty$ вероятность любого отличия $P(\alpha, \beta)$ от $n_1(\alpha, \beta)/n$ стремится к нулю. Этот замечательный результат, известный как закон больших чисел, в дальнейшем будет подробно рассмотрен.

Приступим теперь к более точным построениям.

1.1. Пространство элементарных событий. Алгебра событий

В рассмотренных во введении примерах наглядно выступают все основные элементы общей теоретико-вероятностной схемы. В этом параграфе они будут выделены и точно определены.

1.1.1. Элементарные события. События. Пространство элементарных событий. В общей теоретико-вероятностной схеме для каждого эксперимента со случайным исходом должны быть указаны все элементарные исходы, отвечающие следующему требованию: в результате эксперимента непременно происходит один и только один из этих исходов. Каждый такой исход принято называть элементарным событием; обозначают элементарные события обычно буквой ω . По смыслу элементарные события неразложимы на «более элементарные».

В эксперименте с игральной костью элементарными событиями являются выпадения граней «1», «2», ..., «6». При этом считается, что не может выпасть ребро или вершина кости, хотя в принципе такое явление возможно. В эксперименте с бросанием точки на отрезок $[a, b]$ элементарным событием является точка на $[a, b]$. Соответственно в эксперименте с бросанием двух точек элементарным событием является пара точек на $[a, b]$, или точка в квадрате $[a, b] \times [a, b]$. Наконец в примере с частицей элементарным событием является точка на сфере радиуса v .

Множество всех элементарных событий в теории вероятностей принято называть пространством элементарных событий¹⁾. Пространство элементарных событий будем обозначать буквой Ω . Иногда элементарные события называют точками пространства элементарных событий, хотя на самом деле элементарные события суть одноточечные подмножества Ω .

Всякий исход стохастического эксперимента в теории вероятностей принято называть событием. Среди всех событий элементарные события выделяются таким образом, что для каждого события A и каждого элементарного события $\{\omega\}$ известно, влечет $\{\omega\}$ наступление A или не влечет. Тем самым совокупность всех тех $\{\omega\}$, которые влекут A , полностью характеризуют A . Обратно: произвольное множество A

¹⁾ Элементарный исход эксперимента, по определению, может быть зарегистрирован. Их множество называется выборочным пространством. Далее выборочное пространство отождествляется с пространством элементарных событий.

точек $\omega \in \Omega$ можно рассматривать как событие A , которое происходит или нет в зависимости от того, принадлежит соответственно или нет множеству A элементарное событие $\{\omega\}$, представляющее данный исход эксперимента ¹⁾.

Иными словами, событие A можно считать подмножеством Ω , состоящим из точек $\omega \in \Omega$, представляющих те исходы $\{\omega\}$ эксперимента, при которых происходит A . Далее по этой причине не делается различий между событием A и соответствующим подмножеством $A \subset \Omega$. Событие называется невозможным, если оно никогда не происходит при повторении эксперимента, ему соответствует пустое подмножество Ω , обозначаемое $\emptyset$. С другой стороны событие Ω происходит при любом повторении эксперимента и называется достоверным.

В приведенных во введении примерах стохастических экспериментов события рассматривались как подмножества соответствующих пространств элементарных событий. В эксперименте с игральной костью были выделены три события: $A_1 = \{\langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle\}$, $A_2 = \{\langle 4 \rangle, \langle 5 \rangle\}$ и $A_3 = \{\langle 6 \rangle\}$. В этом случае, например, любое из элементарных событий $\{\omega_1\} = \langle 1 \rangle$, $\{\omega_2\} = \langle 2 \rangle$, $\{\omega_3\} = \langle 3 \rangle$ влечет A_1 . В эксперименте с бро-

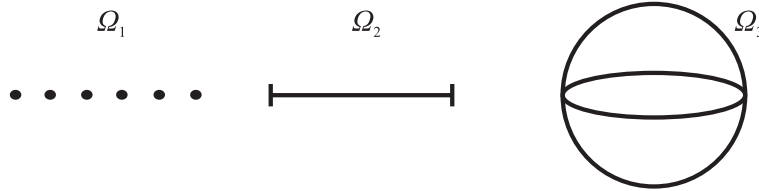


Рис. 1.3. Ω_1 — пространство элементарных событий в эксперименте с игральной костью, Ω_2 — в эксперименте с бросанием точки на отрезок, Ω_3 — в примере с частицей, движущейся в случайном направлении

санием двух точек на отрезок $[a, b]$ событием является заштрихованная область A квадрата $Q = \Omega$. Наконец в примере с частицей событием является полоска на сфере (см. рис. 1.1, 1.2, 1.3).

1.1.2. Операции над событиями. Рассмотрим математические формулировки естественных операций над событиями и их теоретико-множественные аналоги. Приведенные ниже определения и свойства операций над событиями характеризуют *алгебраическую структуру* любой теоретико-вероятностной схемы.

1. Если событие A происходит всякий раз, когда происходит событие B , то будем говорить, что событие A является следствием B , и писать $B \subset A$ или $A \supset B$. В теоретико-множественных терминах

¹⁾ Если ω обозначает точку Ω , т.е. если $\omega \in \Omega$, то соответствующее ω элементарное событие следует обозначать $\{\omega\}$ как одноточечное подмножество Ω и писать $\{\omega\} \subset \Omega$. Далее точки Ω и соответствующие одноточечные подмножества Ω будем обозначать ω и $\{\omega\}$ соответственно.

это означает, что каждая точка $\omega \in B$ содержится в A , или иначе, B является подмножеством A . Эта связь аналогична связи между событиями и элементарными событиями: $\{\omega\}$ влечет A , если $\omega \in A$ или $\{\omega\} \subset A$.

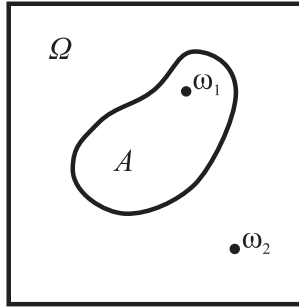


Рис. 1.4. Элементарное событие $\{\omega_1\}$ влечет событие A , $\{\omega_1\} \subset A$, $\{\omega_2\}$ — не влечет, $\{\omega_2\} \not\subset A$.

2. Если $A \subset B$ и $B \subset A$, то события A и B происходят или не происходят одновременно. В таком случае будем писать $A = B$; множества A и B при этом совпадают.

3. Событие, состоящее в том, что не происходит событие A , называется *противоположным* событию A и обозначается $\bar{A}$. Множество $\bar{A}$ состоит из точек Ω , не принадлежащих A , и называется дополнением множества A (в Ω). Очевидно, $\bar{\bar{A}} = A$.

4. Если событие A не содержит ни одного элементарного события, то A называется *невозможным* и обозначается $\emptyset$. Противоположным $\emptyset$ является, очевидно, событие Ω , которое происходит всякий раз и называется *достоверным*. Наоборот, $\bar{\Omega} = \emptyset$. $\emptyset$, очевидно, пустое подмножество Ω .

5. Событие C , происходящее тогда и только тогда, когда происходят события A и B , называется *произведением*, или пересечением событий A и B , и обозначается AB или $A \cap B$. Множество C состоит из точек, принадлежащих как множеству A , так и множеству B , и называется пересечением множеств A и B . При этом $A \cap B$ обозначает пересечение A и B .

6. События A и B называются *несовместными*, если их одновременное наступление невозможно, т. е. если $A \cap B = \emptyset$. Не совместным событиям отвечают непересекающиеся множества.

7. Событие C , состоящее в наступлении хотя бы одного из событий A или B , называется *объединением*, или *суммой* событий A и B . Для объединения будем использовать обозначение $A \cup B$, но в том случае, когда $A \cap B = \emptyset$, условимся писать $C = A + B$. В теоретико-множественных терминах $A \cup B$ — множество, состоящее из тех точек Ω , которые принадлежат хотя бы одному из множеств A или

B . Множество $A \cup B$ также называется объединением или суммой, множеств A и B .

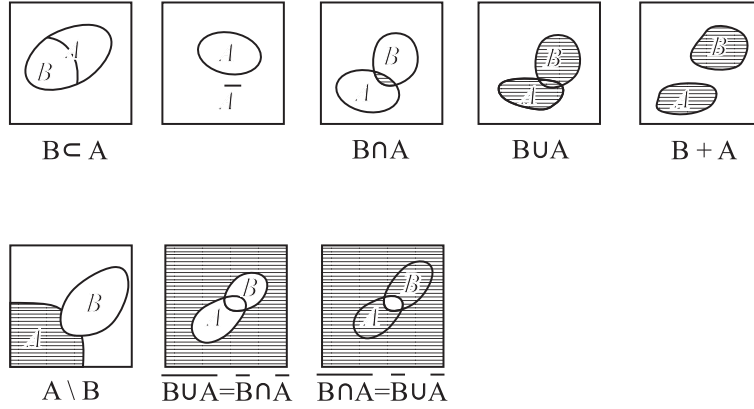


Рис. 1.5. Операции над событиями.

Замечание 1.1.1. Пересечение и объединение определяется для произвольного числа событий. Например, событие $C = A \cap B \cap \dots$ состоит в том, что происходят все события $A, B, \dots$; событие $C = A \cup B \cup \dots$ состоит в том, что происходит хотя бы одно из событий $A, B, \dots$. Операции объединения $\cup$ и пересечения $\cap$, очевидно, ассоциативны и коммутативны по определению: $(A \cup B) \cup C = A \cup (B \cup C)$, $A \cup B = B \cup A$, $(A \cap B) \cap C = A \cap (B \cap C)$, $A \cap B = B \cap A$ для любых событий A, B и C . Это означает, что имеют смысл обозначения $\bigcup_{\alpha \in S} A_\alpha$

и $\bigcap_{\alpha \in S} A_\alpha$ для событий, происходящих тогда и только тогда, когда происходит по крайней мере одно A_α , $\alpha \in S$, и соответственно происходят все A_α , $\alpha \in S$. Заметим, что с точки зрения теоретико-множественной $\bigcup_{\alpha \in S} A_\alpha = \{\omega \in \Omega, \exists \alpha \in S, \omega \in A_\alpha\}$ — множество тех $\omega \in \Omega$, для каждого из которых можно указать $\alpha \in S$, при котором $\omega \in A_\alpha$;

$\bigcap_{\alpha \in S} A_\alpha = \{\omega \in \Omega, \forall \alpha \in S, \omega \in A_\alpha\}$ — множество тех $\omega \in \Omega$, для каждого из которых $\omega \in A_\alpha$ при всех $\alpha \in S$;

8. Событие C , состоящее в том, что событие A происходит, а событие B не происходит, называется *разностью* событий A и B и обозначается $A \setminus B$. В теоретико-множественных терминах множество $A \setminus B = A \cap \overline{B}$ состоит из точек $\omega \in \Omega$, принадлежащих множеству A и не принадлежащих множеству B , и называется разностью множеств A и B . Очевидно, $\overline{A} = \Omega \setminus A$

1.1.3. Свойства операций над событиями. Принцип двойственности. Отметим простейшие свойства операций над событиями. По определению: $\overline{\overline{A}} = A$, $\overline{A} = \Omega \setminus A$ и $A \setminus B = A \cap \overline{B} = \overline{B} \setminus \overline{A}$. Кроме того,

операции над событиями $\cup$ и $\cap$ взаимно дистрибутивны:

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C), \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C). \end{aligned} \quad (1.1)$$

Наконец, в теории вероятностей и ее приложениях важную роль играет так называемый *принцип двойственности*, который может быть выражен следующими соотношениями:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}. \quad (1.2)$$

Докажем, например, первое равенство в (1.2), рассматривая его как равенство событий. Событие $A \cup B$ согласно определению 3 состоит в том, что не происходит событие $A \cup B$, состоящее в наступлении хотя бы одного из событий A или B (определение 7). Но это буквально и означает, что не происходит ни A , ни B . Другими словами, согласно определениям 3 и 5, происходит событие $\overline{A \cap B}$. Следовательно, событие $A \cup B$ влечет событие $\overline{A \cap B}$: $A \cup B \subset \overline{A \cap B}$. Наоборот, если происходит событие $\overline{A \cap B}$, то не происходят ни A , ни B . Следовательно, не происходит событие $A \cup B$, т. е. происходит $A \cup B$. Поэтому $\overline{A \cap B} \subset A \cup B$, и, следовательно, верно первое равенство в (1.2). Второе равенство предлагается проверить читателю.

Докажем теперь то же самое в терминах операций над множествами. Для этого достаточно показать, что всякий элемент множества $A \cup B$ содержится в $\overline{A \cap B}$ (и тем самым $A \cup B \subset \overline{A \cap B}$), и, наоборот, всякий элемент множества $\overline{A \cap B}$ содержится в $A \cup B$ (т. е. $\overline{A \cap B} \subset A \cup B$). Доказательство следует из системы соотношений:

$$\begin{aligned} \omega \in \overline{A \cap B} &\Leftrightarrow \omega \notin A \cap B \Leftrightarrow \omega \notin A, \omega \notin B \Leftrightarrow \\ &\Leftrightarrow \omega \in \overline{A}, \omega \in \overline{B} \Leftrightarrow \omega \in \overline{A \cap B}, \end{aligned} \quad (1.3)$$

в которых символ $\Leftrightarrow$ следует читать как «эквивалентно». Читая соотношения (1.3) слева направо, получим включение $\overline{A \cap B} \subset \overline{A \cap B}$. Если (1.3) прочесть справа налево, то найдем, что $\overline{A \cap B} \subset A \cup B$. Следовательно, $\overline{A \cap B} = A \cup B$.

Разумеется, принцип двойственности справедлив для любого множества событий:

$$\overline{\bigcup_{\alpha \in S} A_{\alpha}} = \bigcap_{\alpha \in S} \overline{A_{\alpha}}; \quad \overline{\bigcap_{\alpha \in S} A_{\alpha}} = \bigcup_{\alpha \in S} \overline{A_{\alpha}};$$

здесь символ $\bigcup_{\alpha \in S} (\bigcap_{\alpha \in S})$ означает объединение (пересечение) множества событий A_{α} , отличающихся индексом $\alpha \in S$, который может пробегать и несчетное множество значений.

К принципу двойственности следует отнести еще одно соотношение

$$A \subset B \Leftrightarrow \overline{A} \supset \overline{B}, \quad (1.4)$$

доказательство которого очевидно.

Роль принципа двойственности в теории вероятностей состоит в том, что для всякого утверждения, относящегося к некоторой системе событий, может быть сформулировано эквивалентное ему двойственное утверждение, в котором события должны быть заменены на соответственно противоположные, объединения — на пересечения и наоборот, и учтено соотношение (1.4). Например, утверждение $(A \cup B) \cap C \subset M \setminus N$ эквивалентно $(A \cup B) \cap C \supset M \setminus N$, а последнее может быть преобразовано к виду $(A \cap \bar{B}) \cup \bar{C} \supset N \cup \bar{M}$, поскольку $(A \cup B) \cap C = (A \cup B) \cup \bar{C} = (\bar{A} \cap \bar{B}) \cup \bar{C}$ и $M \setminus N = M \cap \bar{N} = \bar{M} \cup \bar{N}$.

Возвращаясь к соотношениям дистрибутивности (1.1), заметим, что и для произвольных систем событий

$$A \cup \left(\bigcap_{\alpha \in S} A_\alpha \right) = \bigcap_{\alpha \in S} (A \cup A_\alpha); \quad A \cap \left(\bigcup_{\alpha \in S} A_\alpha \right) = \bigcup_{\alpha \in S} (A \cap A_\alpha). \quad (1.5)$$

В частности,

$$A \cap (B + C + D + \dots) = A \cap B + A \cap C + A \cap D + \dots$$

Докажем, например, первое соотношение в (1.5). Событие $A \cup \left(\bigcap_{\alpha \in S} A_\alpha \right)$ происходит тогда и только тогда, когда происходит либо событие A , либо событие $\bigcap_{\alpha \in S} A_\alpha$ (т. е. все события A_α , $\alpha \in S$), либо, наконец, A и $\bigcap_{\alpha \in S} A_\alpha$. С другой стороны, событие $\bigcap_{\alpha \in S} (A \cup A_\alpha)$ происходит тогда и только тогда, когда происходят все события $A \cup A_\alpha$, $\alpha \in S$, т. е. тогда и только тогда, когда происходят либо A , либо все A_α , $\alpha \in S$, либо, наконец, когда происходят A и все A_α , $\alpha \in S$. Следовательно, события $A \cup \left(\bigcap_{\alpha \in S} A_\alpha \right)$ и $\bigcap_{\alpha \in S} (A \cup A_\alpha)$ равны.

Второе равенство в (1.5) читателю следует доказать самостоятельно.

1.1.4. Алгебра событий. Рассмотренные свойства операций над событиями являются алгебраическими. В теории вероятностей принимается, что класс $\mathcal{F}$ всех событий должен удовлетворять следующим требованиям.

1. Для каждой пары событий A и B из включений $A \in \mathcal{F}$ $B \in \mathcal{F}$ следует включение $A \cup B \in \mathcal{F}$. Иными словами, класс $\mathcal{F}$ вместе с каждой парой событий содержит их объединение.

2. Вместе с каждым событием A класс $\mathcal{F}$ содержит противоположное событие $\bar{A}$.

Если класс $\mathcal{F}$ не пуст, то отсюда следует, что $\Omega \in \mathcal{F}$, так как $\Omega = A + \bar{A}$. Следовательно, $\emptyset \in \mathcal{F}$, поскольку $\emptyset = \bar{\Omega}$. Наконец, так как согласно принципу двойственности $A \cap B = \overline{\bar{A} \cup \bar{B}}$, то класс $\mathcal{F}$

вместе с каждой парой событий A и B содержит их пересечение $A \cap B$ и разность $A \setminus B = A \cap \overline{B}$.

Класс $\mathcal{F}$ событий, удовлетворяющий условиям 1 и 2, называется *алгеброй событий*.

Конструкция алгебры событий позволяет охарактеризовать множество всех возможных результатов любого эксперимента со случайным исходом, если *множество Ω его элементарных исходов конечно*. Например, в эксперименте с игральной костью Ω состоит из шести элементарных событий, а $\mathcal{F}$ состоит из всех подмножеств Ω . Поскольку $\mathcal{F}$ содержит пустое подмножество $\emptyset$, $6 = C_6^1$ одноточечных подмножеств, $15 = C_6^2$ двухточечных, $20 = C_6^3$ трехточечных, ..., одно (C_6^6) шеститочечное, то $\mathcal{F}$ состоит из $2^6 = 1 + C_6^1 + C_6^2 + \dots + C_6^6 = 64$ событий. И вообще, если Ω состоит из n элементарных событий, то $\mathcal{F}$ состоит, очевидно, из $2^n = C_n^0 + C_n^1 + \dots + C_n^n$ событий.

Совсем не так просто обстоит дело в эксперименте с бросанием точки на отрезок. Здесь в качестве системы событий в дальнейшем придется выделить специальный класс подмножеств, более широкий, чем алгебра. Но, с другой стороны, при попытке использовать в качестве событий все подмножества отрезка мы столкнулись бы с утратой свойств непрерывности вероятности, см. §1.3.3

1.2. Классическая теоретико-вероятностная модель

1.2.1. Определение классической вероятности. Примеры.

Пусть Ω – пространство элементарных событий некоторого случайного эксперимента. Предположим, что структура эксперимента такова, что в Ω можно указать n событий $A_1, A_2, \dots, A_n$, обладающих следующими свойствами.

1. События $A_1, \dots, A_n$ попарно несовместны в том смысле, что никакие два из них не могут произойти одновременно. Иначе говоря, $A_i \cap A_j = \emptyset$, $i \neq j$, $i, j = 1, \dots, n$.

2. $A_1, \dots, A_n$ образуют полную группу событий в том смысле, что при любом исходе эксперимента хотя бы одно из них непременно происходит. Это означает, что $A_1 + \dots + A_n = \Omega$.

3. События $A_1, \dots, A_n$ равновероятны, или, иначе говоря, ни одно из них нельзя считать более вероятным, чем любое из остальных. Это, разумеется, доматематическое требование, и не следует стремиться переформулировать его в терминах каких-либо более элементарных свойств. Можно считать, что оно отражает свойство некоторой относительной симметрии событий $A_1, \dots, A_n$, подсказанное здравым смыслом. Именно такая ситуация возникла в эксперименте с игральной костью, когда выпадения всех граней были объявлены равновероятными.

В так называемой классической модели события $A_1, \dots, A_n$, удовлетворяющие условиям 1.1 — 1.3, называются *полной группой попарно несовместных равновероятных событий*.

Вероятность в классической модели определяется лишь для тех исходов эксперимента, которые могут быть представлены в виде объединений некоторых из событий A_i , $i = 1, \dots, n$. Именно если

$$A = A_{i_1} + \dots + A_{i_k} \quad (2.1)$$

и все слагаемые в (2.1) различны, то вероятность события определяется равенством

$$P(A) = k/n, \quad (2.2)$$

в котором k равно числу слагаемых в сумме (2.1). Таково *классическое определение вероятности*.

Для того чтобы определение (2.2) можно было считать корректным, достаточно доказать единственность разложения (2.1). Но для любого события A согласно условиям 1, 2 и свойству дистрибутивности (1.5)

$$A = A \cap \Omega = A \cap (A_1 + \dots + A_n) = A \cap A_1 + \dots + A \cap A_n.$$

Поэтому в рассматриваемом случае разложения (2.1) $A \cap A_j$ либо пусто, если j не совпадает ни с одним из i_s , $s = 1, \dots, k$, либо $A \cap A_j = A_{i_s}$, если $j = i_s$.

В данном случае алгебру событий $\mathcal{F}$ образуют 2^n событий, C_n^k из которых представимы равенством (2.1) при $k = 1, \dots, n$, и невозможное событие $\emptyset$; вероятность определена равенством (2.2) для любого $A \in \mathcal{F}$, пустому множеству отвечает $k = 0$.

Традиционно приложениям классического определения вероятности сопутствует следующая терминология. Эксперимент называют испытанием, полную группу попарно несовместных равновероятных событий называют полной группой возможных исходов испытания, а те из возможных исходов, из которых складывается событие A , называют исходами, благоприятствующими появлению A . В этих терминах согласно определению (2.2) $P(A)$ равно отношению числа исходов, благоприятствующих появлению A , к числу всех возможных исходов.

Как правило, отыскание вероятностей в классической модели сводится к комбинаторным вычислениям. Рассмотрим примеры, позволяющие уяснить технику вычисления классических вероятностей. Напомним вначале простейшие комбинаторные формулы.

Размещения. Перестановки. Имеется n различных объектов $x_1, \dots, x_n$. Сколькими способами можно образовать последовательность $x_{j_1}, \dots, x_{j_r}$, $r < n$, различных объектов?

Объект x_{j_1} можно выбрать n способами. Если x_{j_1} выбран, x_{j_2} можно выбрать $n - 1$ способом из оставшихся объектов и т. д. Всего, таким образом, существуют $n(n - 1) \dots (n - r + 1) = A_n^r$ способов образовать последовательность из r объектов, выбирая объекты из совокупности $x_1, \dots, x_n$.

Иначе эту задачу можно сформулировать следующим образом: сколькими способами можно разместить r из n различных объектов по r местам? A_n^r называется числом размещений из n по r .

Если $r = n$, то $A_n^r = n(n-1)\dots 2 \cdot 1 = n!$ называется числом перестановок (n различных объектов). Поскольку для $n \geq 1$

$$1 < \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} < e^{1/(12n)},$$

то $n!$ при больших n можно вычислить по формуле Стирлинга: $n! \sim \sqrt{2\pi n} n^n e^{-n}$. Здесь символ $\sim$ означает, что отношение левой и правой частей стремится к единице при $n \rightarrow \infty$. Для $n = 0$ полагают $0! = 1$.

Сочетания. Имеется n различных объектов $x_1, \dots, x_n$. Выбирая объекты из $x_1, \dots, x_n$, сколькими способами можно образовать множество X_r , из r объектов, $r \leq n$?

Поскольку в данном случае порядок объектов, образующих множество X_r , несуществен, то искомое число способов равно $C_n^r = A_n^r / r! = \frac{n!}{r!(n-r)!}$. C_n^r называется числом сочетаний из n по r .

Выбор с возвращением. Имеется n различных объектов $x_1, \dots, x_n$, из которых последовательно выбирается объект, регистрируется и возвращается обратно. Сколькими способами может быть образована выборка (совокупность) объектов $x_{j_1}, \dots, x_{j_r}$, зарегистрированная за r шагов?

Поскольку каждый раз объект может быть выбран n способами, существует всего n^r способов образовать выборку $x_{j_1}, \dots, x_{j_r}$.

Задача о днях рождения. Предположим, что в аудитории n студентов. Какова вероятность того, что хотя бы у двоих совпадают дни рождения? Разумеется, такая постановка задачи требует уточнения. Будем считать, что в году 365 дней, у каждого студента есть день рождения, причем им может оказаться любой из 365 дней с одной и той же вероятностью.

Заметим, что для групп из $n = 366$ и более человек искомая вероятность равна единице, т. е. для $n \geq 366$ по меньшей мере у двоих непременно совпадают дни рождения. Поэтому рассмотрим случай $n \leq 365$. Для группы из n человек возможно $(365)^n$ комбинаций дней рождения, поскольку таких возможностей для каждого человека 365. Все эти комбинации образуют полную группу из попарно несовместных, равновероятных событий, причем вероятность каждой комбинации равна $(365)^{-n}$. Число различных комбинаций дней рождения, в которых ни один день рождения не встречается более одного раза, равно $\bar{n} = 365 \cdot 364 \cdot \dots \cdot (365 - n + 1)$. Это число получено следующим образом: первый день рождения можно выбрать 365 способами, после этого второй день рождения можно выбрать $(365 - 1)$ способами, ..., наконец, n -ый день рождения можно выбрать $(365 - n + 1)$ способами. Следовательно, согласно классическому определению, вероятность то-

го, что в группе из n студентов на каждый день приходится не более одного дня рождения, равна $P_n = \bar{n}(365)^{-n}$.

Подсчитав число комбинаций, в которых нет совпадающих дней рождения, нетрудно сообразить, что во всех оставшихся комбинациях имеются по меньшей мере два совпадающих дня рождения. Таких комбинаций $(365)^n - \bar{n}$. Следовательно, искомая вероятность равна $[(365)^n - \bar{n}]/(365)^n = 1 - P_n$, $n = 1, \dots, 365$, некоторые ее значения представлены в таблице.

5	10	20	30	40	50	60	n
0,027	0,117	0,411	0,706	0,891	0,970	0,994	$1 - P_n$

Как видно из приведенной таблицы, искомая вероятность становится довольно близкой к единице для групп, существенно меньших 366 человек. Такой результат априори далеко не очевиден.

Гипергеометрическое распределение. Дана совокупность n объектов, среди которых k отмеченных (например, бракованных изделий, выигрышных билетов и т. п.). Выбираются наугад $n_1 \leq n$ объектов. Какова вероятность того, что среди них окажется k_1 отмеченных?

Выбрать n_1 объектов из n можно $C_n^{n_1}$ различными способами. k_1 отмеченных объектов из общего их числа k можно выбрать $C_k^{k_1}$ способами, причем каждому такому способу соответствует $C_{n-k}^{n_1-k_1}$ способов добрать еще $n_1 - k_1$ объектов до общего числа n_1 выбирая их из $n - k$ неотмеченных. Следовательно, число способов, благоприятствующих появлению k_1 отмеченных объектов среди n_1 выбранных, равно $C_k^{k_1} C_{n-k}^{n_1-k_1}$. Поэтому искомая вероятность равна

$$P_{k,n}(k_1, n_1) = \frac{C_k^{k_1} C_{n-k}^{n_1-k_1}}{C_n^{n_1}}, \quad k_1 = 0, \dots, \min(k, n_1). \quad (2.3)$$

Совокупность вероятностей (2.3) носит название *гипергеометрического распределения*.

Применим гипергеометрическое распределение к анализу вероятностей выигрышей при игре «спортлото». В данном случае $n = 49$ (число наименований видов спорта), $k = 6$ (число отмеченных видов спорта), $n_1 = 6$ (число выбранных видов спорта). Следовательно, согласно (2.3) вероятность угадывания k_1 видов спорта равна $P_{6,49}(k_1, 6) = C_6^{k_1} C_{49-6}^{6-k_1} / C_{49}^6$. В том числе вероятность угадывания всех шести видов спорта равна $P_{6,49}(6, 6) = 1/C_{49}^6 \approx 7,15 \cdot 10^{-8}$, а пяти - $P_{6,49}(5, 6) = C_6^5 C_{43}^1 / C_{49}^6 \approx 1,84 \cdot 10^{-5}$.

Рассмотрим несколько важных задач на размещения, возникающих при изучении некоторых систем частиц в физике и в статистической механике.

Система Максвелла–Больцмана характеризуется как система r различных частиц, каждая из которых может находиться в одной из n ячеек (состояний) вне зависимости от того, где при этом находятся остальные частицы.

В такой системе возможно всего n^r различных размещений r частиц по n ячейкам. Если при этом все такие размещения (состояния системы) считаются равновероятными, то говорят о *статистике Максвелла-Больцмана*. Вероятность каждого состояния системы r различных частиц равна n^{-r} .

Система Бозе–Эйнштейна определяется как система r неразличимых (тождественных) частиц, каждая из которых независимо от остальных может находиться в одной из n ячеек (состояний).

Поскольку частицы неразличимы, каждое состояние такой системы задастся «числами заполнения» $r_1, r_2, \dots, r_n$, где r_j — число частиц в j -й ячейке. Подсчитаем число различных состояний системы (т. е. число размещений частиц, различающихся лишь числами заполнения). Состояние системы удобно условно представить, как это показано на



Рис. 1.6. Система из $r = 7$ частиц и $n = 5$ ячеек.

рис. 1.6. Здесь изображена конфигурация из r точек (частиц) и $n + 1$ черточек (границ ячеек). Понятно, что каждая такая конфигурация задает размещение неразличимых частиц по ячейкам и наоборот, если задано состояние системы в терминах чисел заполнения, то ему соответствует одна конфигурация. Каждая конфигурация, в свою очередь, полностью определяется положениями внутренних $n - 1$ черточек, которые могут находиться в $n + r - 1$ позициях. Следовательно, имеется всего C_{n+r-1}^{n-1} различных конфигураций и столько же различных состояний рассматриваемой системы частиц.

Если все состояния системы равновероятны, то говорят о статистике Бозе–Эйнштейна. При этом вероятность каждого состояния системы равна $1/C_{n+r-1}^{n-1}$.

Заметим, что если дополнительно потребовать, чтобы в каждом состоянии системы ни одна ячейка не оставалась пустой, то число возможных состояний системы сократится до C_{r-1}^{n-1} . При этом, разумеется, частиц должно быть не меньше, чем ячеек, $r \geq n$.

Чтобы получить этот результат, следует воспользоваться найденным ранее числом состояний системы без ограничений, предварительно «приклеив» к каждой из n черточек справа по одной точке, исключив последнюю $(n + 1)$ -ю (правую) черточку. В таком случае, переставляя черточки, мы будем получать состояния, при которых в каждой ячейке будет находиться не менее одной частицы. Теперь однако для $n - 1$ черточек оказывается не $n + r - 1$, а лишь $r - 1$ вакантных мест (на n меньше).

Система Ферми–Дирака определяется как система Бозе–Эйнштейна, в которой дополнительно действует принцип запрета (принцип Паули), согласно которому, в каждой ячейке должно находиться не более одной частицы.

Поскольку и в этом случае частицы неразличимы, состояния системы характеризуются числами заполнения $r_j = 0, 1, j = 1, \dots, n$. Однако в данном случае непременно число частиц $r \leq n$.

Задать состояние системы можно, указав заполненные ячейки. Последние можно выбрать C_n^r способами; столько же состояний системы Ферми–Дирака. Если все состояния равновероятны, то говорят о *статистике Ферми–Дирака*. Вероятность каждого состояния в таком случае равна $1/C_n^r$.

В классической статистической физике статистике Максвелла–Больцмана подчинены, как известно, системы молекул газа. Системы частиц с целым и полуцелым спином подчиняются соответственно статистикам Бозе–Эйнштейна и Ферми–Дирака.

1.2.2. Свойства классической вероятности. Для того чтобы рассмотреть свойства классической вероятности, удобно несколько упростить и формализовать классическую теоретико-вероятностную модель. Будем считать события $A_1, \dots, A_n$, образующие полную группу попарно несовместных равновероятных событий, точками $\omega_1, \dots, \omega_n$ нового пространства элементарных событий, для которого сохраним прежнее обозначение Ω . Другими словами, положим

$$\Omega = \{\omega_1\} + \{\omega_2\} + \dots + \{\omega_n\} = \{\omega_1, \dots, \omega_n\},$$

где $\{\omega_i\}$ обозначает множество, состоящее из одной точки ω_i (элементарное событие), $i = 1, \dots, n$. Для каждого элементарного события $\{\omega_i\}$ определим вероятность

$$P(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Алгебра $\mathcal{F}$ событий в данном случае состоит из невозможного события $\emptyset$ и всевозможных объединений одноточечных множеств $\{\omega_i\}$, $i = 1, \dots, n$, всего, как уже было отмечено, — из $C_n^0 + C_n^1 + \dots + C_n^n = 2^n$ событий.

Для любого события $A \in \mathcal{F}$ его вероятность $P(A)$ определим равенством

$$P(A) = m/n, \quad (2.4)$$

где m — число элементарных событий, объединение которых есть A .

Формально классическая теоретико-вероятностная модель, очевидно, эквивалентна тройке $(\Omega, \mathcal{F}, P)$, состоящей из пространства элементарных событий Ω , содержащего n точек, алгебры $\mathcal{F}$, содержащей 2^n событий, и вероятности $P(\cdot) : \mathcal{F} \rightarrow [0, 1]$, определенной для всех событий из $\mathcal{F}$ равенством (2.4).

Рассмотрим свойства классической вероятности.

1. Для любого $A \in \mathcal{F} : 0 \leq P(A) \leq 1$ (поскольку $0 \leq m \leq n$).
2. Вероятность достоверного события $A = \Omega$ равна единице (так как для $A = \Omega$ $m = n$). Вероятность невозможного события $\emptyset$ равна нулю (так как для $A = \emptyset$ $m = 0$).

3. Для несовместных событий A_1 и A_2

$$P(A_1 + A_2) = P(A_1) + P(A_2). \quad (2.5)$$

Это простейший вариант так называемой *теоремы сложения вероятностей*. Для доказательства равенства (2.5) достаточно заметить, что если m_1 и m_2 — числа элементарных событий, благоприятствующих соответственно событиям A_1 и A_2 , то в силу несовместности A_1 и A_2 ($A_1 \cap A_2 = \emptyset$) сумма $m_1 + m_2$ является числом элементарных событий, благоприятствующих $A_1 + A_2$. Поэтому

$$P(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

4. Вероятность события $\bar{A}$, противоположного A , равна $P(\bar{A}) = 1 - P(A)$. Доказательство следует из замечания, что $A + \bar{A} = \Omega$, и, следовательно, согласно свойствам 2, 3 $P(A + \bar{A}) = P(A) + P(\bar{A}) = 1$.

5. Если событие A влечет B , $A \subset B$, то $P(B \setminus A) = P(B) - P(A)$ и $P(B) \geq P(A)$. Для доказательства заметим, что $B = A + B \cap \bar{A}$, причем события A и $B \cap \bar{A}$ несовместны (так как несовместны события A и $\bar{A}$ и $B \cap \bar{A} \subset \bar{A}$). Поэтому согласно свойству 3 $P(B) = P(A) + P(B \cap \bar{A})$. Отсюда следует, что $P(B) \geq P(A)$, так как согласно свойству 1 $P(B \cap \bar{A}) \geq 0$, а также равенство $P(B \cap \bar{A}) = P(B \setminus A) = P(B) - P(A)$.

6. Для любых событий A_1 и A_2 имеет место равенство

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \quad (2.6)$$

Действительно, $A_1 \cup A_2 = A_1 + A_2 \setminus A_1 = A_1 + A_2 \setminus (A_1 \cap A_2)$, а так как $A_1 \cap A_2 \subset A_2$, то равенство (2.6) следует из свойств 5, 3.

Равенство (2.6) нетрудно обобщить на случай произвольного конечного числа событий. А именно, вероятность $P_{n,1}$ того, что произойдет хотя бы одно из событий $A_1, \dots, A_n$, т.е. — событие $A_1 \cup \dots \cup A_n$,

$$\begin{aligned} P_{n,1} = P(A_1 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{i < j}^n P(A_i \cap A_j) + \\ &+ \sum_{i < j < k}^n P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n). \end{aligned} \quad (2.7)$$

Действительно, если $B = A_1 \cup \dots \cup A_{n-1}$, то искомая вероятность по только что доказанному равенству (2.6) равна $P_{n,1} = P(B) + P(A_n) - P(B \cap A_n)$. Но согласно свойству дистрибутивности (1.5) $B \cap A_n = (A_1 \cap A_n) \cup (A_2 \cap A_n) \cap \dots \cap (A_{n-1} \cap A_n)$ и если считать, что равенство (2.7) верно для объединения $n - 1$ событий, то

$$\begin{aligned} P(B \cap A_n) &= \sum_i P(A_i \cap A_n) - \sum_{i < j} P(A_i \cap A_j \cap A_n) + \dots \\ &\dots + (-1)^{n-2} P(A_1 \cap \dots \cap A_{n-1} \cap A_n), \end{aligned}$$

где при суммировании индексы $i, j, \dots$ пробегают значения $1, \dots, n-1$. Теперь нетрудно увидеть, что верно и (2.7). Согласно (2.7) вероятность того, что не произойдет ни одного события из $A_1, \dots, A_n$, равна $Q_{n,0} = 1 - P_{n,1}$.

Приведем более общие результаты. Вероятность $Q_{n,m}$ того, что осуществится ровно m событий из $A_1, \dots, A_n$, равна

$$Q_{n,m} = S_m - C_{m+1}^1 S_{m+1} + C_{m+2}^2 S_{m+2} - \dots + (-1)^{n-m} C_n^{n-m} S_n.$$

Вероятность $P_{n,m}$ того, что осуществится не менее m событий из $A_1, \dots, A_n$, равна $P_{n,m} = S_m - C_{m+1}^1 S_{m+1} + C_{m+2}^2 S_{m+2} - \dots + (-1)^{n-m} C_{n-1}^{n-m} S_n$. В этих двух формулах, проверка которых рекомендуются читателю в качестве упражнения, $S_j = \sum_{i_1 < i_2 < \dots < i_j} P(A_{i_1} \cap \dots \cap A_{i_j})$, $i_1, \dots, i_j = 1, \dots, n$; $j = 1, \dots, n$; $S_0 = 1$.

Проиллюстрируем сказанное на примере задачи о совпадениях. Пусть имеется n частиц и n ячеек, причем частицы и ячейки отмечены номерами $1, \dots, n$. Частицы случайно размещаются по ячейкам, по одной в каждую ячейку, причем все такие размещения считаются равновероятными. Назовем совпадением любое событие A_i , состоящее в том, что частица с номером i попадает в ячейку с номером i . Какова вероятность $P_{n,1}$ хотя бы одного совпадения? Какова вероятность $Q_{n,m}$ ровно m совпадений?

Событию A_i (ячейка с номером i занята частицей с номером i) благоприятствуют $(n-1)!$ перестановок $n-1$ частиц по $n-1$ свободным ячейкам. Аналогично событию $A_i \cap A_j$ (ячейки с номерами i и j заняты соответственно частицами с номерами i и j) благоприятствует $(n-2)!$ перестановок и т. д. Поскольку всего возможно $n!$ размещений частиц по ячейкам, то

$$P(A_i) = (n-1)!/n!, \quad P(A_i \cap A_j) = (n-2)!/n!, \dots$$

Так как сумма S_j содержит C_n^j одинаковых слагаемых $P(A_{i_1} \cap \dots \cap A_{i_j})$, то

$$S_j = ((n-j)!/n!) C_n^j = 1/j!, \quad j = 1, \dots, n.$$

Следовательно, вероятность $P_{n,1}$ хотя бы одного совпадения равна

$$P_{n,1} = 1 - 1/2! + 1/3! - \dots + (-1)^{n-1} 1/n!.$$

Любопытно, что $\lim_{n \rightarrow \infty} P_{n,1} = 1 - e^{-1} = 0,6321 \dots$ ($e = 2,718 \dots$), так как выражение для $P_{n,1}$ представляет сумму первых $n+1$ членов ряда для $1 - e^{-1} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots$.

Аналогично для вероятности $Q_{n,m}$ ровно m совпадений $\lim_{n \rightarrow \infty} Q_{n,m} = \frac{1}{m!} e^{-1}$, так что для больших n : $Q_{n,m} \approx \frac{1}{m!} e^{-1}$ и не зависит от n .

Заклячая изучение классической теоретико-вероятностной модели, приведем поучительный пример, известный как *парадокс шевалье де Мере*. Пусть одновременно бросаются три игральные кости. Какая комбинация более вероятна: дающая в сумме 11 очков или 12?

По мнению де Мере, эти комбинации равновероятны, поскольку 11 очков (событие A) можно получить шестью способами: $4 + 4 + 3$, $5 + 3 + 3$, $5 + 4 + 2$, $5 + 5 + 1$, $6 + 3 + 2$, $6 + 4 + 1$, и столькими же способами можно получить 12 очков (событие B): $4 + 4 + 4$, $5 + 4 + 3$, $5 + 5 + 2$, $6 + 3 + 3$, $6 + 4 + 2$, $6 + 5 + 1$. Де Мере полагал, что поскольку число способов, позволяющих получить событие A и B , одно и то же, то равны и вероятности $P(A)$ и $P(B)$. Однако в результате многочисленных наблюдений за игрой в кости шевалье отметил, что комбинация, дающая в сумме 12 очков, выпадает реже, чем дающая 11. Он обратился за разъяснениями к знаменитому Паскалю, который указал, что рассматриваемые де Мере «способы» не равновероятны, поскольку кроме выпадающих очков следует учитывать, на каких именно костях они выпали.

Действительно, занумеруем кости и будем выписывать значения выпадающих очков в соответствии с нумерацией костей. Тогда комбинация $6 + 4 + 1$ реализуется в шести случаях (641, 614, 461, 416, 164, 146), комбинация $5 + 3 + 3$ — в трех (533, 353, 335), а комбинация $4 + 4 + 4$ — лишь в одном (444). Поскольку в данном случае, очевидно, равновероятны все $6 \times 6 \times 6 = 216$ исходов $x y z$ ($x = 1, \dots, 6$, $y = 1, \dots, 6$, $z = 1, \dots, 6$), то понятно, что «способы» де Мере не равновероятны. На самом деле 11 очкам благоприятствуют 27 исходов, а 12 очкам — 25, так что $P(A) = 27/216$, $P(B) = 25/216$.

Классическая теоретико-вероятностная модель была построена в течение XVII — XIX вв. на пути естественной формализации некоторых из тех интуитивных представлений о вероятности, которые обсуждались во введении. Основателями математической теории вероятностей считаются Пьер Ферма (1601 — 1665) и Блез Паскаль (1623 — 1662). Размышляя о математических проблемах, возникающих в связи с азартными играми, в 1654 г. они установили некоторые из основных положений теории вероятностей. Ознакомившись с результатами Ферма и Паскаля, в разработке проблем теории вероятностей принимает участие Христиан Гюйгенс (1629 — 1695) и в 1657 г. издает первый трактат по теории вероятностей «О расчетах при азартных играх». В это время Гюйгенс уже полностью отдает себе отчет в том, что на самом деле речь идет не об играх, а о глубокой математической теории. Следующий крупный шаг был сделан Якобом Бернулли (1654 — 1705). Его посмертный труд «Искусство предположения» содержит много новых результатов. Наконец, «Учение о случае» Авраама де Муавра (1667 — 1754) и фундаментальный труд «Аналитическая теория вероятностей» Пьера Симона Лапласа (1749 — 1827) придают этой науке в известном смысле законченный вид. Однако после Лапласа интерес к теории вероятностей значительно упал, и в продолжение первых десятилетий

XIX в. ее даже перестали относить к математическим дисциплинам. Одна из главных причин этого в том, что теория вероятностей, построенная на неудовлетворительных основаниях, изобиловала парадоксами и противоречиями. В частности, лапласовское определение вероятности события A как отношения $n(A)/n$, где n — общее число равновероятных исходов, а $n(A)$ — число исходов, влекущих A , исходило из порочного круга понятий, поскольку использовало понятие равновероятности. Кроме того, оставался широкий круг случайных явлений, которые не удавалось понять в рамках классической модели. Это относится и к задачам на геометрическую вероятность, рассмотренным во введении.

Положение самостоятельной математической дисциплины теория вероятностей достигает лишь в трудах русского математика середины XIX в. Пафнутия Львовича Чебышева (1821 — 1894) и его учеников, А. А. Маркова (1856 — 1922) и А. М. Ляпунова (1857 — 1918). А в результате последующих фундаментальных исследований советских математиков А. Я. Хинчина, А. Н. Колмогорова, Е. Е. Слуцкого и С. Н. Бернштейна теория вероятностей, по существу, приобрела тот вид, какой она имеет на сегодняшний день. В частности, аксиоматика теории вероятностей, построенная А.Н. Колмогоровым, в настоящее время считается общепринятой.

1.3. Аксиоматическое построение теории вероятностей

Пусть Ω — пространство элементарных событий, $\mathcal{F}$ — алгебра событий (подмножеств Ω). Следующие пять условий образуют систему аксиом теории вероятностей.

1.3.1. Аксиомы теории вероятностей. 1. Алгебра $\mathcal{F}$ является σ -алгеброй событий.

Алгебра событий $\mathcal{F}$ называется σ -алгеброй, если для всякой последовательности событий $A_j \in \mathcal{F}$, $j = 1, 2, \dots$, их объединение $A = A_1 \cup A_2 \cup \dots = \bigcup_{j=1}^{\infty} A_j$ также принадлежит $\mathcal{F}$, т. е. является событием. Согласно принципу двойственности отсюда следует, что и $B = \bigcap_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} \bar{A}_j \in \mathcal{F}$.

Подчеркнем, что речь идет лишь о счетных объединениях и пересечениях. Если A_α , $\alpha \in S$, произвольная система событий, то, например, их объединение $\bigcup_{\alpha \in S} A_\alpha$ может и не быть событием.

2. На σ -алгебре $\mathcal{F}$ определяется функция $P(\cdot)$, принимающая значения $P(A) \in [0, 1]$, $A \in \mathcal{F}$, $P(\cdot) : \mathcal{F} \rightarrow [0, 1]$, называемая вероятностью и обладающая следующими свойствами.

3. Для всяких двух событий A и B , таких что $A \cap B = \emptyset$, $P(A + B) = P(A) + P(B)$ (аксиома сложения вероятностей).

Отсюда следует, что для произвольного конечного числа попарно несовместных событий $A_1, \dots, A_n$ $P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n)$.

4. Пусть события A_j , $j = 1, 2, \dots$, попарно несовместны: $A_i \cap A_j = \emptyset$, $i \neq j$, $i, j = 1, 2$ и $A = A_1 + A_2 + \dots$. Тогда

$$P(A) = P(A_1) + P(A_2) + \dots = \sum_{i=1}^{\infty} P(A_i) \quad (3.1)$$

(заметим, что согласно аксиоме 1 $A \in \mathcal{F}$).

Эта аксиома определяет *счетную аддитивность вероятности*. Может быть, более привычно ее называть *аксиомой непрерывности вероятности*. Для этого рассмотрим последовательность событий $B_1 = A_1$, $B_2 = A_1 + A_2, \dots$. Событие A , как будет показано в §1.3.3, можно понимать как предел последовательности $B_1 \subset B_2 \subset \dots$, $A = \lim_{n \rightarrow \infty} B_n$. При этом равенство (3.1) можно понимать как свойство непрерывности вероятности ¹⁾

$$P(A) = P(\lim_{n \rightarrow \infty} B_n) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{j=1}^n P(A_j) = \sum_{j=1}^{\infty} P(A_j).$$

5. $P(\Omega) = 1$.

Пространство элементарных событий Ω , σ -алгебра событий $\mathcal{F}$ и вероятность $P(\cdot) : \mathcal{F} \rightarrow [0, 1]$, удовлетворяющие аксиомам 1.-5., образуют так называемое *вероятностное пространство*, которое принято обозначать $(\Omega, \mathcal{F}, P)$.

Система аксиом теории вероятностей не противоречива, так как существуют Ω , $\mathcal{F}$ и P , удовлетворяющие этим аксиомам, и не полна, так как вероятность можно определить многими способами в рамках аксиом 1 ÷ 5. В качестве примера укажем классическую теоретико-вероятностную модель, в которой Ω — конечное множество, $\mathcal{F}$ — алгебра всех подмножеств Ω и вероятность P определена для каждого подмножества $A \in \mathcal{F}$ как отношение числа точек Ω , образующих A , к числу всех точек Ω .

Понятно, что для произвольного пространства элементарных событий Ω система всех его подмножеств образует σ -алгебру. Но такая σ -алгебра может оказаться столь обширной, что на ней невозможно определить вероятность, удовлетворяющую свойству счетной аддитивности 4. Требование счетной аддитивности P и стремление выбрать систему множеств $\mathcal{F}$ как можно более широкой взаимно ограничивают друг друга.

Приведем несколько примеров σ -алгебр. Если Ω — произвольное пространство элементарных событий, то $\mathcal{F} = \{\emptyset, \Omega\}$, а также множество $\mathcal{F}$ всех подмножеств Ω — тривиальные примеры σ -алгебр.

¹⁾ Определение непрерывности $P(\cdot) : \mathcal{F} \rightarrow [0, 1]$ см. в §1.3.3.

Многочисленные примеры σ -алгебр могут быть получены на следующем пути. Пусть $\mathcal{A}$ — произвольная система множеств и $\Omega = \bigcup_{A \in \mathcal{A}} A$. Если $\mathcal{F}$ — σ -алгебра всех подмножеств Ω , то, очевидно, $\mathcal{A} \subset \mathcal{F}$. Следовательно, существуют σ -алгебры подмножеств Ω , содержащие $\mathcal{A}$. Если $\mathcal{F}_1$ и $\mathcal{F}_2$ — σ -алгебры подмножеств Ω , содержащие $\mathcal{A}$, то $\mathcal{F}_1 \cap \mathcal{F}_2$ также содержит $\mathcal{A}$ и, кроме того, является σ -алгеброй. Следовательно, пересечение $\bigcap_{\alpha} \mathcal{F}_{\alpha} = \mathcal{F}_{\mathcal{A}}$ всех σ -алгебр, содержащих $\mathcal{A}$,

- 1) является σ -алгеброй,
- 2) является минимальной σ -алгеброй в том смысле, что для всякой σ -алгебры $\mathcal{F}$, содержащей $\mathcal{A}$, непременно $\mathcal{F}_{\mathcal{A}} \subset \mathcal{F}$ (ибо $\mathcal{F}$ совпадает с одной из σ -алгебр $\mathcal{F}_{\alpha}$, как σ -алгебра, содержащая $\mathcal{A}$),
- 3) содержит все множества из $\mathcal{A}$, $\mathcal{A} \subset \mathcal{F}$. В частности, если $\mathcal{A}$ — σ -алгебра, то $\mathcal{F}_{\mathcal{A}} = \mathcal{A}$, а если $\mathcal{A}$ состоит из двух множеств A и B , то $\mathcal{F}_{\mathcal{A}}$ состоит из восьми множеств: $\emptyset$, A , B , $A \cap B$, $A \cup B = \Omega$, $A \setminus B$, $B \setminus A$, $(A \setminus B) \cup (B \setminus A)$.

$\mathcal{F}_{\mathcal{A}}$ называется σ -алгеброй, порожденной системой множеств $\mathcal{A}$. При определении σ -алгебры событий порождающая ее система событий $\mathcal{A}$, как правило, составляется из событий, наблюдаемых в эксперименте. Так, например, в эксперименте с бросанием точки на отрезок $[a, b] = \Omega$ естественно считать событием любой отрезок $[\alpha, \beta]$, содержащийся в $[a, b]$. В таком случае $\mathcal{A}$ — система всех отрезков $[\alpha, \beta]$, содержащихся в $[a, b]$. σ -алгебра $\mathcal{F}_{\mathcal{A}}$ порожденная $\mathcal{A}$, в данном случае называется σ -алгеброй борелевских подмножеств отрезка $[a, b]$, или борелевской σ -алгеброй отрезка $[a, b]$.

Посмотрим, насколько обширна борелевская σ -алгебра $[a, b]$. Отдельные точки $[a, b]$, как одноточечные подмножества, являются борелевскими множествами, так как все они имеют вид $[\alpha, \alpha]$. Любое счетное подмножество $[a, b]$, например множество рациональных точек, является борелевским, равно как и его дополнение. Любой интервал $(\alpha, \beta) \subset [a, b]$ является борелевским множеством, поскольку $(\alpha, \beta) = \bigcup_{n=1}^{\infty} \left[\alpha + \frac{1}{n}, \beta - \frac{1}{n} \right]$. Для доказательства этого равенства заметим, что если $x \in (\alpha, \beta)$, то есть $\alpha < x < \beta$, то найдется n , для которого $\alpha + \frac{1}{n} \leq x \leq \beta - \frac{1}{n}$. Следовательно, $x \in \bigcup_{n=1}^{\infty} \left[\alpha + \frac{1}{n}, \beta - \frac{1}{n} \right]$. Наоборот, если выполнено последнее включение, то для некоторого n : $x \in \left[\alpha + \frac{1}{n}, \beta - \frac{1}{n} \right]$ и, следовательно, тем более $x \in (\alpha, \beta)$. Отсюда следует, что любое открытое подмножество $[a, b]$ — борелевское, поскольку всякое открытое множество, как известно, является не более чем счетным объединением интервалов. Все замкнутые подмножества $[a, b]$, как дополнения открытых, также борелевские. Наконец, борелевскими являются множества, полученные счетным объединением и (или) пересечением открытых и (или) замкнутых множеств, и т. д.

Класс борелевских множеств, как видно из приведенного перечисления, весьма обширен. Легко понять, что пример множества, не являющегося борелевским, не может быть простым. Однако такие множества есть, и, следовательно, не всякое подмножество $[a, b]$ является событием.

Для того чтобы завершить построение вероятностного пространства в эксперименте с бросанием точки на отрезок $[a, b]$, осталось определить вероятность на борелевской алгебре событий. Если для $[\alpha, \beta] \subset [a, b]$ положить $P([\alpha, \beta]) = (\beta - \alpha)/l$, $l = b - a$, и для любых содержащихся в $[a, b]$ непересекающихся отрезков $[\alpha, \beta]$ и $[\gamma, \delta]$ определить $P([\alpha, \beta] + [\gamma, \delta]) = ((\beta - \alpha) + (\delta - \gamma))/l$, то $P([\alpha, \alpha]) = 0$, $P([\alpha, \beta]) = P((\alpha, \beta]) = P((\alpha, \beta)) = (\beta - \alpha)/l$ и $P([a, b]) = 1$. Но это определение не содержит указаний на то, каким образом определить вероятность для произвольного борелевского множества. Этот вопрос полностью решается в теории меры. Можно показать, что на некоторой (лебеговской) σ -алгебре, содержащей борелевскую алгебру, можно определить вероятность, значение которой для отрезка $[\alpha, \beta] \subset [a, b]$ равно $(\beta - \alpha)/l$.

1.3.2. Дискретное вероятностное пространство. Существует важный для приложений класс так называемых *дискретных вероятностных пространств*, теория которых достаточно проста. Пример такого вероятностного пространства естественно возникает в стохастическом эксперименте с бросанием игральной кости. Рассмотрим «игру в кости до первого проигрыша». Условимся, что выпадение любого числа очков, кроме шести, означает выигрыш и продолжение игры, а выпадение шести очков означает проигрыш, и при этом игра заканчивается. Построим вероятностное пространство как математическую модель стохастического эксперимента «игра в кости до первого проигрыша». Начнем с того, что перечислим все исходы, приводящие к окончанию игры:

6;
16, 26, 36, 46, 56;
116, 126, 136, 146, 156, 226, ..., 256, 316, ..., 356, 416, ..., 456, 516, ..., 556;
1116, ...

Легко заметить, что имеется один исход, при котором игра заканчивается после первого бросания, 5 исходов (из 6^2), при которых игра заканчивается после второго бросания, и вообще 5^{n-1} исходов (из 6^n), при которых игра заканчивается после n -го бросания. В принципе игра может продолжаться неопределенно долго и, следовательно, не является стохастическим экспериментом с конечным числом исходов, если под исходом понимать выпадение некоторого числа очков при очередном бросании.

Введем пространство элементарных событий Ω , в котором точка ω_1 обозначает исход «6», точка ω_2 обозначает множество последовательностей двух исходов «16, 26, ..., 56», ω_3 соответственно «116, ..., 556» и т.

д. Согласно классической модели припишем событию $\{\omega_1\}$ вероятность $1/6$, событию $\{\omega_2\}$ — вероятность $5/(6^2)$, событию $\{\omega_3\}$ — вероятность $(5^2)/(6^3), \dots$, событию $\{\omega_n\}$ — вероятность $(5^{n-1})/(6^n)$. Обозначим ω_∞ множество всех бесконечных последовательностей (исходов), не содержащих шестерки, и припишем событию $\{\omega_\infty\}$ вероятность, равную $\lim_{n \rightarrow \infty} (5^{n-1})/(6^n) = 0$. Событие $\{\omega_\infty\}$ называется невероятным, но в отличие от $\emptyset$ не является невозможным.

Поскольку $\Omega = \{\omega_1\} + \{\omega_2\} + \dots + \{\omega_n\} + \dots$, то согласно аксиоме 4 о счетной аддитивности вероятности

$$P(\Omega) = \sum_{j=1}^{\infty} P(\{\omega_j\}) = \frac{1}{6} \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^n = 1,$$

причем значение $P(\Omega)$ не зависит от порядка суммирования (и порядка перечисления точек Ω), так как ряд справа сходится абсолютно.

Объявим событием любое подмножество Ω и для любого события $A \subset \Omega$ определим вероятность

$$P(A) = \sum_{j: \omega_j \in A} P(\{\omega_j\}). \quad (3.2)$$

Понятно, что это определение корректно, так как и в этом случае ряд справа сходится абсолютно, и значение $P(A)$ не зависит от порядка суммирования.

Подсчитаем, например, вероятность того, что игра закончится на четном бросании. Речь идет о вероятности события $A = \{\omega_2\} + \{\omega_4\} + \dots$, и согласно (3.2)

$$P(A) = \sum_{j=1}^{\infty} P(\{\omega_{2j}\}) = \frac{5}{36} \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^{2n} = \frac{5}{11}.$$

Оказывается, что вероятность (3.2) обладает свойством счетной аддитивности (и аддитивности) на σ -алгебре $\mathcal{F}$ всех подмножеств Ω , и, следовательно, построено вероятностное пространство $(\Omega, \mathcal{F}, P)$, где $\mathcal{F} = \mathcal{P}(\Omega)$ — класс всех подмножеств Ω , полностью характеризующее вероятностные свойства стохастического эксперимента «игра в кости до первого проигрыша».

Факт счетной аддитивности построенной вероятности является следствием *леммы о суммировании по блокам для абсолютно сходящихся рядов*.

Лемма 1.3.1 (о суммировании по блокам). Пусть $I = \{1, 2, \dots\}$ — множество всех чисел натурального ряда, I_α , $\alpha = 1, 2, \dots$, — непесекающиеся подмножества I , такие, что $I = I_1 + I_2 + \dots$ (этих подмножеств может быть и конечное число). Пусть числовой ряд $c_1 + c_2 + \dots + c_n + \dots$ сходится абсолютно и его сумма равна S , $S = \sum_{k=1}^{\infty} c_k$. Если $S_\alpha = \sum_{k \in I_\alpha} c_k$, $\alpha = 1, 2, \dots$, то числовой ряд

$S_1 + S_2 + \dots + S_n + \dots$ сходится абсолютно и его сумма равна S ,
 $S = \sum_{\alpha=1}^{\infty} S_{\alpha}$.

Доказательство. Понятно, что все S_{α} , $\alpha = 1, 2, \dots$, ограничены:

$$|S_{\alpha}| \leq \sum_{k \in I_{\alpha}} |c_k| \leq \sum_{k=1}^{\infty} |c_k| < \infty.$$

Покажем, что ряд $S_1 + S_2 + \dots$ сходится абсолютно и его сумма равна S . Абсолютная сходимость следует из оценки $\sum_{\alpha=1}^n |S_{\alpha}| \leq \sum_{\alpha=1}^n \sum_{k \in I_{\alpha}} |c_k| \leq \sum_{k=1}^{\infty} |c_k|$, верной для любого $n \geq 1$. Покажем теперь, что его сумма равна S . Зафиксируем любое $\varepsilon > 0$. Следует показать, что найдется такой номер $N_0 = N_0(\varepsilon)$, что при $n \geq N_0$

$$\left| \sum_{\alpha=1}^n S_{\alpha} - S \right| < \varepsilon. \quad (3.3)$$

Так как по условию ряд $\sum_{k=1}^{\infty} c_k$ сходится абсолютно, то для данного $\varepsilon > 0$ найдется такой номер $N_1 = N_1(\varepsilon)$, что при $n \geq N_1$

$$\left| \sum_{k=1}^n c_k - S \right| < \varepsilon/2, \quad \sum_{k=n+1}^{\infty} |c_k| < \varepsilon/2. \quad (3.4)$$

Выберем N_2 столь большим, чтобы в сумму $\sum_{\alpha=1}^{N_2} S_{\alpha} = \sum_{\alpha=1}^{N_2} \sum_{k \in I_{\alpha}} c_k$ вошли все члены c_k с номерами $k < N_1$ и положим $N_0 = \max(N_1, N_2)$. Тогда при $n \geq N_0$, используя (3.4), найдем

$$\begin{aligned} \left| \sum_{\alpha=1}^n S_{\alpha} - S \right| &\leq \left| \sum_{k=1}^n c_k - S \right| + \left| \sum_{\alpha=1}^n S_{\alpha} - \sum_{k=1}^n c_k \right| \leq \\ &\leq \left| \sum_{k=1}^n c_k - S \right| + \sum_{k=N_1+1}^{\infty} |c_k| < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Неравенство (3.3) доказано. ■

Замечание 1.3.1. Требование абсолютной сходимости ряда $c_1 + c_2 + \dots$ существенно. Например, ряд $1 - 1/2 + 1/3 - 1/4 + \dots$ сходится (условно), но $S_1 = 1 + 1/3 + 1/5 + \dots$ расходится.

Замечание 1.3.2. Обратная теорема неверна. Если для некоторой системы подмножеств $\{I_{\alpha}\}$ $\sum_{\alpha=1}^{\infty} |S_{\alpha}| < \infty$, то отсюда не следует абсолютная сходимость ряда $c_1 + c_2 + \dots$. Действительно, если $S_1 = 1 - 1/2$, $S_2 = 1/3 - 1/4$, ..., $S_{\alpha} = \frac{1}{2\alpha-1} - \frac{1}{2\alpha} = \frac{1}{2\alpha(2\alpha-1)}$, то $\sum_{\alpha=1}^{\infty} S_{\alpha} = \sum_{\alpha=1}^{\infty} |S_{\alpha}| < \infty$, но ряд $1 + 1/2 + 1/3 + \dots$ расходится.

Замечание 1.3.3. Если $c_k \geq 0$ (или $c_k \leq 0$), $k = 1, 2, \dots$, то теорема верна «в обе стороны». Действительно, пусть для некоторой системы подмножеств

$$I_\alpha, \alpha = 1, 2, \dots, I = I_1 + I_2 + \dots, \sum_{\alpha=1}^{\infty} S_\alpha = \sum_{\alpha=1}^{\infty} |S_\alpha| = S.$$

Тогда $\sum_{k=1}^n c_k \leq \sum_{\alpha=1}^{N(n)} S_\alpha \leq \sum_{\alpha=1}^{\infty} S_\alpha < \infty$, где $N(n)$ выбрано так, чтобы в сумму $S_1 + \dots + S_{N(n)}$ вошли все c_k , $k = 1, \dots, n$. Отсюда следует, что последовательность частичных сумм $\sum_{k=1}^n c_k$, $k = 1, 2, \dots$, монотонна и ограничена. Поэтому ряд $c_1 + c_2 + \dots$ сходится и его сумма совпадает с суммой ряда $S_1 + S_2 + \dots$ (в силу прямой теоремы).

Замечание 1.3.4. Если $\sum_{i,j} |x_i y_j| < \infty$, то равенства $\sum_{i,j} (x_i y_j) = \sum_i \sum_j (x_i y_j) = \sum_i x_i \sum_j y_j$ выполнены, если дополнительно руководствоваться «правилом Лебега», согласно которому $0 \cdot \infty = \infty \cdot 0 = 0$. Иллюстрацией является случай $x_i = 0$, $i = 1, 2, \dots$, $y_j = 1$, $j = 1, 2, \dots$, в котором $\sum_{i,j} |x_i y_j| = 0$.

Рассмотрим теперь общий случай дискретных вероятностных пространств.

Определение 1.3.1. Вероятностное пространство $(\Omega, \mathcal{F}, P)$ называется *дискретным*, если $\Omega = \{\omega_1, \omega_2, \dots\}$ конечно или счетно, $\mathcal{F}$ — σ -алгебра всех подмножеств Ω (включая пустое множество $\emptyset$), вероятность P определена для каждого одноточечного подмножества Ω :

$$P(\{\omega_j\}) = p_j \geq 0, \quad j = 1, 2, \dots, \quad \sum_{j=1}^{\infty} p_j = 1. \quad (3.5)$$

При этом вероятность любого события $A \in \mathcal{F}$ определяется равенством

$$P(A) = \sum_{j: \omega_j \in A} p_j. \quad (3.6)$$

Проверим, что для дискретного вероятностного пространства выполнены аксиомы теории вероятностей. Поскольку аксиомы 1, 2 и 5, очевидно, выполнены, достаточно проверить лишь аксиомы 3 и 4. Но обе эти аксиомы следуют из леммы о суммировании по блокам.

1.3.3. Свойства вероятности. Непрерывность. Остановимся на свойствах вероятности, которые вытекают из аксиом. Так же как при доказательстве свойств классической вероятности, найдем, что:

- 1) $P(\bar{A}) = 1 - P(A)$, так как $\bar{A} + A = \Omega$;
- 2) если $A \subset B$, то $P(B \setminus A) = P(B) - P(A)$, так как $B = A + B \setminus A$. Следовательно, включение $A \subset B$ влечет неравенство $P(A) \leq P(B)$ (монотонность вероятности);

3) для любых событий $A_1, \dots, A_n$ имеет место равенство (2.7), причем приведенное там доказательство верно и в общем случае. Сохраняют силу и выражения для вероятности $Q_{n,m}$ того, что осуществится ровно m событий из $A_1, \dots, A_n$, а также для вероятности $P_{n,m}$ того, что осуществится не менее m событий из $A_1, \dots, A_n$;

4) пусть $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$ — последовательность событий, каждое из которых влечет все последующие. Если $A = \bigcup_{j=1}^{\infty} A_j$ — событие, состоящее в том, что происходит хотя бы одно из событий A_j , $j = 1, 2, \dots$, то

$$P(A) = \lim_{n \rightarrow \infty} P(A_n). \quad (3.7)$$

Действительно, положим $A_0 = \emptyset$. Тогда $A = \bigcup_{j=1}^{\infty} A_j = (A_1 \setminus A_0) + (A_2 \setminus A_1) + \dots + (A_n \setminus A_{n-1}) + \dots$ и

$$\begin{aligned} P(A) &= \sum_{j=1}^{\infty} P(A_j \setminus A_{j-1}) = \lim_{n \rightarrow \infty} \sum_{j=1}^n P(A_j \setminus A_{j-1}) = \\ &= \lim_{n \rightarrow \infty} \sum_{j=1}^n (P(A_j) - P(A_{j-1})) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

5) Если $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$ — последовательность событий, каждое из которых влечет все предыдущие, и $A = \bigcap_{j=1}^{\infty} A_j$ событие, состоящее в том, что происходят все события $A_1, \dots, A_n, \dots$ то

$$P(A) = \lim_{n \rightarrow \infty} P(A_n). \quad (3.8)$$

Это утверждение может быть доказано с помощью принципа двойственности. Действительно, для противоположных событий: $\overline{A_1} \subset \overline{A_2} \subset \dots \subset \overline{A_n} \subset \dots$ и $\overline{A} = \bigcap_{j=1}^{\infty} \overline{A_j}$. Поэтому $P(\overline{A}) = \lim_{n \rightarrow \infty} P(\overline{A_n})$, и, следовательно,

$$P(A) = 1 - P(\overline{A}) = \lim_{n \rightarrow \infty} (1 - P(\overline{A_n})) = \lim_{n \rightarrow \infty} P(A_n).$$

Свойства 4 и 5 можно понимать как свойства непрерывности вероятности относительно монотонных предельных переходов. Действительно, если $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$, то $\bigcup_{j=1}^n A_j = A_n$, и множество

$A = \bigcup_{j=1}^{\infty} A_j$ естественно назвать пределом монотонной последовательности множеств $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$: $A = \lim_{j \rightarrow \infty} A_j$. Тогда согласно свойству 4: $P(A) = P(\lim_{j \rightarrow \infty} A_j) = \lim_{j \rightarrow \infty} P(A_j)$.

Точно так же, если $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$, то $A_n = \bigcap_{j=1}^n A_j$, и множество $A = \bigcap_{j=1}^{\infty} A_j$, называется пределом монотонной последовательности множеств $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots : A = \lim_{j \rightarrow \infty} A_j$. В данном случае свойство 5 означает, что $P(A) = P(\lim_{j \rightarrow \infty} A_j) = \lim_{j \rightarrow \infty} P(A_j)$.

Определим понятия сходимости и предела для произвольной последовательности множеств $A_1, A_2, \dots$. Рассмотрим множества $\tilde{A} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$ и $\tilde{A} = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k$, называемые верхним и соответственно нижним пределами последовательности $A_1, A_2, \dots$, которая называется сходящейся, если $\tilde{A} = \tilde{A}$, множество $A = \tilde{A} = \tilde{A}$ называется ее пределом и обозначается $\lim_{n \rightarrow \infty} A_n$. Заметим, что событие $\tilde{A}$ происходит, если и только если происходят все события монотонно убывающей последовательности

$$\bigcup_{k \geq 1} A_k \supset \bigcup_{k \geq 2} A_k \supset \dots \supset \bigcup_{k \geq n} A_k \supset \dots, \quad (3.9)$$

а это происходит тогда и только тогда, когда для любого $n = 1, 2, \dots$ происходит хотя бы одно A_k , $k \geq n$, т.е. $\tilde{A}$ происходит тогда и только тогда, когда происходит бесконечно много событий последовательности $A_1, A_2, \dots$. Событие $\tilde{A}$ происходит, если и только если происходит хотя бы одно из событий монотонно возрастающей последовательности

$$\bigcap_{k \geq 1} A_k \subset \bigcap_{k \geq 2} A_k \subset \dots \subset \bigcap_{k \geq n} A_k \subset \dots, \quad (3.10)$$

т.е. $\tilde{A}$ происходит тогда и только тогда, когда для некоторого $n = 1, 2, \dots$ происходят все события A_k при $k \geq n$, или, иначе говоря, происходят все события последовательности $A_1, A_2, \dots$, за исключением, быть может, конечного их числа. Нетрудно заметить, что

$$\begin{aligned} P(\tilde{A}) &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right) \geq \overline{\lim}_{n \rightarrow \infty} P(A_n) \geq \\ &\geq \underline{\lim}_{n \rightarrow \infty} P(A_n) \geq \lim_{n \rightarrow \infty} P\left(\bigcap_{k \geq n} A_k\right) = P(\tilde{A}). \end{aligned} \quad (3.11)$$

Первое и последнее равенства в (3.11) следуют из (3.8), (3.9) и соответственно из (3.7), (3.10), первое и третье неравенства в (3.11), следуют из включений $\bigcup_{k \geq n} A_k \supset A_n$ и $A_n \supset \bigcap_{k \geq n} A_k$ соответственно, наконец, второе неравенство в (3.11) следует из определений верхнего $\overline{\lim}$ и нижнего $\underline{\lim}$ пределов. Если последовательность $A_1, A_2, \dots$ сходится,

то $\tilde{A} = A$ и согласно (3.11) $\overline{\lim}_{n \rightarrow \infty} P(A_n) = \underline{\lim}_{n \rightarrow \infty} P(A_n)$. Итак, если $A = \lim_{n \rightarrow \infty} A_n$, то $P(A) = \lim_{n \rightarrow \infty} P(A_n)$, т.е. счетно-аддитивная вероятность непрерывна относительно сходимости последовательности множеств $A_1, A_2, \dots$. Читателю полезно проверить, что для монотонных последовательностей $A_1 \subset A_2 \subset \dots$ и $A_1 \supset A_2 \supset \dots$ $\tilde{A} = A = \bigcup_{n=1}^{\infty} A_n$ и соответственно $\tilde{A} = A = \bigcap_{n=1}^{\infty} A_n$.

1.4. Условная вероятность. Независимость

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, P)$. Рассмотрим задачу определения вероятности события A , если известно, что произошло событие B , причем $P(B) > 0$. При этих условиях пространством элементарных событий естественно считать не Ω , а B , поскольку тот факт, что B произошло, означает, что речь идет лишь о тех $\omega \in \Omega$, которые принадлежат множеству B . В общем случае событие $A \cap B$ влечет A , но, если известно, что событие B произошло, то при этом условии A влекут те и только те элементарные события, которые принадлежат $A \cap B$. Поскольку ранее мы условились отождествлять событие A и множество элементарных событий, влекущих A , то теперь событие A следует отождествить с множеством $A_B = A \cap B$. Можно сказать, что множество $A_B = A \cap B$ есть событие A , рассматриваемое с точки зрения, согласно которой пространством элементарных событий объявлено событие $\Omega_B = B$.

На новом пространстве элементарных событий Ω_B σ -алгебра событий $\mathcal{F}_B$ определяется, или, как говорят, индуцируется, σ -алгеброй событий $\mathcal{F}$. Именно $\mathcal{F}_B$ состоит из событий вида $A_B = A \cap B$, где $A \in \mathcal{F}$. Проверим, что $\mathcal{F}_B$ действительно σ -алгебра. Пусть $A_B, C_B, C_B^j \in \mathcal{F}_B$ ($A, C, C^j \in \mathcal{F}$), $j = 1, 2, \dots$. Тогда, используя свойства (1.5) взаимной дистрибутивности операций объединения $\cup$ и пересечения $\cap$, найдем

$$\begin{aligned} A_B \cup C_B &= (A \cap B) \cup (C \cap B) = (A \cup C) \cap B = (A \cup C)_B, \\ \bigcup_{j=1}^{\infty} C_B^j &= \bigcup_{j=1}^{\infty} (C^j \cap B) = \left(\bigcup_{j=1}^{\infty} C^j \right) \cap B = \left(\bigcup_{j=1}^{\infty} C^j \right)_B, \\ A_B \cap C_B &= (A \cap C)_B. \end{aligned}$$

Следовательно, $A_B \cup C_B, \bigcup_{j=1}^{\infty} C_B^j, A_B \cap C_B \in \mathcal{F}_B$. Наконец, по определению $\bar{A}_B = B \setminus A_B = B \setminus A = B \cap \bar{A} = (\bar{A})_B$, поэтому $\bar{A}_B \in \mathcal{F}_B$, если $A_B \in \mathcal{F}_B$.

Определим на σ -алгебре $\mathcal{F}_B$ вероятность $P_B(\cdot)$:

$$P_B(A_B) = P(A_B)/P(\Omega_B) = P(A \cap B)/P(B), \quad A_B \in \mathcal{F}_B.$$

Без труда можно проверить, что $P_B(\cdot)$, так же как и $P(\cdot)$, удовлетворяет аксиомам 2.—5. теории вероятностей.

Наглядный смысл вероятности $P_B(\cdot)$ можно пояснить на примере классической вероятности. В этом случае $\Omega = \{\omega_1, \dots, \omega_n\}$, причем события $\{\omega_1\}, \dots, \{\omega_n\}$ равновероятны. Пусть $\Omega_B = B = \{\omega_{j_1}, \dots, \omega_{j_k}\}$, так что $P(B) = k/n$, и $A_B = A \cap B = \{\omega_{i_1}, \dots, \omega_{i_s}\} \subset \{\omega_{j_1}, \dots, \omega_{j_k}\}$, так что $P(A_B) = P(A \cap B) = s/n$. Если Ω_B рассматривать как новое пространство элементарных событий, то вероятность события A_B определяется как отношение числа s исходов, благоприятствующих A_B , к общему числу исходов k . Теперь общее число исходов — число исходов, приводящих к событию B . Таким образом, в согласии с классической схемой

$$P_B(A_B) = \frac{s}{k} = \frac{s/n}{k/n} = \frac{P(A \cap B)}{P(B)}.$$

Тройка $(B, \mathcal{F}_B, P_B)$ является новым вероятностным пространством, построенным в связи с поставленной задачей. Но вероятность $P_B(\cdot)$ можно рассматривать и на σ -алгебре $\mathcal{F}$. На $\mathcal{F}$ $P_B(\cdot)$ также является вероятностью и обозначается $P(\cdot|B)$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{F}, \quad (4.1)$$

$P(\cdot|B)$ как функция на $\mathcal{F}$ называется условной при условии, что событие B произошло, вероятностью.

Что касается свойств вероятности $P(\cdot|B)$, то они вполне аналогичны соответствующим свойствам вероятности $P(\cdot)$: $P(\Omega|B) = 1$, $P(\bar{A}|B) = 1 - P(A|B)$, $P(A_1 + A_2|B) = P(A_1|B) + P(A_2|B)$, так как $(A_1 + A_2) \cap B = A_1 \cap B + A_2 \cap B$. Если воспользоваться равенством $P(A' \cup C') = P(A') + P(C') - P(A' \cap C')$, то, положив $A' = A \cap B$, $C' = C \cap B$, нетрудно проверить, что

$$\begin{aligned} P(A \cup C|B) &= P((A \cup C) \cap B)/P(B) = \\ &= P((A \cap B) \cup (C \cap B))/P(B) = P(A|B) + P(C|B) - P(A \cap C|B). \end{aligned}$$

Наконец, очевидно, $P(A_1 + A_2 + \dots|B) = \sum_{j=1}^{\infty} P(A_j|B)$ (счетная аддитивность условной вероятности).

Для иллюстрации техники вычисления условных вероятностей рассмотрим *пример*. Известно, что вероятность аварии при запуске ракеты равна 0,1, в том числе вероятность аварии на старте — 0,09. Какова вероятность аварии в случае успешного старта? Обозначим A событие «авария при запуске», B — событие «авария на старте». Требуется вычислить вероятность $P(A|\bar{B})$. По смыслу ясно, что B влечет A , $B \subset A$. Поэтому $\bar{B} \supset \bar{A}$. Следовательно, $\bar{A} \cap \bar{B} = \bar{A}$. Это равенство позволяет подсчитать вероятность $P(\bar{A}|\bar{B}) = P(\bar{A} \cap \bar{B})/P(\bar{B}) = P(\bar{A})/P(\bar{B})$. Поскольку $P(A|\bar{B}) + P(\bar{A}|\bar{B}) = 1$, то искомая условная вероятность равна

$P(A|\bar{B}) = 1 - \frac{1 - 0,1}{0,1 - 0,09} = \frac{1}{91}$. На рис. 1.7 приведено «графическое» решение задачи.

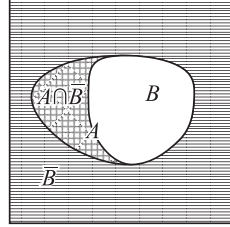


Рис. 1.7. $P(A) = 0,1$, $P(B) = 0,09$, $P(A|\bar{B}) = P(A \cap \bar{B})/P(\bar{B}) = (0,1 - 0,09)/(1,0 - 0,09)$.

Из определения (4.1) следует так называемая *теорема умножения вероятностей*

$$P(A \cap B) = P(A|B)P(B), P(B) > 0. \quad (4.2)$$

Вполне аналогично можно записать

$$P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C),$$

где $P(B \cap C) > 0$ и, следовательно, $P(C) > 0$.

С точки зрения математической этот результат тривиален. Но его роль на самом деле не математическая. Эта теорема применяется при определении вероятностей в тех случаях, когда по смыслу задачи легко определяются условные вероятности. Поясним это следующим *примером*. Имеются 30 экзаменационных билетов, среди которых 5 «счастливых». Кому выгоднее тянуть билет — первому или второму? Обозначим A событие «первый вытягивает «счастливый» билет», B — «второй вытягивает «счастливый» билет». Тогда $\bar{A}$ означает, что первый не вытянет «счастливый» билет, а $\bar{B}$ — что второй. Заметим, что $A + \bar{A} = \Omega$, поэтому $B = B \cap \Omega = B \cap A + B \cap \bar{A}$. Следовательно, $P(B) = P(B \cap A) + P(B \cap \bar{A}) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$. Что касается вероятностей, которые входят в эту формулу, то

$$P(A) = \frac{5}{30} = \frac{1}{6}, \quad P(\bar{A}) = \frac{25}{30}, \quad P(B|A) = \frac{4}{29}, \quad P(B|\bar{A}) = \frac{5}{29}$$

и, следовательно, $P(B) = \frac{1}{6} \cdot \frac{4}{29} + \frac{5}{6} \cdot \frac{5}{29}$. Таким образом, $P(B) = 1/6 = P(A)$. Читателю рекомендуется выяснить, может быть третьему больше повезет?

Одним из важнейших понятий теории вероятностей является понятие (*стохастической*) независимости.

События A и B называются *независимыми*, если

$$P(A \cap B) = P(A)P(B). \quad (4.3)$$

Если $P(B) > 0$, то согласно (4.2) и (4.3) $P(A|B) = P(A)$, как и должно быть по смыслу условной вероятности. Аналогично, если $P(A) > 0$, то $P(B|A) = P(B)$. Однако определение независимости A и B на основе равенств $P(A|B) = P(A)$, $P(B|A) = P(B)$ не эквивалентно (4.3), так как в (4.3) не предполагается *существования* условных вероятностей.

Из определения независимости непосредственно следует, что

1. Ω и любое событие A независимы.
2. Любое событие A и событие B независимы, если $P(B) = 0$. В самом деле, поскольку $A \cap B \subset B$, то $0 \leq P(A \cap B) \leq P(B) = 0$. Следовательно, $P(A \cap B) = 0 = P(A)P(B)$.
3. Если A и B независимы, то независимы также следующие пары событий: $\bar{A}$ и B , A и $\bar{B}$, $\bar{A}$ и $\bar{B}$.

Докажем, например, независимость $\bar{A}$ и B . Так как $A + \bar{A} = \Omega$, то $A \cap B + \bar{A} \cap B = B$ и, следовательно, $P(B) = P(A)P(B) + P(\bar{A} \cap B)$. Поэтому $P(\bar{A} \cap B) = P(B)(1 - P(A)) = P(B)P(\bar{A})$. Остальные утверждения доказываются аналогично.

В определении (4.3) *независимости событий* A и B , существенную роль играет вероятность P , изменив которую можно нарушить равенство (4.3), а вместе с ним и независимость A и B . В связи с этим замечанием, которое призвано подчеркнуть, что стохастическая независимость — нечто более определенное, чем независимость вообще, рассмотрим модель двух стохастических экспериментов (испытаний) и выясним, при каких условиях они стохастически независимы. Моделью i -го эксперимента $\mathcal{E}_i$ является вероятностное пространство $(\Omega_i, \mathcal{F}_i, P_i)$, в котором $\Omega_i = \{\omega_1^{(i)}, \omega_2^{(i)}, \dots\}$, $\mathcal{F}_i$ — σ -алгебра всех подмножеств Ω_i , P_i — вероятность, определенная на $\mathcal{F}_i$ своими значениями на элементарных событиях, $P_i(\{\omega_j^{(i)}\})$, $j = 1, 2, \dots$, моделирующих элементарные исходы i -го эксперимента, $i = 1, 2$.

Определим новое вероятностное пространство $(\Omega, \mathcal{F}, P)$, в котором Ω состоит из всевозможных упорядоченных пар $\omega_{ij} = (\omega_i^{(1)}, \omega_j^{(2)})$, $i, j = 1, 2, \dots$:

$$\Omega = \{\omega_{ij} \stackrel{\text{def}}{=} (\omega_i^{(1)}, \omega_j^{(2)}), \omega_i^{(1)} \in \Omega_1, \omega_j^{(2)} \in \Omega_2, i, j = 1, 2, \dots\}.$$

Так определенное Ω называется *произведением* Ω_1 и Ω_2 и обозначается $\Omega_1 \times \Omega_2$. Событиями назовем все подмножества Ω и соответствующую σ -алгебру обозначим $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. Наконец, вероятность P на $\mathcal{F}$ определим, задав ее значения на элементарных событиях равенствами $P(\{\omega_{ij}\}) = P_1(\{\omega_i^{(1)}\})P_2(\{\omega_j^{(2)}\})$, $i, j = 1, 2, \dots$, так что для всякого события $A \in \mathcal{F}$

$$P(A) = \sum_{i,j: \omega_{ij} \in A} P(\{\omega_{ij}\}) = \sum_{i,j: (\omega_i^{(1)}, \omega_j^{(2)}) \in A} P_1(\{\omega_i^{(1)}\})P_2(\{\omega_j^{(2)}\}). \quad (4.4)$$

Построенное вероятностное пространство называется *произведением пространств* $(\Omega_1, \mathcal{F}_1, P_1)$ и $(\Omega_2, \mathcal{F}_2, P_2)$ и обозначается $(\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2)$.

Рассмотрим событие $A \in \mathcal{F}$, которое состоит из тех пар $(\omega_i^{(1)}, \omega_j^{(2)})$, в которых $\omega_i^{(1)} \in A \in \mathcal{F}_1$, $\omega_j^{(2)}$ произвольно, $\omega_j^{(2)} \in \Omega_2$. Такое событие назовем *цилиндрическим* и обозначим $A = A_1 \times \Omega_2$. Согласно определению (4.4) и теореме о суммировании по блокам, и см. рис. 1.8,

$$\begin{aligned} P(A) &= \sum_{i,j: \omega_i^{(1)} \in A_1, \omega_j^{(2)} \in \Omega_2} P_1(\{\omega_i^{(1)}\})P_2(\{\omega_j^{(2)}\}) = \\ &= \sum_{i: \omega_i^{(1)} \in A_1} P_1(\{\omega_i^{(1)}\}) \sum_{j: \omega_j^{(2)} \in \Omega_2} P_2(\{\omega_j^{(2)}\}) = P_1(A_1). \end{aligned}$$

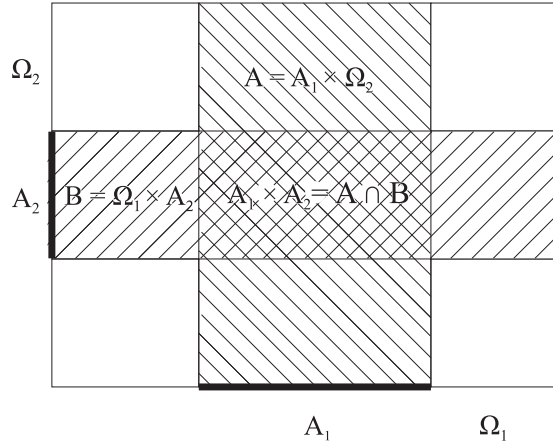


Рис. 1.8. События $A_1 \subset \Omega_1$, $A_2 \subset \Omega_2$, $A = A_1 \times \Omega_2 \subset \Omega_1 \times \Omega_2$, $B = \Omega_1 \times A_2 \subset \Omega_1 \times \Omega_2$ и $A_1 \times A_2 = A \cap B \subset \Omega_1 \times \Omega_2$.

Аналогично для $B = \Omega_1 \times A_2$ найдем $P(B) = P_2(A_2)$. Событие $A \cap B$ состоит из тех пар $(\omega_i^{(1)}, \omega_j^{(2)})$, для которых $\omega_i^{(1)} \in A_1$, $\omega_j^{(2)} \in A_2$. Поэтому

$$\begin{aligned} P(A \cap B) &= \sum_{i,j: \omega_i^{(1)} \in A_1, \omega_j^{(2)} \in A_2} P_1(\{\omega_i^{(1)}\})P_2(\{\omega_j^{(2)}\}) = \\ &= \sum_{i: \omega_i^{(1)} \in A_1} P_1(\{\omega_i^{(1)}\}) \sum_{j: \omega_j^{(2)} \in A_2} P_2(\{\omega_j^{(2)}\}) = P_1(A_1)P_2(A_2). \end{aligned}$$

Следовательно, $P(A \cap B) = P(A)P(B)$, т. е. события A и B независимы при любых вероятностях P_1 и P_2 .

Вероятностное пространство $(\Omega_1, \mathcal{F}_1, P_1)$ моделирует первый стохастический эксперимент, $(\Omega_2, \mathcal{F}_2, P_2)$ — второй, никак не связанный с первым. Вероятностное пространство $(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times$

$\times (\Omega_2, \mathcal{F}_2, P_2)$ моделирует два стохастически независимых стохастических эксперимента, выполняемых либо один за другим, либо — одновременно. Цилиндрические подмножества Ω определяют события в одном из экспериментов при произвольном исходе другого.

Пусть, *например*, бросаются две кости. Вероятность каждого из 36 возможных исходов определим равной $1/36$. Тогда события «на первой кости выпало четное число очков, на второй — что угодно» и «на первой кости выпало что угодно, на второй — три очка» независимы. Этот результат, конечно, является следствием того, что вероятности совместных элементарных исходов определены как произведения их вероятностей.

Однако в некоторых ситуациях, анализируя условия стохастического эксперимента, бывает трудно заключить, независимы те или иные события или нет. В таких случаях следует действовать согласно определению независимости. Пусть, *например*, из колоды, содержащей 52 карты, вытягиваются карты. Рассмотрим события: A_1 — дама, A_2 — карта пиковой масти. Тогда

$$P(A_1) = 4/52 = 1/13, \quad P(A_2) = 13/52 = 1/4, \\ P(A_1)P(A_2) = 1/52.$$

С другой стороны, вероятность вытянуть даму пиковой масти равна $P(A_1 \cap A_2) = 1/52$. Следовательно, события A_1 и A_2 независимы.

Определение 1.4.1. События $A_1, \dots, A_n$ называются *взаимно независимыми* или *независимыми в совокупности*, если при любом выборе различных событий $A_{i_1}, \dots, A_{i_k}$ из данной совокупности выполняется равенство

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}). \quad (4.5)$$

Разумеется, из независимости событий в совокупности следует их попарная независимость. Но обратное утверждение неверно. Проиллюстрируем этот факт *на примере*, принадлежащем С.Н. Бернштейну. Пусть три грани правильного тетраэдра окрашены соответственно в красный (К), зеленый (З) и синий (С) цвета, а четвертая — в три цвета (КЗС). Ясно, что вероятность упасть тетраэдру гранью, на которой есть, скажем, красный цвет, равна $P(K) = 1/2$. Условная вероятность оказаться на этой грани красному цвету при условии, что на ней уже есть зеленый, равна $P(K|З) = 1/2$. Аналогично $P(K) = P(З) = P(С) = P(K|З) = \dots = P(С|З) = 1/2$. Следовательно, события К, З и С попарно независимы. Однако вероятность упасть гранью, на которой есть все три цвета, равна $P(K \cap З \cap С) = 1/4 \neq 1/8$. Отсюда следует, что события К, З и С не являются независимыми в совокупности.

1.4.1. Формулы полной вероятности и Байеса. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, $A_1, \dots, A_n$ — полная группа

попарно несовместных событий, т. е. $A_i \cap A_j = \emptyset$, $i \neq j$, $i, j = 1, \dots, n$, $A_1 + \dots + A_n = \Omega$.

Если $B \in \mathcal{F}$ — произвольное событие, то из последнего равенства следует, что $B = B \cap \Omega = B \cap A_1 + \dots + B \cap A_n$. Поэтому $P(B) = \sum_{j=1}^n P(B \cap A_j)$, или в терминах условных вероятностей (в силу равенства (4.2))

$$P(B) = \sum_{j=1}^n P(B|A_j)P(A_j), \quad (4.6)$$

если $P(A_j) > 0$, $j = 1, \dots, n$. Равенство (4.6) называется *формулой полной вероятности*. Если полная группа попарно несовместных событий состоит из счетного множества событий $A_1, \dots, A_n, \dots$, то $B = B \cap A_1 + \dots + B \cap A_n + \dots$, и в силу счетной аддитивности вероятности отсюда следует *формула полной вероятности для счетного множества событий*

$$P(B) = \sum_{j=1}^{\infty} P(B|A_j)P(A_j), \quad (4.7)$$

если $P(A_j) > 0$, $j = 1, 2, \dots$.

Если $P(B) > 0$, то согласно определению условной вероятности

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)}, \quad k = 1, 2, \dots, \quad (4.8)$$

или, воспользовавшись формулой (4.6) или (4.7)

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_j P(B|A_j)P(A_j)}, \quad k = 1, 2, \dots \quad (4.9)$$

Формулы (4.8) и (4.9) называются *формулами Байеса* и обычно используются следующим образом. Пусть известны вероятности $P(A_j)$ попарно несовместных событий A_j , $j = 1, 2, \dots$, образующих полную группу, и условные вероятности $P(B|A_j)$, $j = 1, 2, \dots$, события B . Вероятности $P(A_j)$, $j = 1, 2, \dots$, носят название *априорных*, или *доопытных*. Если в результате опыта происходит событие B , то априорные вероятности по формуле Байеса могут быть пересчитаны в *апостериорные*, или *послеопытные*, вероятности $P(A_j|B)$, $j = 1, 2, \dots$.

Байесовский пересчет вероятностей наглядно сводится к заменам:

$$\begin{aligned} \Omega &\rightarrow B, \quad A_k \rightarrow A_k \cap B, \\ P(A_k) &\rightarrow P(A_k \cap B)/P(B) = P(A_k|B), \quad k = 1, 2, \dots \end{aligned}$$

В качестве иллюстрации рассмотрим *задачу*, известную как *задача о разорении игрока*. Некто при выпадении единицы получает одну копейку, при выпадении минуса единицы платит одну копейку (получает минус одну копейку). По условию игра заканчивается при выполнении

хотя бы одного из двух событий: либо игрок набирает капитал a копеек, либо разоряется, т. е. набирает 0 копеек. Пусть x — начальный капитал, $x \leq a$. Обозначим $P(x)$ вероятность разорения игрока. Будем считать, что «1» и «-1» выпадают с вероятностью $1/2$. Определим события: A — разорение игрока, A_1 — выпадение «1», A_2 — выпадение «-1». Тогда $A = A \cap A_1 + A \cap A_2$, и по формуле полной вероятности

$$\begin{aligned} P(x) &= P(x|A_1)P(A_1) + P(x|A_2)P(A_2) = \\ &= [P(x+1) + P(x-1)] \cdot 1/2. \end{aligned} \quad (4.10)$$

Общее решение этого разностного уравнения имеет вид $P(x) = cx + d$, что совместно с граничными условиями $P(0) = 1$, $P(a) = 0$, дает $P(x) = 1 - x/a$, $0 \leq x \leq a$.

Если вероятности выпадения «1» и «-1» различны и равны соответственно p и $q = 1 - p$, то вместо уравнения (4.10), определяющего вероятность $P(x)$, $x \in [0, a]$, разорения игрока при условиях $P(0) = 1$, $P(a) = 0$, по формуле полной вероятности найдем

$$P(x) = pP(x+1) + qP(x-1), \quad x \in (0, a), \quad P(0) = 1, \quad P(a) = 0. \quad (4.11)$$

Общее решение уравнения (4.11) при $p \neq q$ определяется как линейная комбинация $\alpha z_1^x + \beta z_2^x$ его частных решений z_1^x и z_2^x , $x \in [0, a]$, где z_1 и z_2 суть корни q/p и 1 квадратного уравнения $pz + q/z = z$, а α и β определяются краевыми условиями в (4.11). В этом более общем случае вероятность разорения определяется равенством

$$P(x) = \frac{(q/p)^x - (q/p)^a}{1 - (q/p)^a}, \quad x \in [0, a].$$

Читателю рекомендуется рассмотреть задачу о разорении при разных ставках.

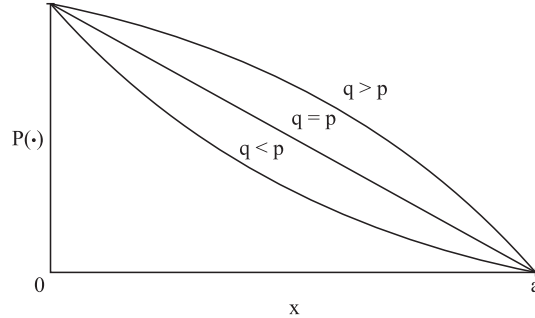


Рис. 1.9. Вероятность разорения как функция начального капитала.

1.5. Последовательность независимых испытаний

Рассмотрим дискретное вероятностное пространство $(\Omega, \mathcal{F}, P)$, в котором $\Omega = \{\omega_1, \omega_2, \dots\}$, $\mathcal{F}$ состоит из всех подмножеств Ω и веро-

ятность P определена для каждого элементарного события $\{\omega_j\} \subset \Omega$ равенством $P(\{\omega_j\}) = p_j, j = 1, 2, \dots, \sum_{j=1}^{\infty} p_j = 1$. Будем считать, что вероятностное пространство $(\Omega, \mathcal{F}, P)$ является моделью испытания T , а $\{\omega_1\}, \{\omega_2\}, \dots$ — элементарные исходы T . Тогда моделью испытания T , повторенного дважды, должно быть вероятностное пространство $(\Omega_2, \mathcal{F}_2, P_2) = (\Omega, \mathcal{F}, P) \times (\Omega, \mathcal{F}, P)$. Элементарными событиями теперь будут упорядоченные пары $\{(\omega_i, \omega_j)\} \subset \Omega^2 = \Omega_2, i, j = 1, 2, \dots, \mathcal{F}_2 = \mathcal{F} \times \mathcal{F}$ — σ -алгеброй всех подмножеств Ω_2 . Вероятность P_2 на $\mathcal{F}_2$ можно определить многими способами, которые, в свою очередь, определяются условиями проведения двух испытаний. Однако, если исходы первого испытания никак не влияют на исходы второго, т.е. испытания выполнены независимо, то согласно результатам предыдущего параграфа следует положить $P_2(\{(\omega_i, \omega_j)\}) = p_i p_j, i, j = 1, 2, \dots$. Там было показано, что так определенная вероятность на $\mathcal{F}_2$ отвечает условию независимости испытаний. События, связанные с каждым испытанием, являются цилиндрическими множествами Ω_2 .

Очевидно, что моделью n взаимно независимых испытаний T является вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P)^n$, в котором $\Omega_n = \Omega \times \dots \times \Omega = \Omega^n, \mathcal{F}_n = \mathcal{F} \times \dots \times \mathcal{F} = \mathcal{F}^n$, а вероятность P_n задана равенствами

$$P_n(\{(\omega_i \omega_j \dots \omega_k)\}) = p_i p_j \dots p_k, i, j, \dots, k = 1, 2, \dots \quad (5.1)$$

Определение 1.5.1. Пусть $(\Omega, \mathcal{F}, P)$ — дискретное вероятностное пространство, $\Omega = \{\omega_1, \omega_2, \dots\}, P(\{\omega_j\}) = p_j, j = 1, 2, \dots, \sum_{j=1}^{\infty} p_j = 1$.

*Последовательностью n взаимно независимых испытаний*¹⁾ называется вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n)$, элементарными событиями в котором являются последовательности $\{(\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n})\}$ и вероятность определена для каждого элементарного события равенством (5.1). Последовательность взаимно независимых испытаний называется последовательностью испытаний Бернулли, или *схемой Бернулли*, если $\Omega = \{\omega_1, \omega_2\}$, т. е. если испытание T имеет лишь два исхода.

Обычно в схеме Бернулли исход $\{\omega_1\}$ называют успехом и обозначают соответствующую вероятность $P(\{\omega_1\}) = p, \{\omega_2\}$ называют неудачей, $P(\{\omega_2\}) = 1 - p = q$. В серии из n независимых испытаний в схеме Бернулли Ω_n состоит из 2^n элементарных событий, причем $P_n(\{(\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n})\}) = p^k q^{n-k}$ где k — число успехов в последовательности $\{\omega_{i_1}, \dots, \omega_{i_n}\}$.

В качестве примеров схемы Бернулли можно привести опыт с бросанием монеты или игральной кости. В последнем случае можно

¹⁾ или — моделью последовательности n взаимно независимых испытаний.

считать, что $\{\omega_1\}$ — выпадение одного очка, $\{\omega_2\}$ — невыпадение одного очка.

В связи со схемой Бернулли найдем вероятность того, что в серии из n взаимно независимых испытаний успех наступит $k \leq n$ раз. Поскольку при этом не имеет значения, когда именно в этих испытаниях будут наблюдаться k успехов, то событие A_k , состоящее в наступлении k успехов, является объединением всех различных элементарных событий $\{(\omega_{i_1}\omega_{i_2}\dots\omega_{i_n})\}$, в которых ω_1 встречается k раз. Всего таких элементарных событий в A_k C_n^k , а поскольку они несовместны, искомая вероятность дается равенством

$$p_n(A_k) = P_n(k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \dots, n. \quad (5.2)$$

В частности, вероятность не иметь ни одного успеха равна q^n , и, следовательно, вероятность иметь хотя бы один успех равна $1 - q^n$. Совокупность вероятностей $P_n(k)$, $k = 0, 1, \dots, n$, называется *биномиальным распределением*. Такое название связано с тем, что вероятность (5.2) является общим членом разложения бинома

$$1 = (p + q)^n = \sum_{j=0}^n C_n^j p^j q^{n-j} = \sum_{k=0}^n P_n(k).$$

Это равенство в данном случае является отражением того факта, что в серии из n испытаний Бернулли исходы, содержащие $0, 1, \dots, n$ успехов образуют полную группу попарно несовместных событий.

Рассмотрим отношение

$$\frac{P_n(k+1)}{P_n(k)} = \frac{n! p^{k+1} q^{n-k-1}}{(k+1)!(n-k-1)!} \frac{k!(n-k)!}{n! p^k q^{n-k}} = \frac{p(n-k)}{q(k+1)}.$$

Так как неравенство $P_n(k+1)/P_n(k) > 1$ эквивалентно неравенству $np - q > k$, то вероятность $P_n(k)$ возрастает при переходе от k к $k+1$, если $np - q > k$, и, наоборот, убывает при переходе от k к $k+1$, если $np - q < k$. Если же $np - q = k$, то $P_n(k) = P_n(k+1)$. Заметим, что при достаточно малых вероятностях успеха p , таких, что $np - q \leq 0$, вероятность $P_n(k)$ не возрастает ни при каких $k \neq 0$. В этом случае вероятнее всего, что при n испытаниях не будет ни одного успеха.

Возвращаясь к общей схеме последовательности взаимно независимых испытаний, рассмотрим случай конечного $\Omega = \{\omega_1, \dots, \omega_r\}$ и положим $P(\{\omega_j\}) = p_j$, $j = 1, \dots, r$. В каждом испытании теперь возможны r различных исходов, и при n испытаниях вероятность исхода $\omega_{i_1}, \dots, \omega_{i_n}$ равна $p_1^{s_1} \dots p_r^{s_r}$, где $s_1, s_2, \dots, s_r$ — количества исходов $\omega_1, \omega_2, \dots, \omega_r$ соответственно в последовательности $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n}$, $s_1 + s_2 + \dots + s_r = n$. Нетрудно заметить, что при n испытаниях вероятность того, что ω_1 наблюдается s_1 раз, ω_2 — s_2 раз, ..., ω_r — s_r раз,

равна

$$P_n(s_1, \dots, s_r) = \frac{n!}{s_1! \dots s_r!} p_1^{s_1} \dots p_r^{s_r}, \quad s_1 \geq 0, \dots, s_r \geq 0, \quad (5.3)$$

$$s_1 + \dots + s_r = n. \quad \sum_{s_1 + \dots + s_r = n} P_n(s_1, \dots, s_r) = 1.$$

Это так называемое *полиномиальное распределение*. Название объясняется тем, что вероятность (5.3) является общим членом разложения полинома $(p_1 + \dots + p_r)^n$. Множитель $n!/(s_1! \dots s_r!)$ равен числу различных исходов в n испытаниях, при которых ω_j наблюдается s_j раз, $j = 1, \dots, r$. Для того чтобы получить это число, рассмотрим последовательность $\{\omega_1, \omega_2, \dots, \omega_n\}$, содержащую s_j исходов $\{\omega_j\}$, $j = 1, \dots, r$. Всего существует $n!$ возможностей расположить исходы $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_n\}$, однако среди них лишь $n!/(s_1! \dots s_r!)$ различных расположений, так как $s_1!$ перестановок s_1 исходов $\{\omega_1\}, \dots, s_r!$ перестановок s_r исходов $\{\omega_r\}$, приводят к эквивалентным расположениям.

С рассмотренными распределениями тесно связано так называемое *отрицательно-биномиальное распределение*, называемое также *распределением Паскаля*. Речь идет о вероятности того, что в последовательности испытаний Бернулли для достижения n успехов потребуется $n + k$ испытаний. Искомая вероятность равна

$$p(n, n + k) = C_{n+k-1}^{n-1} p^n q^k, \quad k = 0, 1, \dots \quad (5.4)$$

Действительно, выражение (5.4) дает вероятность того, что в серии $n + k - 1$ испытаний будет ровно k неудач, а следующее, $n + k$ -ое, испытание приведет к успеху, $C_{n+k-1}^k q^k p^{n-1} p = C_{n+k-1}^{n-1} p^n q^k$. Название распределения (5.4) связано со следующим равенством:

$$C_{-n}^k = \frac{(-n)(-n-1)\dots(-n-k+1)}{k!} = (-1)^k C_{n+k-1}^k$$

которое позволяет переписать распределение (5.4) в виде $p(n, n + k) = C_{-n}^k p^n (-q)^k$, $k = 0, 1, \dots$. Заметим, что

$$\sum_{k=0}^{\infty} C_{-n}^k (-q)^k = (1 - q)^{-n} = p^{-n},$$

и, следовательно, $\sum_{k=0}^{\infty} p(n, n + k) = 1$. При $n = 1$ распределение (5.4) называется *геометрическим*, $p(1, k + 1) = pq^k$, $k = 0, 1, \dots$

В качестве иллюстрации рассмотрим следующую *задачу С. Банаха*. Пусть a и b — две коробки, содержащие по n спичек. Некто пользуется ими, выбирая коробку a или b соответственно с вероятностями $p(a) = p$ или $p(b) = 1 - p = q$ и доставая каждый раз по одной спичке. Какова вероятность того, что в тот момент, когда выбранная коробка окажется пустой, в другой будет $r \leq n$ спичек?

Для решения задачи заметим, что если вынутая коробка пуста, а в другой в этот момент r спичек, то это означает, что до этого коробки вынимались $2n - r$ раз, причем одна из них — n раз. Пусть A — событие, состоящее в том, что вынутая коробка пуста. Такой коробкой может оказаться как a , так и b . Обозначим $P(A|a)$ и $P(A|b)$ соответствующие условные вероятности. Тогда искомая вероятность равна

$$P(A) = P(A|a)p(a) + P(A|b)p(b).$$

Поскольку $P(A|a)$ — вероятность вынуть пустую коробку при условии, что вынута коробка a , то речь идет о вероятности того, что перед этим коробка a доставалась n раз из общего числа $2n - r$ выборов. Следовательно, $P(A|a) = C_{2n-r}^n p^n q^{n-r}$. Аналогично $P(A|b) = C_{2n-r}^n q^n p^{n-r}$. Поэтому $P(A) = C_{2n-r}^n (p^{n+1} q^{n-r} + q^{n+1} p^{n-r})$.

Следует отличать найденную вероятность от вероятности того, что в тот момент, когда одна из коробок опустела, во второй оказалось r спичек. И вообще, коробка, вынутая пустой, не обязательно опустела первой. Читателю предлагается самостоятельно подсчитать вероятность, о которой идет речь.

Рассмотрим задачу Банаха, пользуясь отрицательно-биномиальным распределением. Искомая вероятность совпадает с вероятностью того, что для $(n + 1)$ -го «выбора» коробки a или b потребуется $2n - r + 1$ «испытаний». Обозначим A событие, состоящее в том, что для $n + 1$ «выбора» a или b потребуется $2n - r + 1$ «испытаний». Тогда $A = A \cap a + A \cap b$ и согласно формуле (5.4) для каждого из событий $A \cap a$ или $A \cap b$ найдем $P(A \cap a) = C_{2n-r}^n p^{n+1} q^{n-r}$, $P(A \cap b) = C_{2n-r}^n q^{n+1} p^{n-r}$, что приводит к найденному ранее значению вероятности $P(A)$.

1.6. Распределение Пуассона

Формулы биномиального распределения вероятностей приводят при больших n к громоздким вычислениям. Важно поэтому иметь приближенные, но зато достаточно простые формулы для вычисления соответствующих вероятностей. В частности, нередко встречаются задачи, в которых рассматривается большое число независимых испытаний, в которых вероятность наступления некоторого события A при каждом отдельном испытании мала.

В этом случае вероятности $P_n(k)$, $k = 0, \dots, n$, в (5.2) могут быть приближенно вычислены по так называемой *формуле Пуассона*. Эта формула получается как предельная для биномиального распределения, когда число испытаний стремится к бесконечности, а вероятность успеха стремится к нулю, однако в пределах одной последовательности независимых испытаний ни то, ни другое, по определению, невозможно. Поэтому естественно введение следующей конструкции.

Представим себе серию независимых испытаний, состоящую из конечного числа испытаний, затем новую серию, затем еще новую и т. д. Речь идет о *последовательности серий испытаний*. Пусть при каждом испытании в каждой серии может наступить или не наступить некоторое событие A , и вероятность наступления A при отдельном испытании остается постоянной в пределах каждой серии (как это требуется для последовательности независимых испытаний), но может меняться от серии к серии. В этих условиях справедлива

Теорема 1.6.1. (Теорема Пуассона). Пусть дана последовательность $\{s_n\}$ серий взаимно независимых испытаний, состоящих соответственно из $1, 2, \dots, n, \dots$ испытаний, и пусть вероятность p события A при каждом испытании n -й серии равна λ/n , где λ — постоянная (не зависящая от n). Тогда вероятность $P_n(m)$ того, что число наступлений события A в n -й серии будет равно m , при $n \rightarrow \infty$ и фиксированном m стремится к $(\lambda^m/m!)e^{-\lambda}$, $m = 0, 1, \dots$
Доказательство. Согласно равенствам (5.2) и условию теоремы

$$\begin{aligned} \lim_{n \rightarrow \infty} P_n(m) &= \lim_{n \rightarrow \infty} \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} = \\ &= \lim_{n \rightarrow \infty} \frac{n!}{m!(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} = \lim_{n \rightarrow \infty} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \times \\ &\times \left\{ \frac{n(n-1) \dots (n-m+1)}{n^m} \left(1 - \frac{\lambda}{n}\right)^{-m} \right\} = \frac{\lambda^m}{m!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \times \\ &\times \lim_{n \rightarrow \infty} \left\{ 1 \left(1 - \frac{\lambda}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-m} \right\} = \frac{\lambda^m}{m!} e^{-\lambda}. \blacksquare \end{aligned} \quad (6.1)$$

Распределение вероятностей, определяемое формулой

$$P(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, 2, \dots, \quad \lambda > 0, \quad (6.2)$$

называется *распределением Пуассона* и является одним из важнейших в теории вероятностей. Эта формула получена нами как предельная для последовательности серий независимых испытаний, в которой число испытаний стремится к бесконечности, а вероятность наступления события A стремится к нулю. Поэтому ясно, что если мы хотим воспользоваться приближенной формулой

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \quad \lambda = np, \quad m = 0, 1, \dots, m \quad (6.3)$$

для одной серии из n испытаний, то n должно быть велико, а вероятность наступления события A в этой серии мала, т. е. формула (6.3) применима для «редких событий». Оценим погрешность, возникающую при замене биномиального распределения на распределение Пуассона

при конечных n и при $0 \leq m \leq n$. Согласно (6.1)

$$\begin{aligned} j(m, n) &\equiv \ln \frac{P_n(m)}{P(m)} = \ln \left\{ \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \right. \\ &\quad \left. \dots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-m} e^\lambda \right\} = \\ &= \sum_{k=1}^{m-1} \ln \left(1 - \frac{k}{n}\right) + (n-m) \ln \left(1 - \frac{\lambda}{n}\right) + \lambda. \end{aligned} \quad (6.4)$$

Оценим сверху и снизу $\ln(1-x)$ при $0 \leq x < 1$:

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots < -x - \frac{x^2}{2} < -x, \quad (6.5)$$

$$\begin{aligned} \ln(1-x) &= -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots > -x - \frac{x^2}{2} - \frac{x^3}{2} - \frac{x^4}{2} - \dots = \\ &= -x - \frac{x^2}{2}(1+x+x^2+\dots) = -x - \frac{x^2}{2} \frac{1}{1-x} > -\frac{1}{1-x}. \end{aligned} \quad (6.6)$$

Используя (6.5) и (6.6), оценим (6.4):

$$\begin{aligned} j(m, n) &< -\frac{1}{n} \sum_{k=1}^{m-1} k + \lambda + (n-m) \left(-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2} \right) = \\ &= -\frac{m(m-1)}{2n} + \frac{\lambda m}{n} - \frac{\lambda^2}{2n} + \frac{m\lambda^2}{2n^2} = -\frac{(m-\lambda)^2}{2n} + \frac{m}{2n} + \frac{m\lambda^2}{2n^2} \end{aligned} \quad (6.7)$$

и

$$\begin{aligned} j(m, n) &> -\frac{1}{1 - \frac{m-1}{n}} \frac{1}{n} \sum_{k=1}^{m-1} k + \lambda + (n-m) \left(-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2 \left(1 - \frac{\lambda}{n}\right)} \right) = \\ &= -\frac{m(m-1)}{2(n-m+1)} + \frac{m\lambda}{n} - \frac{\lambda^2}{2(n-\lambda)} + \frac{m\lambda^2}{2n(n-\lambda)} \geq \\ &\geq -\frac{(m-\lambda)^2}{2(n-m+1)} + \frac{(m-\lambda-1)\lambda^2}{2(n-\lambda)(n-m+1)} - \frac{m(m-1)\lambda}{n(n-m+1)}. \end{aligned} \quad (6.8)$$

Представив теперь биномиальное распределение в виде

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda} e^{j(m,n)}, \quad m = 0, 1, \dots, n, \quad (6.9)$$

из (6.7) и (6.8) заключаем, что множитель $e^{j(m,n)}$ отличается от единицы на величину порядка $1/n$ при $n \rightarrow \infty$ и любом фиксированном m . На практике формула (6.3) служит хорошим приближением, если $n \geq 100$, а $0 \leq np < 10$ ($m = 0, 1, \dots, n$).

Обратимся к распределению Пуассона. Оно определено для всех целых неотрицательных $m = 0, 1, 2, \dots$

$$\sum_{m=0}^{\infty} P(m) = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = 1. \quad (6.10)$$

Пусть λ фиксировано, рассмотрим поведение $P(m)$ как функции m . Поскольку $P(m)/P(m-1) = \lambda/m$, то $P(m)$ возрастает при увеличении m от нуля до $m_0 = [\lambda]$ ($[\lambda]$ — целая часть числа λ), а затем убывает; $\max_m P(m) = \frac{\lambda^{[\lambda]}}{[\lambda]!} e^{-\lambda}$. Если λ — целое, то $P(m)$ имеет максимальное значение при $m = \lambda$ и $m = \lambda - 1$.

Распределение Пуассона описывает вероятности маловероятных (редких во времени) событий, таких, как число отказов радиоэлектронной аппаратуры за время t , число радиоактивных атомов, распавшихся к моменту t , число вызовов на телефонной станции и т. д.

Проиллюстрируем сказанное на примерах.

Пример 1.6.1. В лотерее в среднем разыгрывается один выигрыш на 1000 номеров. Какова вероятность, имея 100 билетов, получить не менее 2 выигрышей?

Это схема Бернулли с $n = 100$, вероятностью успеха $p = 1/1000$, так что $\lambda = np = 0,1$. Искомая вероятность

$$p = \sum_{m=2}^{100} P_{100}(m) = 1 - P_{100}(0) - P_{100}(1).$$

По формуле Пуассона $P_{100}(0) \approx P(0) = e^{-\lambda} \approx 0,9$, $P_{100}(1) \approx P(1) = \lambda e^{-\lambda}$. Таким образом, $p \approx 1 - 0,9 - 0,09 = 0,01$.

Пример 1.6.2. Вероятность зарегистрировать частицу счетчиком равна 10^{-4} . Какое наименьшее число частиц должно вылететь из источника для того, чтобы с вероятностью, не меньшей 0,99, счетчик зарегистрировал более 3 частиц?

Обозначим n искомое число частиц, а событие {счетчик зарегистрировал более трех частиц} обозначим A . Так как $p(A) = 1 - p(\bar{A})$, то по формуле Пуассона

$$p(\bar{A}) = P_n(0) + P_n(1) + P_n(2) + P_n(3) \approx P(0) + P(1) + P(2) + P(3) = e^{-\lambda}(1 + \lambda + \lambda^2/2 + \lambda^3/6) \leq 0,01, \quad (6.11)$$

ибо по условию $p(A) \geq 0,99$. По таблицам для распределения Пуассона находим λ , удовлетворяющее неравенству (6.11), $\lambda = 10,7$. Отсюда $n = \lambda/p = 10,7 \cdot 10^4$.

Для полиномиального распределения справедлив аналог теоремы Пуассона. Рассмотрим его на примере следующей задачи. Пусть T — область евклидова пространства, в которую взаимно независимо наугад бросаются N точек, обозначим $t_0, t_1, \dots, t_n$ — непересекающиеся области, такие, что $\bigcup_{j=0}^n t_j = T$. Какова вероятность того, что в области $t_1, \dots, t_n$, попадет соответственно $k_1, \dots, k_n$ точек?

Каждый исход опыта состоит в попадании точки в одну из областей t_j , $j = 0, 1, \dots, n$. Вероятности исходов равны соответственно

$$\begin{aligned} p_1 &= \frac{|t_1|}{|T|}, \dots, p_n = \frac{|t_n|}{|T|}, \\ p_0 &= 1 - p_1 - \dots - p_n = \frac{|T| - |t_1| - \dots - |t_n|}{|T|}, \end{aligned} \quad (6.12)$$

где $|t_j|$ — объем области t_j , $|T|$ — объем области T . Эту задачу естественно рассматривать как схему N независимых испытаний с $(n+1)$ -м исходом. Обозначим $\xi(t_j)$ число частиц, попавших в j -ю область, $j = 0, \dots, n$. Воспользовавшись полиномиальным распределением, получим

$$\begin{aligned} P\{\xi(t_1) = k_1, \dots, \xi(t_n) = k_n, \xi(t_0) = k_0\} &= \frac{N!}{k_0! k_1! \dots k_n!} p_0^{k_0} p_1^{k_1} \dots p_n^{k_n}, \\ k_0 + k_1 + \dots + k_n &= N. \end{aligned} \quad (6.13)$$

Предположим теперь, что в (6.12) $p = p_1 + \dots + p_n \ll q = 1 - p_0$, и как в доказательстве теоремы 1.6.1, зафиксируем $k = k_1 + \dots + k_n$, среднее число точек на единицу объема $\lambda = N/|T|$ и устремим $N \rightarrow \infty$. Тогда в (6.13) согласно формуле Стирлинга при $N \rightarrow \infty$

$$\frac{N!}{\sqrt{2\pi N} e^{-N} N^N} \rightarrow 1, \quad \frac{(N-k)!}{\sqrt{2\pi(N-k)} (N-k)^{N-k} e^{-N+k}} \rightarrow 1. \quad (6.14)$$

и при этом будет выполнено условие $p \ll 1 - p_0$, ибо

$$\begin{aligned} p_j^{k_j} &= \left(\frac{|t_j|}{|T|} \right)^{k_j} = (|t_j| \lambda)^{k_j} N^{-k_j} \rightarrow 0, \quad j = 1, \dots, n, \\ p_0^{k_0} &= \left(1 - \frac{|t_1| + \dots + |t_n|}{|T|} \right)^{N-k} = \\ &= \left(1 - \lambda \frac{|t_1| + \dots + |t_n|}{N} \right)^{N-k} \rightarrow e^{-\lambda(|t_1| + \dots + |t_n|)} \end{aligned} \quad (6.15)$$

Учитывая (6.14) и (6.15) в (6.13), найдем

$$\begin{aligned}
\lim_{N \rightarrow \infty} \frac{N!}{k_1! \dots k_n! (N-k)!} p_1^{k_1} \dots p_n^{k_n} p_0^{k_0} &= \lim_{N \rightarrow \infty} \left[\frac{\sqrt{2\pi N}}{\sqrt{2\pi(N-k)}} \times \right. \\
&\times \frac{N^{N-k}}{(N-k)^{N-k}} \frac{e^{-N} N^k}{e^{-(N-k)}} \frac{(|t_1|\lambda)^{k_1}}{N^{k_1}} \dots \frac{(|t_n|\lambda)^{k_n}}{N^{k_n}} \times \\
&\times \left(1 - \frac{\lambda(|t_1| + \dots + |t_n|)}{N} \right)^{N-k} \frac{1}{k_1! \dots k_n!} \Big] = \\
&= \frac{(|t_1|\lambda)^{k_1} \dots (|t_n|\lambda)^{k_n}}{k_1! \dots k_n!} \lim_{N \rightarrow \infty} \left[e^k \left(1 - \frac{k}{N} \right)^{-N} \times \right. \\
&\times \left. \left(1 - \frac{\lambda(|t_1| + \dots + |t_n|)}{N} \right)^{N-k} \right] = \\
&= \frac{(|t_1|\lambda)^{k_1} \dots (|t_n|\lambda)^{k_n}}{k_1! \dots k_n!} e^{-\lambda(|t_1| + \dots + |t_n|)} = \prod_{j=1}^n \frac{(|t_j|\lambda)^{k_j}}{k_j!} e^{-\lambda|t_j|}.
\end{aligned} \tag{6.16}$$

Таким образом, в пределе при $N \rightarrow \infty$ и фиксированных λ и k для вероятности того, что в области $t_1, \dots, t_n$ попадет соответственно $k_1, \dots, k_n$ точек, получено выражение

$$\begin{aligned}
P\{\xi(t_1) = k_1, \dots, \xi(t_n) = k_n\} &= \\
\prod_{j=1}^n \frac{(|t_j|\lambda)^{k_j}}{k_j!} e^{-\lambda|t_j|}, \quad k_1 + \dots + k_n = k,
\end{aligned} \tag{6.17}$$

представляющее собой n -мерное распределение Пуассона.

1.7. Локальная и интегральная предельные теоремы Муавра–Лапласа

Как мы уже отмечали в начале предыдущего параграфа, при большом числе испытаний формулы для вычисления вероятностей $P_n(m)$, $m = 0, \dots, n$, (5.2) биномиального распределения весьма громоздки. В связи с этим удобно иметь более простые приближенные формулы. Для случая, когда вероятность p успеха при каждом испытании мала, мы установили приближенные равенства $P_n(m) \approx P(m)$, $m = 0, 1, \dots, n$, (6.3), — распределение Пуассона. Сейчас мы рассмотрим другую предельную форму биномиального распределения для случая, когда вероятность p успеха отлична от нуля и единицы.

Теорема 1.7.1. (Локальная теорема Муавра–Лапласа). Если вероятность события A в n взаимно независимых испытаниях равна p , $0 < p < 1$, то вероятность $P_n(m)$ того, что в этих испытаниях событие A наступит m раз, удовлетворяет при $n \rightarrow \infty$ соотношению

$$\lim_{n \rightarrow \infty} \frac{\sqrt{npq} P_n(m)}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}}} = 1, \tag{7.1}$$

в котором $q = 1 - p$, $x_m = \frac{m - np}{\sqrt{npq}}$, $x_m \in [a, b]$, где $a < b$, a и b — любые конечные фиксированные числа.

Стремление к пределу в (7.1) равномерно относительно всех m , для которых $x_m \in [a, b]$.

Доказательство. Так как

$$\sqrt{npq} P_n(m) = \sqrt{npq} \frac{n!}{m!(n-m)!} p^m q^{n-m},$$

то, воспользовавшись формулой Стирлинга

$$k! = \sqrt{2\pi k} k^k e^{-k} e^{\vartheta_k}, \quad |\vartheta_k| \leq \frac{1}{12k},$$

получим

$$\begin{aligned} \sqrt{npq} P_n(m) &= \sqrt{npq} \frac{1}{\sqrt{2\pi}} \frac{n^n \sqrt{n} e^{\vartheta_n - \vartheta_m - \vartheta_{n-m}} p^m q^{n-m}}{m^m (n-m)^{n-m} \sqrt{m} \sqrt{n-m}} = \\ &= \frac{1}{\sqrt{2\pi}} \left[\left(\frac{np}{m} \right)^m \left(\frac{nq}{n-m} \right)^{n-m} \right] \sqrt{\frac{n^2 pq}{m(n-m)}} e^{\vartheta_n - \vartheta_m - \vartheta_{n-m}} \equiv \\ &\equiv \frac{1}{\sqrt{2\pi}} A_n(x_m) B_n(x_m) C_n(x_m). \end{aligned} \quad (7.2)$$

Найдем пределы выражений $A_n(x_m)$, $B_n(x_m)$, $C_n(x_m)$ при $n \rightarrow \infty$.

Пусть $[a, b]$ — произвольный конечный интервал; будем рассматривать такие m , для которых $x_m \in [a, b]$. Так как $x_m = \frac{m - np}{\sqrt{npq}}$, то

$$\begin{aligned} m &= np + x_m \sqrt{npq}, \quad n - m = n(1 - p) - x_m \sqrt{npq} = \\ &= nq - x_m \sqrt{npq}, \quad a \leq x_m \leq b. \end{aligned} \quad (7.3)$$

Начнем с $C_n(x_m) = e^{\vartheta}$, $\vartheta = \vartheta_n - \vartheta_m - \vartheta_{n-m}$.

$$\begin{aligned} |\vartheta| &\leq |\vartheta_n| + |\vartheta_m| + |\vartheta_{n-m}| \leq \frac{1}{12} \left(\frac{1}{n} + \frac{1}{np + x_m \sqrt{npq}} + \right. \\ &\left. + \frac{1}{nq - x_m \sqrt{npq}} \right) = \frac{1}{12n} \left(1 + \frac{1}{p + x_m \sqrt{\frac{pq}{n}}} + \frac{1}{q - x_m \sqrt{\frac{pq}{n}}} \right). \end{aligned} \quad (7.4)$$

Согласно оценке в (7.4) $\vartheta \rightarrow 0$ при $n \rightarrow \infty$, причем, как нетрудно видеть, при $n \rightarrow \infty$ $x_m \sqrt{\frac{pq}{n}} \rightarrow 0$ равномерно относительно $x_m \in [a, b]$. Таким образом,

$$C_n(x_m) \rightarrow 1 \text{ при } n \rightarrow \infty \quad (7.5)$$

равномерно относительно $x_m \in [a, b]$. Далее, в силу (7.2), (7.3)

$$\begin{aligned} B_n(x_m) &= \sqrt{\frac{n^2 pq}{m(n-m)}} = \\ &= \sqrt{\frac{1}{\left(1 + x_m \sqrt{\frac{q}{np}}\right) \left(1 - x_m \sqrt{\frac{p}{nq}}\right)}} \rightarrow 1 \text{ при } n \rightarrow \infty, \end{aligned} \quad (7.6)$$

как и в (7.5) равномерно относительно $x_m \in [a, b]$.

Рассмотрим, наконец, $A_n(x_m)$. Пользуясь формулой

$$\ln(1+z) = z - \frac{z^2}{2} + O(z^3), \quad |z| < 1,$$

получим

$$\begin{aligned} \ln A_n(x_m) &= -m \ln \left(\frac{m}{np} \right) - (n-m) \ln \left(\frac{n-m}{nq} \right) = \\ &= -(np + x_m \sqrt{npq}) \ln \left(1 + x_m \sqrt{\frac{q}{np}} \right) - \\ &\quad -(nq - x_m \sqrt{npq}) \ln \left(1 - x_m \sqrt{\frac{p}{nq}} \right) = \\ &= - \left[(np + x_m \sqrt{npq}) \left(x_m \sqrt{\frac{q}{np}} - \frac{x_m^2}{2np} + O(n^{-3/2}) \right) + \right. \\ &\quad \left. + (nq - x_m \sqrt{npq}) \left(-x_m \sqrt{\frac{p}{nq}} - \frac{x_m^2 p}{2nq} + O(n^{-3/2}) \right) \right] = \\ &= - \left[x_m \sqrt{npq} + x_m^2 q - \frac{x_m^2 q}{2} + O(n^{-1/2}) - x_m \sqrt{npq} + \right. \\ &\quad \left. + x_m^2 p - \frac{x_m^2 p}{2} + O(n^{-1/2}) \right] = -\frac{1}{2} x_m^2 + O(n^{-1/2}), \end{aligned} \quad (7.7)$$

причем, поскольку при $n \rightarrow \infty$ $x_m \sqrt{\frac{q}{np}}$, $x_m \sqrt{\frac{p}{nq}}$ стремятся к нулю равномерно по $x_m \in [a, b]$, оценку O -членов можно выбрать независимой от m . Итак,

$$A_n(x_m) = e^{-\frac{1}{2} x_m^2 + O\left(\frac{1}{\sqrt{n}}\right)}, \quad n \rightarrow \infty. \quad (7.8)$$

Теперь утверждение (7.1) следует из сопоставления формул (7.2), (7.5), (7.6) и (7.8). ■

Установленная теорема дает оценку вероятности $P_n(m)$ при больших n и при фиксированном m :

$$P_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x_m^2}{2}}, \quad x_m = \frac{m - np}{\sqrt{npq}}. \quad (7.9)$$

Однако практически при большом числе испытаний n и не слишком малой вероятности p нас редко интересует вероятность того, что данное

событие наступит точно m раз. Важнее уметь оценивать вероятность того, что число наступлений события лежит в некоторых границах. Такую оценку можно получить с помощью *интегральной предельной теоремы Муавра–Лапласа*.

Теорема 1.7.2. (*Интегральная предельная теорема Муавра–Лапласа*). Пусть m — число наступлений события A в серии из n взаимно независимых испытаний, p — вероятность наступления A при каждом испытании, $0 < p < 1$, a и b — любые фиксированные числа, $a < b$. Тогда

$$\lim_{n \rightarrow \infty} P \left\{ a \leq \frac{m - np}{\sqrt{npq}} \leq b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \quad (7.10)$$

причем стремление к пределу равномерно относительно a и b , $-\infty < a < b < +\infty$.

Хотя доказательство этой теоремы может быть установлено на основании локальной предельной теоремы, мы предпочтем получить его как совсем простое следствие центральной предельной теоремы (см. §1.11).

Практическое применение интегральной предельной теоремы основано на приближенном равенстве

$$P \left\{ a \leq \frac{m - np}{\sqrt{npq}} \leq b \right\} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \quad (7.11)$$

Оценка соответствующей погрешности показывает, что приближенная формула (7.11) обеспечивает хорошую точность уже при значениях $npq \geq 10$ и, разумеется, тем лучшую точность, чем больше эта величина.

Рассмотрим некоторые **типичные задачи**, связанные с интегральной теоремой Муавра–Лапласа.

Задача 1.7.1. Заданы число испытаний n и вероятность успеха p при каждом испытании. Требуется найти вероятность того, что число успехов m будет заключено между заданными числами m_1 и m_2 , $0 \leq m_1 < m_2 \leq n$.

$$\begin{aligned} P\{m_1 \leq m \leq m_2\} &= P \left\{ \frac{m_1 - np}{\sqrt{npq}} \leq \frac{m - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}} \right\} \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{m_1 - np}{\sqrt{npq}}}^{\frac{m_2 - np}{\sqrt{npq}}} e^{-\frac{x^2}{2}} dx. \end{aligned} \quad (7.12)$$

Задача 1.7.2. Игральная кость бросается 12000 раз. Какова вероятность того, что число выпадений единицы будет заключено между

1900 в 2150? Здесь $n = 12000$, $p = 1/6$, $q = 5/6$, $\sqrt{npq} = 100/\text{sqrt}6$, $m_1 - np = -100$, $m_2 - np = 150$. Искомая вероятность p_x приближенно равна

$$\begin{aligned} p_x &\approx \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{6}}^{\frac{3\sqrt{6}}{2}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{3\sqrt{6}}{2}} e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{6}} e^{-\frac{x^2}{2}} dx = \\ &= \Phi\left(\frac{3\sqrt{6}}{2}\right) + \Phi(\sqrt{6}) = 0,99. \end{aligned}$$

Функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$, $x \in (0, \infty)$, называется *интегралом ошибок*. Иногда интегралом ошибок называют функцию $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$, $x \in (-\infty, \infty)$.

Задача 1.7.3. Пусть заданы числа p , a и β . Требуется определить, какое наименьшее число n испытаний надо произвести для того, чтобы с вероятностью, не меньшей β , частота m/n появлений успеха отклонялась от вероятности p не больше, чем на α . Таким образом, надо найти n из условия

$$P\left\{\left|\frac{m}{n} - p\right| \leq \alpha\right\} \geq \beta.$$

Поскольку

$$\begin{aligned} P\left\{\left|\frac{m}{n} - p\right| \leq \alpha\right\} &= P\left\{-\alpha\sqrt{\frac{n}{pq}} \leq \frac{m - np}{\sqrt{npq}} \leq \alpha\sqrt{\frac{n}{pq}}\right\} \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\alpha\sqrt{\frac{n}{pq}}}^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}} dx = 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right), \end{aligned}$$

то задача состоит в определении n из условия

$$2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right) \geq \beta. \quad (7.13)$$

Ее решение находим с помощью таблиц для интегралов ошибок.

Задача 1.7.4. Сколько раз надо бросить монету для того, чтобы с вероятностью, не меньшей 0,99, частота появления герба отличалась от вероятности $p = 1/2$ не больше чем на 0,01.

Имеем согласно (7.13) $2\Phi(y) \geq 0,99$. Из таблиц находим, что

$$\begin{aligned} y &\geq 2,58. \text{ Таким образом, } y = \alpha\sqrt{\frac{n}{pq}} \geq 2,58, \text{ т. е. } \sqrt{n} \geq 2,58 \frac{\sqrt{\frac{1}{2} \cdot \frac{1}{2}}}{0,01} = \\ &= 129 \text{ и } n \geq 16641. \end{aligned}$$

Задача 1.7.5. Пусть заданы числа n , p и β . Требуется определить границы возможных отклонений частоты появления успеха от вероятности p , т. е. надо найти α , для которого

$$P\left\{\left|\frac{m}{n} - p\right| \leq \alpha\right\} = \beta.$$

Согласно предыдущему примеру

$$P\left\{\left|\frac{m}{n} - p\right| \leq \alpha\right\} \approx 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right) = \beta,$$

отсюда по таблицам определяем α .

Задача 1.7.6. Вероятность попадания в цель $1/10$. Сделано 100 выстрелов, в каких пределах с вероятностью 0,8 будет лежать относительная частота попаданий. Здесь $p = 0,1$, $q = 0,9$, $\beta = 0,8$, $\sqrt{\frac{n}{pq}} = 33,33$, $2\Phi(\alpha \cdot 33,33) = 0,8$. Отсюда $\alpha = 0,04$. Таким образом, частота m/n попаданий с вероятностью 0,8 лежит в интервале $(1/10 - 0,04, 1/10 + 0,04)$.

Задача 1.7.7. Телефонная станция A , обслуживающая 2000 абонентов, должна соединять их с другой станцией B . Какое наименьшее число x линий должно связывать A с B , чтобы в 99% случаев вызовов нашлась свободная линия. Пусть в течение наиболее напряженного часа дня каждый абонент разговаривает с B в среднем 2 минуты.

Найдем x . Естественно рассматривать описанную ситуацию как схему Бернулли с $n = 2000$, p — вероятность вызова, $p = 1/30$. Число x определяется из условия: вероятность того, что число вызовов $\geq x$, должна быть меньше, чем 0,01, т. е.

$$P\{m \geq x\} < 0,01, \text{ или } P\{m \geq x\} = P\left\{\frac{m - np}{\sqrt{npq}} \geq \frac{x - np}{\sqrt{npq}}\right\} \approx \\ \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{x - np}{\sqrt{npq}}}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{2} - \Phi\left(\frac{x - np}{\sqrt{npq}}\right) < 0,01,$$

отсюда по таблицам

$$\frac{x - np}{\sqrt{npq}} \geq 2,327, \quad x \geq \frac{200}{3} + 2,327 \cdot 8,027 = 85,4.$$

Таким образом, $x = 86$.

1.8. Случайные величины, функции распределения

1.8.1. Случайные величины. Изучая схему взаимно независимых испытаний, мы, по существу, имели дело с типичным примером случайной величины, когда рассматривали число успехов в серии из n испытаний. Примерами случайных величин являются: число вызовов в

единицу времени на телефонной станции, время ожидания очередного вызова, число молекул газа, продиффундировавших из одного объема газа в другой, и т.д.

Для случайной величины характерно, что нельзя заранее указать значение, которое она примет, хотя, с другой стороны, множество ее возможных значений считается известным. Это множество может быть конечным, как в упомянутом случае числа успехов, может совпадать с положительной полупрямой $[0, \infty)$, как в случае времени ожидания, и т.д. Однако для полного задания случайной величины следует еще указать вероятности тех значений, которые она может принимать, точнее вероятность на множестве ее значений. (Конечно, вероятность задается на некоторой σ -алгебре подмножеств множества ее значений, но ради краткости говорят и так.)

Прежде чем приступить к изучению математического понятия случайной величины, вернемся к схеме независимых испытаний и проанализируем известную нам случайную величину — число успехов. В рассматриваемом случае пространство элементарных событий Ω состоит из 2^n элементарных событий — последовательностей вида $\{\omega\} = \{\omega_1, \omega_2, \dots, \omega_n\}$. В схеме испытаний Бернулли нас интересовали события $A_k, k = 0, \dots, n$, где A_k составляют те последовательности, в которых успех ω_1 встречается k раз. Следовательно, A_k содержит C_n^k таких элементарных событий $\{\omega\}$, а поскольку вероятность каждого из них есть $P_n(\{\omega\}) = p^k q^{n-k}$, то $P_n(A_k) = C_n^k p^k q^{n-k}, k = 0, \dots, n$. Рассмотрим функцию $\xi = \xi(\omega)$, определенную на данном Ω равенствами

$$\xi(\omega) = k, \omega \in A_k, k = 1, \dots, n. \quad (8.1)$$

То, что эти равенства действительно определяют $\xi(\omega)$ для каждого $\omega \in \Omega$, следует из равенства $A_0 + A_1 + \dots + A_n = \Omega$. Так определенная функция $\xi = \xi(\omega), \omega \in \Omega$, определяет число успехов в серии из n независимых испытаний Бернулли в том смысле, что число успехов в каждой последовательности испытаний ω равно (по определению) значению $\xi(\omega)$. Функция $\xi = \xi(\omega), \omega \in \Omega$, называется *случайной величиной*. В данном случае эта случайная величина — число успехов в серии из n испытаний Бернулли.

Обозначим $\{\omega : \xi(\omega) = k\}$ множество тех $\omega \in \Omega$, для которых $\xi(\omega) = k$. Следовательно, по определению $A_k = \{\omega : \xi(\omega) = k\}$ и $P_n(\{\omega : \xi(\omega) = k\}) = C_n^k p^k q^{n-k}$. Последние вероятности обычно записываются короче:

$$P_n(k) = C_n^k p^k q^{n-k} = P(\xi = k), k = 0, 1, \dots, n, \quad (8.2)$$

и называют распределением вероятностей (значений) случайной величины ξ . Таким образом, случайная величина — число успехов в серии из n испытаний Бернулли — имеет биномиальное распределение.

Формула (8.2) задает вероятность на алгебре всех подмножеств множества значений случайной величины ξ . Последнее в данном слу-

чае состоит из $n + 1$ точек, $\tilde{\Omega} = \{0, 1, \dots, n\}$, алгебра событий состоит из всех 2^{n+1} подмножеств $\tilde{\Omega}$.

Таким образом, со случайной величиной ξ оказывается связанным новое вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, в котором пространство элементарных событий есть множество значений случайной величины, $\tilde{\mathcal{F}}$ — алгебра всех подмножеств $\tilde{\Omega}$, а вероятность $\tilde{P}$ связана с вероятностью на исходном вероятностном пространстве формулой (8.2): $\tilde{P}(\xi = k) = P_n(k)$.

Случайная величина $\xi = \xi(\omega)$, $\omega \in \Omega$, задает отображение вероятностного пространства $(\Omega, \mathcal{F}, P)$ на вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, определяет последнее и в этом смысле может быть названа канонической для $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. При этом каждой точке $k \in \tilde{\Omega}$ отвечает ее прообраз в Ω — множество $\{\omega : \xi(\omega) = k\} \subset \Omega$. Задание функции $\xi(\cdot)$ эквивалентно разбиению пространства элементарных событий: $\Omega = \{\omega : \xi(\omega) = 0\} + \{\omega : \xi(\omega) = 1\} + \dots + \{\omega : \xi(\omega) = n\}$. Утверждения « ξ попадает в $\tilde{A} \in \tilde{\mathcal{F}}$ » и « ω попадает в $A \in \mathcal{F}$ » эквивалентны.

Характерно, что в теоретико-вероятностных задачах явная зависимость $\xi = \xi(\omega)$ от $\omega \in \Omega$, как правило, не играет существенной роли.

В связи с распределением Пуассона можно рассматривать пространство элементарных событий Ω , состоящее из бесконечных последовательностей $\omega = (\omega_1, \omega_1, \omega_2, \dots)$ испытаний. Пусть ω содержит k успехов ω_1 , а событие A_k состоит из всех таких ω (которые содержат k раз ω_1). Тогда положим по определению $P(A_k) = \lambda^k e^{-\lambda}/k!$, $k = 1, \dots$, и определим случайную величину ξ равенствами $\xi(\omega) = k, \omega \in A_k$. Эта случайная величина имеет распределение Пуассона, так как $P\{\xi = k\} = (\lambda^k/k!)e^{-\lambda}$, $k = 0, 1, \dots$, $\lambda > 0$. (Заметим, что здесь естественно считать возможным значение $\xi(\omega) = \infty$, причем $P(\{\xi = \infty\}) = 0$.) Случайная величина $\xi = \xi(\omega)$, $\omega \in \Omega$, порождает новое вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, в котором $\tilde{\Omega} = \{0, 1, \dots, \infty\}$ — множество значений ξ , $\tilde{\mathcal{F}}$ — σ -алгебра всех подмножеств $\tilde{\Omega}$ и вероятность $\tilde{P}$ определена для каждого одноточечного подмножества $\tilde{\Omega}$ равенством $\tilde{P}(\xi = k) = (\lambda^k/k!)e^{-\lambda}$, $k = 0, 1, \dots$, $\tilde{P}(\xi = \infty) = 0$.

Это были примеры так называемых дискретных случайных величин. В общем случае случайная величина определяется следующим образом.

Определение 1.8.1. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство. Случайной величиной ξ (определенной на $(\Omega, \mathcal{F}, P)$) называется действительная функция $\xi = \xi(\omega)$, $\omega \in \Omega$, для которой множество $\{\omega : \xi(\omega) < x\}$ является событием (т.е. $\in \mathcal{F}$) для каждого действительного числа $x \in \mathcal{R}^1$.

1.8.2. Функции распределения и их свойства. В определении 1.8.1 требуется, чтобы для каждого $x \in \mathcal{R}^1$ множество $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$, и это условие гарантирует, что для каждого x определена вероятность события $\{\xi < x\}$: $F(x) = P(\{\xi < x\})$ (запись $\{\xi < x\}$

здесь и далее означает то же самое, что и $\{\omega \in \Omega, \xi(\omega) < x\}$ или $\{\omega : \xi(\omega) < x\}$.

Определение 1.8.2. Функция $F_\xi(x) \equiv F(x) = P(\{\xi < x\})$, $-\infty < x < \infty$, называется *функцией распределения* случайной величины¹⁾ ξ .

Заметим, что, как будет видно из дальнейшего, функция $F(\cdot)$ определяет вероятность на множестве значений ξ , являясь в то же время конструкцией, существенно более простой, чем вероятность $\tilde{P}$.

Поясним понятия случайной величины и функции распределения на примерах.

Пример 1.8.1. Пусть ξ — число успехов в серии из n испытаний Бернулли. Тогда функция распределения ξ определена равенством

$$F(x) = \begin{cases} 0, & x \leq 0, \\ \sum_{k \leq x} C_n^k p^k q^{n-k}, & 0 < x \leq n, \\ 1, & x > n. \end{cases} \quad (8.3)$$

Пример 1.8.2. Если ξ распределена по закону Пуассона, то ее функция распределения

$$F(x) = \begin{cases} 0, & x \leq 0, \\ \sum_{k \leq x} \frac{\lambda^k e^{-\lambda}}{k!}, & x > 0. \end{cases} \quad (8.4)$$

Пример 1.8.3. Говорят, что случайная величина ξ имеет *нормальное*, или *гауссово*, распределение $\mathcal{N}(\mu, \sigma^2)$, если ее функция распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz, \quad -\infty < x < \infty. \quad (8.5)$$

При этом определяемое ξ вероятностное пространство устроено следующим образом: $\tilde{\Omega}$ — действительная прямая $-\infty < x < \infty$, σ -алгебра событий $\tilde{\mathcal{F}}$ — σ -алгебра борелевских множеств на прямой. Для каждого события $A \in \tilde{\mathcal{F}}$

$$P(A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz. \quad (8.6)$$

Однако адекватное понимание последнего равенства возможно только в терминах интеграла Лебега. Мы ограничиваемся далее множествами A простой структуры: интервалами, конечными объединениями

¹⁾ В обозначении $F_\xi(\cdot)$ явно указана случайная величина, для которой $F_\xi(\cdot)$ — функция распределения. Дело в том, что случайная величина не определяется ее функцией распределения, см. замечание в конце §1.8.2.

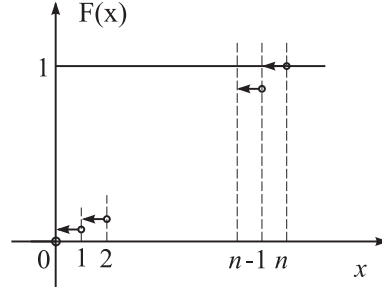


Рис. 1.10. Функция распределения бернуллиевской случайной величины.

ями интервалов и т.п. В этом случае интеграл (8.6) можно понимать как интеграл Римана (в том числе и как несобственный).

Перечислим основные **свойства функций распределения**.

1. Для $x_2 > x_1$ $\{\xi < x_2\} = \{\xi < x_1\} + \{x_1 \leq \xi < x_2\}$. Поскольку события в правой части этого равенства несовместны, то $P(\{\xi < x_2\}) = P(\{\xi < x_1\}) + P(\{x_1 \leq \xi < x_2\})$, следовательно,

$$P(\{x_1 \leq \xi < x_2\}) = F(x_2) - F(x_1). \quad (8.7)$$

Так как $P(\{x_1 \leq \xi < x_2\}) \geq 0$, то из равенства (8.7) следует, что $F(x)$, $x \in \mathcal{R}^1$, — *неубывающая функция*.

2. Из определения $F(\cdot)$ следует, что

$$0 \leq F(x) \leq 1, x \in \mathcal{R}^1.$$

3. $F(\cdot)$ непрерывна слева в каждой точке $x \in \mathcal{R}^1$, т.е.

$$F(x) = F(x - 0) = \lim_{x_k \uparrow x} F(x_k), \quad (8.8)$$

где последнее равенство — равенство по определению. Действительно, пусть $x_1 < x_2 < \dots < x_n < \dots < x$ — любая последовательность, стремящаяся к x слева, $x_k \uparrow x$, $\lim_{k \rightarrow \infty} x_k = x$. Событие $\xi < x$ можно представить в виде

$$\{\xi < x\} = \{\xi < x_1\} \cup \{\xi < x_2\} \cup \dots = \{\xi < x_1\} + \{x_1 \leq \xi < x_2\} + \dots$$

В силу σ -аддитивности вероятности и равенства (8.7) отсюда следует, что

$$\begin{aligned} F(x) &= P(\{\xi < x\}) = P(\{\xi < x_1\}) + P(\{x_1 \leq \xi < x_2\}) + \dots = \\ &= F(x_1) + [F(x_2) - F(x_1)] + \dots + [F(x_{k+1}) - F(x_k)] + \dots = \\ &= \lim_{k \rightarrow \infty} F(x_k), \end{aligned}$$

и (8.8) доказано.

4. Правое предельное значение $F(\cdot)$ в точке x равно $P(\{\xi \leq x\})$, т.е.

$$F(x + 0) \equiv \lim_{x_k \downarrow x} F(x_k) = P(\{\xi \leq x\}). \quad (8.9)$$

Действительно, пусть $x_1 > x_2 > \dots > x_n > \dots > x$ — любая последовательность, стремящаяся к x справа, $x_k \downarrow x$, $\lim_{k \rightarrow \infty} x_k = x$. Тогда на основании (8.7) и σ -аддитивности вероятности

$$\begin{aligned} \lim_{k \rightarrow \infty} F(x_k) &= F(x_1) - [F(x_1) - F(x_2)] - \dots - [F(x_k) - F(x_{k+1})] - \dots = \\ &= P(\{\xi < x_1\}) - [P(\{x_2 \leq \xi < x_1\}) + \dots + P(\{x_{k+1} \leq \xi < x_k\}) + \dots] = \\ &= P(\{\xi < x_1\}) - P(\{x < \xi < x_1\}) = P(\{\xi \leq x\}), \end{aligned}$$

что и доказывает (8.9).

Свойства 3 и 4 функции распределения можно получить, воспользовавшись непрерывностью вероятности, исследованной в §1.3.3. Действительно, если при $n \rightarrow \infty$ $x_{2,n} \downarrow x_1$, то в согласии с (8.9) $\lim_{x_{2,n} \downarrow x_1} P(\{\xi \in [x_1, x_{2,n}]\}) = \lim_{x_{2,n} \downarrow x_1} F(x_{2,n}) - F(x_1) = F(x_1 + 0) - F(x_1)$. С другой стороны, в силу непрерывности вероятности $\lim_{x_{2,n} \downarrow x_1} P(\{\xi \in [x_1, x_{2,n}]\}) = P(\{\xi \in \bigcap_{n=1}^{\infty} [x_1, x_{2,n}]\}) = P(\{\xi = x_1\})$, так как $\lim_{n \rightarrow \infty} [x_1, x_{2,n}] = \bigcap_{n=1}^{\infty} [x_1, x_{2,n}] = \{x_1\}$. Поэтому $F(x_1 + 0) - F(x_1) = P(\{\xi = x_1\})$. Если же при $n \rightarrow \infty$ $x_{1,n} \uparrow x_2$, то $\lim_{x_{1,n} \uparrow x_2} P(\{\xi \in [x_{1,n}, x_2]\}) = F(x_2) - F(x_2 - 0)$, а в силу непрерывности вероятности $\lim_{x_{1,n} \uparrow x_2} P(\{\xi \in [x_{1,n}, x_2]\}) = P(\{\xi \in \bigcap_{n=1}^{\infty} [x_{1,n}, x_2]\}) = P(\{\xi \in \emptyset\}) = 0$, ибо $\lim_{n \rightarrow \infty} [x_{1,n}, x_2] = \bigcap_{n=1}^{\infty} [x_{1,n}, x_2] = \emptyset$.

5. Справедливы следующие соотношения ($x_1 \leq x_2$ — любые):

$$P(\{x_1 \leq \xi \leq x_2\}) = F(x_2 + 0) - F(x_1), \quad (8.10)$$

так как $\{\xi < x_1\} + \{x_1 \leq \xi \leq x_2\} = \{\xi \leq x_2\}$. В частности, если $x_2 = x_1 = x$, то

$$P(\{\xi = x\}) = F(x + 0) - F(x). \quad (8.11)$$

Так как $\{\xi \leq x_1\} + \{x_1 < \xi \leq x_2\} = \{\xi \leq x_2\}$, то

$$P(\{x_1 < \xi \leq x_2\}) = F(x_2 + 0) - F(x_1 + 0), \quad (8.12)$$

и так как $\{\xi \leq x_1\} + \{x_1 < \xi < x_2\} = \{\xi < x_2\}$, то

$$P(\{x_1 < \xi < x_2\}) = F(x_2) - F(x_1 + 0). \quad (8.13)$$

6. Ввиду неубывания $F(\cdot)$ определим значения $F(-\infty)$ и $F(+\infty)$ равенствами

$$F(-\infty) = \lim_{n \rightarrow \infty} F(-n), \quad F(+\infty) = \lim_{n \rightarrow \infty} F(n).$$

Тогда

$$F(-\infty) = 0, \quad F(+\infty) = 1. \quad (8.14)$$

В самом деле,

$$\begin{aligned} F(+\infty) &= \lim_{n \rightarrow \infty} F(n) = F(0) + [F(1) - F(0)] + \dots + \\ &+ [F(n+1) - F(n)] + \dots = P(\{\xi < 0\}) + P(\{0 \leq \xi < 1\}) + \dots + \\ &+ P(\{n \leq \xi < (n+1)\}) + \dots = P(\{\xi < +\infty\}) = 1. \end{aligned}$$

Аналогично $F(-\infty) = 0$.

Таким образом, каждая функция распределения не убывает, непрерывна слева и удовлетворяет условиям (8.14). Верно и обратное утверждение: если $F(x)$, $-\infty < x < \infty$ удовлетворяет перечисленным условиям, то она может рассматриваться как функция распределения некоторой случайной величины.

Заметим, что существует сколько угодно случайных величин с данной функцией распределения. Например, функция распределения случайной величины ξ , принимающей значения -1 и $+1$ с вероятностью $1/2$, совпадает с функцией распределения $\eta = -\xi$, хотя $\xi \neq \eta$ — с вероятностью единица. Отметим также, что из неубывания произвольной функции распределения $F(\cdot)$ и неравенств $0 \leq F(x) \leq 1$, $x \in (-\infty, \infty)$, следует, что $F(\cdot)$ имеет не более счетного числа скачков. Действительно, можно пронумеровать все скачки следующим образом: сначала все скачки, большие $1/2$ (их ≤ 1), затем скачки, большие $1/3$ (их ≤ 2) и т.д. Таким образом, $F(\cdot)$ непрерывна всюду на $\mathcal{R}^1$, за исключением не более чем счетного множества точек.

1.8.3. Дискретные и абсолютно непрерывные случайные величины. Рассмотренные выше случайные величины, распределенные нормально и по биномиальному закону, являются характерными примерами двух основных классов случайных величин: непрерывных и дискретных.

Определение 1.8.3. Случайная величина ξ называется дискретной, если множество ее значений конечное или счетное.

Пусть $(\Omega, \mathcal{F}, P)$ — некоторое вероятностное пространство, $\Omega = \sum_{k=1}^{\infty} \Omega_k$ — некоторое разбиение Ω на (непересекающиеся) подмно-

жества $\Omega_1, \Omega_2, \dots$, $P(\Omega_k) = p_k$ и $\chi_k(\omega) = \begin{cases} 1, & \text{если } \omega \in \Omega_k, \\ 0, & \text{если } \omega \in \Omega \setminus \Omega_k \end{cases}$ — индикаторная функция Ω_k , $k = 1, 2, \dots$

Любую дискретную случайную величину ξ , определенную на $(\Omega, \mathcal{F}, P)$, можно представить как функцию

$$\xi(\omega) = \sum_{k=1}^{\infty} x_k \chi_k(\omega), \quad \omega \in \Omega, \quad (8.15)$$

принимающую значение x_k на множестве Ω_k с вероятностью

$$p_k = P(\{\omega \in \Omega, \xi(\omega) = x_k\}) = P(\Omega_k), \quad k = 1, 2, \dots \quad (8.16)$$

Как уже было отмечено, при изучении вероятностных свойств дискретной случайной величины можно не обращаться к исходному вероятностному пространству $(\Omega, \mathcal{F}, P)$. Достаточно рассматривать дискретное вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, порожденное этой случайной величиной, в котором $\tilde{\Omega} = \{x_1, x_2, \dots\}$ — множество значений ξ , $\tilde{\mathcal{F}}$ — σ -алгебра всех подмножеств $\tilde{\Omega}$ и $\tilde{P}$ — вероятность на $\tilde{\mathcal{F}}$, определенная для каждого одноточечного подмножества $\tilde{\Omega}$ равенством

$$\tilde{P}(\{x_k\}) = p_k = \tilde{P}(\xi = x_k) = P(\Omega_k), k = 1, 2, \dots \quad (8.16^*)$$

Согласно (8.16*) для любого $\tilde{A} \in \tilde{\mathcal{F}}$

$$\tilde{P}(\tilde{A}) \equiv \tilde{P}(\xi \in \tilde{A}) = \sum_{k: x_k \in \tilde{A}} p_k = P(\{\omega \in \Omega, \xi(\omega) \in \tilde{A}\}). \quad (8.17)$$

Корректность определения вероятности $\tilde{P}$ на $\tilde{\mathcal{F}}$ следует из леммы о суммировании по блокам, см. §1.3.2.

Случайная величина $\xi(\cdot) : \Omega \rightarrow \tilde{\Omega}$ называется *канонической* для $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$.

Функция распределения (вероятностей значений) ξ определяется равенством

$$F_\xi(x) \triangleq P(\{\omega \in \Omega, \xi(\omega) < x\}) = \sum_{k: x_k < x} p_k, -\infty < x < \infty, \quad (8.18)$$

согласно которому она кусочно-постоянна, возрастает скачками в точках $x_1, x_2, \dots$, и определяет вероятность $\tilde{P}$ на $\tilde{\mathcal{F}}$, поскольку $F(x_k + 0) - F(x_k) = p_k, k = 1, 2, \dots$; ее значения не зависят от способа нумерации значений ξ .

Определение 1.8.4. Случайная величина ξ называется *непрерывной*, точнее, *абсолютно непрерывной*, если ее функция распределения представима в виде

$$F(x) = \int_{-\infty}^x p(y) dy, \quad x \in \mathcal{R}^1 = (-\infty, \infty). \quad (8.19)$$

Функция $p(y), -\infty < y < \infty$, называется *плотностью распределения вероятностей* (или, короче — *плотностью*) случайной величины ξ и далее предполагается неотрицательной и кусочно-непрерывной. Плотность $p(\cdot)$ полностью определяет функцию распределения $F(\cdot)$, а в точках непрерывности плотность определяется функцией распределения, так как в этих точках $p(x) = \frac{dF(x)}{dx}, x \in \mathcal{R}^1$, и, таким образом, в этих точках свойство неотрицательности плотности является следствием неубывания $F(\cdot)$.

Для любых $x_1 < x_2$

$$P(\{x_1 \leq \xi < x_2\}) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(y) dy. \quad (8.20)$$

Если, в частности, плотность $p(\cdot)$ непрерывна на интервале $[x, x + \Delta x]$, $\Delta x > 0$, то согласно (8.20) и теореме о среднем для интеграла

$$P(\{x \leq \xi < x + \Delta x\}) = p(x)\Delta x + o(\Delta x). \quad (8.21)$$

Заметим, что для непрерывной случайной величины для любого $x \in \mathcal{R}^1$

$$P(\{\xi = x\}) = F(x+0) - F(x) = 0, \quad (8.22)$$

так как функция $F(\cdot)$ непрерывна, т.е. вероятность того, что ξ примет любое фиксированное значение из $\mathcal{R}^1$, равна нулю. Поэтому в формулах (8.20) и (8.21) строгие неравенства могут быть заменены на нестрогие. Очевидно, что

$$\int_{-\infty}^{\infty} p(y) dy = F(+\infty) = 1. \quad (8.23)$$

Примером непрерывной случайной величины является нормальная $\mathcal{N}(\mu, \sigma^2)$ случайная величина с плотностью

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathcal{R}^1, \quad (8.24)$$

(пример 1.8.3).

С непрерывной случайной величиной, определенной на $(\Omega, \mathcal{F}, P)$, также связано новое вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, где $\tilde{\Omega}$ — действительная прямая $\mathcal{R}^1$, $\tilde{\mathcal{F}}$ — σ -алгебра борелевских множеств на прямой. Для каждого интервала (x_1, x_2) вероятность $\tilde{P}$ определяется по формуле (8.20). В теории меры доказывается, что тем самым вероятность $\tilde{P}$ определяется и для всякого события $A \in \tilde{\mathcal{F}}$, т.е. для любого борелевского множества на прямой. Отсюда следует, что и в этом случае *вероятность полностью P (8.20) определяется функцией распределения $F(\cdot)$* (8.19).

В дальнейшем будем считать, что случайная величина ξ задана, если задано вероятностное пространство $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, или, иначе, если задана функция распределения $F(\cdot)$, когда речь идет о дискретных или непрерывных случайных величинах.

Следует сказать, что, конечно, существуют случайные величины, которые не являются ни дискретными, ни непрерывными, ни их

комбинациями с функцией распределения вида: $F(x) = q_1 \sum_{k: x_k < x} p_k + q_2 \int_{-\infty}^x p(y) dy$, где $q_1, q_2 > 0$, $q_1 + q_2 = 1$, $\sum_k p_k = 1$, $\int_{-\infty}^{\infty} p(y) dy = 1$.

Речь идет о так называемых *сингулярных случайных величинах*.

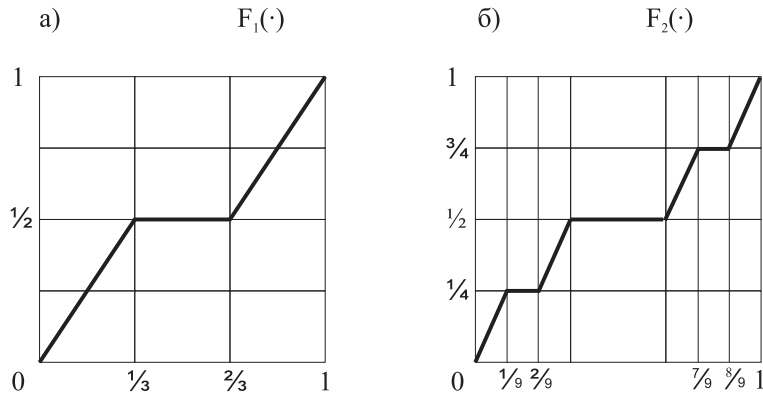


Рис. 1.11. Предел последовательности $F_1(x)$, $F_2(x)$, ..., называемый «канторовой лестницей», является примером функции распределения сингулярной случайной величины.

Пример 1.8.4. Приведем пример функции распределения сингулярной случайной величины, известной как *канторова лестница*. Определим ее как предел последовательности непрерывных монотонно неубывающих функций, определенных на $[0, 1]$ и принимающих значения в $[0, 1]$.

Первая функция последовательности $F_1(\cdot)$ строится так. Область определения $F_1(\cdot)$ (отрезок $[0, 1]$) представим как $[0, 1/3] \cup [1/3, 2/3] \cup [2/3, 1]$ и на отрезках $[0, 1/3]$, $[1/3, 2/3]$ и $[2/3, 1]$ определим $F_1(\cdot)$ так, как показано на графике 1.11а. Затем на три равных части делим отрезки $[0, 1/3]$ и $[2/3, 1]$ и на каждом из отрезков $[0, 1/9]$, $[1/9, 2/9]$, $[2/9, 1/3]$, $[1/3, 2/3]$, $[2/3, 7/9]$, $[7/9, 8/9]$ и $[8/9, 1]$ определяем $F_2(\cdot)$ так, как показано на графике 1.11б, далее аналогично переопределяем $F_2(\cdot)$ на отрезках $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$ и $[8/9, 1]$, поделив каждый из них на три равных части, и т.д. Функция $F_\xi(x) = \lim_{n \rightarrow \infty} F_n(x)$, $x \in [0, 1]$, монотонно не убывает и непрерывна на $[0, 1]$ (проверьте). Суммарная длина отрезков, на которых $F_\xi(\cdot) = \text{const}$, равна $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots = \frac{1}{3}(1 + \frac{2}{3} + \frac{4}{9} + \dots) = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1$, т.е. $F_\xi(\cdot)$, будучи непрерывной и возрастающей на множестве меры ноль, принимает значения, заполняющие $[0, 1]$!

Так как $F'_\xi(\cdot) = 0$ на множестве отрезков, сумма длин которых равна единице, ее нельзя представить в виде интеграла от плотности, как в (8.19), и в виде суммы, как в (8.18), поскольку $F_\xi(\cdot)$ непрерывна. Сингулярные распределения представляют собой некоторую «экзотику» и в реальных задачах практически не встречаются. Мы исключаем случайные величины, подобные ξ , из дальнейшего изучения.

Замечание 1.8.1. Все сказанное о функциях распределения автоматически переносится на случай условных вероятностей. Если $P(B) > 0$, то $F(x|B) = P(\xi < x|B)$, $-\infty < x < \infty$, называется *условной функцией распределения* случайной величины ξ . Она обладает всеми указанными выше свойствами функции распределения.

1.8.4. Векторные (многомерные) случайные величины. В стохастических экспериментах обычно приходится учитывать взаимодействие различных случайных факторов. Это естественным образом приводит к рассмотрению многомерных случайных величин.

Определение 1.8.5. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, и $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)$, $\omega \in \Omega$, — случайные величины, определенные на $(\Omega, \mathcal{F}, P)$. Векторно-значная функция $\xi(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega))$, $\omega \in \Omega$, называется *случайным вектором*, или *n -мерной случайной величиной со значением в $\mathcal{R}^n$* , $\xi_j(\omega)$, $\omega \in \Omega$, $j = 1, \dots, n$, называют *координатами* случайного вектора $\xi(\cdot)$.

Поскольку все $\xi_j(\cdot)$, $j = 1, 2, \dots, n$, заданы на одном и том же вероятностном пространстве, а $\mathcal{F}$ замкнуто относительно операции конечного и счетного пересечения событий, то множество $\{\omega : \xi_1(\omega) < x_1, \dots, \xi_n(\omega) < x_n\} \in \mathcal{F}$ для любого набора чисел $x_1, \dots, x_n$.

Определение 1.8.6. Функция $F(x_1, x_2, \dots, x_n) = P(\{\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n\})$, $x_j \in \mathcal{R}^1$, $j = 1, \dots, n$, называется *функцией распределения n -мерной случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_n)$* .

Ради наглядности и краткости будем рассматривать двумерные случайные величины. Геометрически двумерная функция распределения $F(\cdot, \cdot)$, определенная равенством

$$F(x, y) = P(\{\xi < x, \eta < y\}), \quad (x, y) \in \mathcal{R}^2 = \mathcal{R}^1 \times \mathcal{R}^1, \quad (8.25)$$

задает вероятность попадания случайной точки (ξ, η) в бесконечный прямоугольник $\xi < x, \eta < y$ (заштрихованная часть на рис. 1.12).

Коротко перечислим основные свойства двумерной функции распределения $F(x, y)$, $(x, y) \in \mathcal{R}^2$:

- 1) $F(x, y)$ не убывает по x и по y ,
- 2) $F(\cdot, \cdot)$ непрерывна слева по каждому аргументу,
- 3) $F(\infty, \infty) = 1$, $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, где, по определению, $F(\infty, \infty) = \lim_{n, m \rightarrow \infty} F(n, m)$, $F(-\infty, y) = \lim_{n \rightarrow \infty} F(n, y)$, $F(x, -\infty) = \lim_{n \rightarrow \infty} F(x, -n)$; все эти пределы существуют ввиду неубывания $F(x, y)$ по $x \in \mathcal{R}^1, y \in \mathcal{R}^1$ (доказательства свойств 1–3 аналогичны проведенным в §1.8.3);

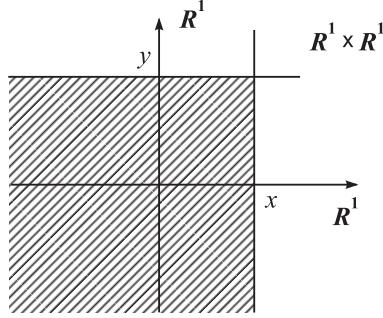


Рис. 1.12. Значение $F(x, y)$ равно вероятности включения случайного вектора (ξ, η) в заштрихованную область.

4) вероятность включения (ξ, η) в заштрихованную область на рис. 1.13

$$\begin{aligned} P(\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\}) = \\ = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1), \end{aligned} \quad (8.26)$$

$(x_i, y_j) \in \mathcal{R}^2, i, j, = 1, 2;$

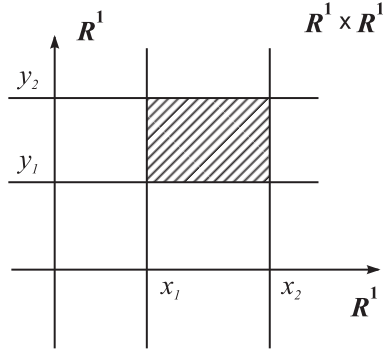


Рис. 1.13. Иллюстрация равенства (8.26).

5) пользуясь двумерной функцией распределения, можно найти функции распределения координат ξ, η (так называемые *маргинальные функции распределения*):

$$F_\xi(x) = F(x, \infty), \quad x \in \mathcal{R}^1, \quad F_\eta(y) = F(\infty, y), \quad y \in \mathcal{R}^1, \quad (8.27)$$

где $F(x, \infty) = \lim_{n \rightarrow \infty} F(x, n)$, $F(\infty, y) = \lim_{n \rightarrow \infty} F(n, y)$.

В самом деле, $\{\xi < x\} = \sum_{k=-\infty}^{\infty} \{\xi < x, k \leq \eta < k+1\}$. Поскольку события под знаком суммы попарно несовместны, то в силу

σ -аддитивности вероятности

$$\begin{aligned} F_\xi(x) &= P(\{\xi < x\}) = \sum_{k=-\infty}^{\infty} P(\{\xi < x, k \leq \eta < k+1\}) = \\ &= \sum_{k=-\infty}^{\infty} [F(x, k+1) - F(x, k)] = \lim_{N_1, N_2 \rightarrow \infty} \sum_{k=-N_1}^{N_2} [F(x, k+1) - \\ &\quad - F(x, k)] = \lim_{N_1, N_2 \rightarrow \infty} [F(x, N_2+1) - F(x, -N_1)] = \\ &= F(x, \infty) - F(x, -\infty) = F(x, \infty), \quad x \in \mathcal{R}^1, \end{aligned}$$

и первая формула в (8.27) обоснована, другая проверяется аналогично.

Определение 1.8.7. Случайный вектор называется *дискретным*, если каждая его координата — дискретная случайная величина, и абсолютно непрерывным, если существует кусочно-непрерывная неотрицательная функция $p(x, y)$, $(x, y) \in \mathcal{R}^2$, такая, что

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(z_1, z_2) dz_1 dz_2, \quad x \in \mathcal{R}^1, \quad y \in \mathcal{R}^1. \quad (8.28)$$

Функция $p(x, y)$, $(x, y) \in \mathcal{R}^1 \times \mathcal{R}^1$, называется *плотностью распределения вероятностей* значений случайного вектора (ξ, η) , короче — плотностью вероятности, или плотностью случайного вектора.

Плотность вероятности обладает следующими свойствами.

1. В точках непрерывности $p(\cdot, \cdot)$ справедливо равенство $p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$, $x, y \in \mathcal{R}^1$, и, таким образом, в этих точках тот факт, что $p(x, y) \geq 0$, следует из неубывания $F(x, y)$ по каждой переменной x и y , $x, y \in \mathcal{R}^1$.

2. Для любой квадратируемой области $D \subset \mathcal{R}^2$

$$P\{(\xi, \eta) \in D\} = \iint_D p(x, y) dx dy. \quad (8.29)$$

В самом деле, если D представляет собой произвольный прямоугольник $\Pi = \{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\}$, то (8.29) есть очевидное следствие равенств (8.26) и (8.28). В общем случае, как обычно, аппроксимируем область D суммой входящих и выходящих квадратов, для них справедливость (8.29) доказана только что, а затем устремим сторону квадратов к нулю. В силу предположенной квадратируемости области D и кусочной непрерывности $p(\cdot, \cdot)$ этот предельный переход приведет нас к (8.29).

В частности, если плотность $p(x, y)$ непрерывна при $x_1 \leq x \leq x_1 + \Delta x$, $y_1 \leq y \leq y_1 + \Delta y$, то с помощью равенства (8.29) и теоремы о среднем для интеграла, получим

$$P\{x_1 \leq \xi \leq x_1 + \Delta x, y_1 \leq \eta \leq y_1 + \Delta y\} = p(x, y) \Delta x \Delta y + o(\Delta x \Delta y). \quad (8.30)$$

3. Вследствие равенства $F(\infty, \infty) = 1$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1. \quad (8.31)$$

4) На основании равенств (8.28) и (8.27) заключаем, что если двумерная случайная величина (ξ, η) имеет плотность, то каждая ее координата имеет плотность, например,

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p(x, y) dy, \quad x \in \mathcal{R}^1, \quad p_{\eta}(y) = \int_{-\infty}^{\infty} p(x, y) dx, \quad y \in \mathcal{R}^1. \quad (8.32)$$

Рассмотрим несколько примеров.

Пример 1.8.5. Случайный вектор (ξ, η) называется равномерно распределенным в области D , если $p(x, y) = \begin{cases} 1/\mu(D), & (x, y) \in D, \\ 0, & (x, y) \notin D, \end{cases}$ где $\mu(D)$ — площадь области D . Пусть D^* — квадратируемая область на плоскости. Тогда, как в примерах на геометрические вероятности в §1.1,

$$P\{(\xi, \eta) \in D^*\} = \frac{\mu(D^* \cap D)}{\mu(D)}.$$

Пример 1.8.6. Случайный вектор $(\xi_1, \dots, \xi_n)$ называется *нормально распределенным*, если его плотность равна

$$p(x_1, \dots, x_n) = \left(\frac{\det A}{(2\pi)^n} \right)^{1/2} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n a_{ij} (x_i - \mu_i)(x_j - \mu_j) \right\}, \quad (8.33)$$

$$x_i \in \mathcal{R}^1, \quad i = 1, \dots, n,$$

где $A = \|a_{ij}\|$ — положительно определенная матрица (обратная к матрице ковариаций случайных величин $\xi_1, \dots, \xi_n$, см. ниже §1.9). Интегрируя (8.33) и пользуясь равенствами (8.32) (или их аналогами, если $n > 2$), можно показать, что каждая координата $\xi_j, j = 1, \dots, n$, нормально распределенного вектора имеет нормальное распределение.

В двумерном случае плотность (8.33) имеет вид

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{(x-a)^2}{\sigma_1^2} - \frac{2r(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} \right] \right\}, \quad x, y \in \mathcal{R}^1 \quad (8.34)$$

где

$$\sigma_1^2 = D\xi, \quad \sigma_2^2 = D\eta, \quad r = \frac{\text{cov}(\xi, \eta)}{\sigma_1 \sigma_2},$$

см. ниже, §1.9.

Рассмотрим в заключение этого пункта понятие *плотности условного распределения*. Пусть плотность $p(x, y)$, $(x, y) \in \mathcal{R}^1 \times \mathcal{R}^1$, случайного вектора (ξ, η) непрерывна и событие $B = \{y \leq \eta \leq y + \Delta y\}$, его вероятность $P(B) = P(\{y \leq \eta \leq y + \Delta y\}) = \int_y^{y+\Delta y} p_\eta(z) dz$, где согласно

$$(8.32) \quad p_\eta(z) = \int_{-\infty}^{\infty} p(x, z) dx, \quad z \in \mathcal{R}^1. \quad \text{Далее } P(\{\xi < x, y \leq \eta \leq y + \Delta y\}) = \int_{-\infty}^x \int_y^{y+\Delta y} p(t, z) dt dz, \text{ и, следовательно, в силу определения условной вероятности (см. §1.4), если } P(B) > 0, \text{ то}$$

$$\begin{aligned} F_\xi(x|B) &\equiv P(\{\xi < x|B\}) = \frac{P(\{\xi < x, y \leq \eta \leq y + \Delta y\})}{P(B)} = \\ &= \frac{\int_{-\infty}^x \int_y^{y+\Delta y} p(t, z) dt dz}{\int_y^{y+\Delta y} p_\eta(z) dz}, \quad x \in \mathcal{R}^1. \end{aligned} \quad (8.35)$$

Продифференцируем (8.35) по x , а затем устремим $\Delta y \rightarrow 0$ и воспользуемся теоремой о среднем для интеграла по промежутку $[y, y + \Delta y]$, в результате найдем:

$$p_\xi(x|y) \equiv \frac{dF_\xi(x|\eta=y)}{dx} = \frac{p(x, y)}{p_\eta(y)}, \quad p_\eta(y) \neq 0, \quad x, y \in \mathcal{R}^1. \quad (8.36)$$

Функция $p_\xi(x|y)$, $x \in \mathcal{R}^1$ называется *плотностью вероятности условного распределения* ξ при условии $\eta = y \in \mathcal{R}^1$. Равенство (8.36) можно переписать в виде

$$p(x, y) = p_\eta(y) p_\xi(x|y), \quad x, y \in \mathcal{R}^1, \quad (8.37)$$

напоминающем теорему умножения вероятностей. Если в (8.35) устремим $\Delta y \rightarrow 0$ и воспользуемся определением (8.36), то получим

$$F_\xi(x|\eta=y) = P(\{\xi < x|\eta=y\}) = \int_{-\infty}^x p_\xi(t|y) dt, \quad x, y \in \mathcal{R}^1. \quad (8.38)$$

Интегрируя равенство (8.37) по переменной y , найдем в силу (8.32)

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p_{\eta}(y) p_{\xi}(x|y) dy, \quad x \in \mathcal{R}^1. \quad (8.39)$$

Это непрерывный аналог формулы полной вероятности.

Аналогичные рассуждения справедливы и для дискретного случайного вектора (ξ, η) . Пусть ξ принимает значения $x_i, i = 1, 2, \dots$, а η принимает значения $y_j, j = 1, 2, \dots$, и $p_{ij} = P(\{\xi = x_i, \eta = y_j\})$. Тогда аналогом формул (8.32) являются равенства:

$$p_i = P(\{\xi = x_i\}) = \sum_{j=1}^{\infty} p_{ij}, \quad q_j = P(\{\eta = y_j\}) = \sum_{i=1}^{\infty} p_{ij}, \quad i, j = 1, 2, \dots \quad (8.40)$$

Условные распределения вероятностей определяются равенствами:

$$\begin{aligned} p_{i|j} &= P(\{\xi = x_i | \eta = y_j\}) = \frac{P(\{\xi = x_i, \eta = y_j\})}{P(\{\eta = y_j\})} = \frac{p_{ij}}{q_j}, \\ q_{j|i} &= P(\{\eta = y_j | \xi = x_i\}) = \frac{p_{ij}}{p_i}, \quad i, j = 1, 2, \dots \end{aligned} \quad (8.41)$$

И как следствие (8.40) и (8.41)

$$p_i = \sum_{j=1}^{\infty} q_j p_{i|j}, \quad q_j = \sum_{i=1}^{\infty} p_i q_{j|i}, \quad i = 1, 2, \dots \quad (8.42)$$

1.8.5. Независимость случайных величин. В §1.4 рассмотрена конструкция, связанная со стохастическим экспериментом, повторяемым при неизменных условиях, в которой естественно возникают независимые события. Если в каждом эксперименте измеряется какая-либо случайная величина, например, координата частицы, совершающей броуновское движение, то ее значения в различных экспериментах не зависят друг от друга. Понятие независимости случайных величин является одним из важнейших в теории вероятностей.

Определение 1.8.8. Случайные величины $\xi_1, \dots, \xi_n$ называются независимыми в совокупности (или — взаимно независимыми), если для любых $x_1, \dots, x_n$ события $\{\xi_1 < x_1\}, \dots, \{\xi_n < x_n\}$ независимы в совокупности, т.е. если

$$P(\{\xi_1 < x_1\} \cap \dots \cap \{\xi_n < x_n\}) = P(\{\xi_1 < x_1\}) \dots P(\{\xi_n < x_n\}), \quad (8.43)$$

или, что то же самое,

$$F(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n), \quad x_1, \dots, x_n \in \mathcal{R}^1. \quad (8.44)$$

Для независимых случайных величин ξ и η при любых $x_1 < x_2$ и $y_1 < y_2$

$$\begin{aligned} P(\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\}) &= P(\{\xi < x_2, \eta < y_2\}) - \\ &- P(\{\xi < x_1, \eta < y_2\}) - P(\{\xi < x_2, \eta < y_1\}) + P(\{\xi < x_1, \eta < y_1\}) = \\ &= P(\{\xi < x_2\})[P(\{\eta < y_2\}) - P(\{\eta < y_1\})] - \\ &- P(\{\xi < x_1\})[P(\{\eta < y_2\}) - P(\{\eta < y_1\})] = \\ &= P(\{x_1 \leq \xi < x_2\})P(\{y_1 \leq \eta < y_2\}). \end{aligned}$$

Итак,

$$P(\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\}) = P(\{x_1 \leq \xi < x_2\})P(\{y_1 \leq \eta < y_2\}), \quad (8.45)$$

$x_1 < x_2, y_1 < y_2$ — любые из $\mathcal{R}^1$.

Аналогичный результат справедлив, разумеется, и для произвольного числа $\xi_1, \dots, \xi_n$ взаимно независимых случайных величин: $P(\cap_{i=1}^n \{x_i^{(1)} \leq \xi_i < x_i^{(2)}\}) = \prod_{i=1}^n P(\{x_i^{(1)} \leq \xi_i < x_i^{(2)}\})$.

Пусть, теперь, выполнены равенства (8.45) при любых $x_1 < x_2$ и $y_1 < y_2$. Устремив в (8.45) $x_1, y_1 \rightarrow -\infty$, получим

$$P(\{\xi < x_2, \eta < y_2\}) = P(\{\xi < x_2\})P(\{\eta < y_2\}).$$

Это равенство ввиду произвольности x_2 и y_2 означает независимость случайных величин ξ и η . Поэтому условие (8.45) необходимо и достаточно для независимости случайных величин ξ и η .

Пусть теперь ξ и η — дискретные случайные величины, $x_1 < x_2 < \dots < x_n < \dots$ — значения, принимаемые ξ , и $y_1 < y_2 < \dots < y_n < \dots$ — значения η . Тогда $P(\{x_k \leq \xi < x_{k+1}\}) = P(\{\xi = x_k\})$, $P(\{y_j \leq \eta < y_{j+1}\}) = P(\{\eta = y_j\})$, $k, j = 1, 2, \dots$. Подставляя эти выражения в (8.45), приходим к выводу, что условие

$$P(\{\xi = x_k, \eta = y_j\}) = P(\{\xi = x_k\})P(\{\eta = y_j\}), \quad k, j = 1, 2, \dots, \quad (8.46)$$

является необходимым и достаточным условием независимости дискретных случайных величин ξ и η .

Равенство (8.44) служит определением независимости случайных величин $\xi_1, \dots, \xi_n$. Однако, условие независимости $\xi_1, \dots, \xi_n$ может быть сформулировано в форме (8.44), даже если при этом не предполагается, что стоящие справа множители являются маргинальными функциями распределения. Пусть, например, для всех $x, y \in \mathcal{R}^1$ функция распределения имеет вид: $F(x, y) = F_1(x)F_2(y)$, $x, y \in \mathcal{R}^1$, причем $\lim_{x \rightarrow \infty} F_1(x) = 1$. Тогда ξ и η независимы и $F_1(x) = F_\xi(x)$, $F_2(y) = F_\eta(y)$. Действительно, $F_\eta(y) = F(\infty, y) = F_1(\infty)F_2(y) = F_2(y)$. Таким образом, $F_\eta(y) = F_2(y)$, $y \in \mathcal{R}^1$, поэтому $F_2(\infty) = 1$, а отсюда, в свою очередь, следует, что $F_1(x) = F_\xi(x)$, $x \in \mathcal{R}^1$, и ξ и η независимы в силу (8.44).

Если независимые случайные величины ξ и η имеют соответственно плотности вероятности $p_\xi(\cdot)$ и $p_\eta(\cdot)$, то вектор (ξ, η) имеет плотность

$p(x, y) = p_\xi(x)p_\eta(y)$, $x, y \in \mathcal{R}^1$. Это равенство является следствием (8.44) в точках $x, y \in \mathcal{R}^1$, в которых $p_\xi(x)$ и $p_\eta(y)$ непрерывны, в остальных точках оно является определением $p(x, y)$, $x, y \in \mathcal{R}^1$.

Важным является обратное утверждение: если плотность $p(\cdot, \cdot)$ случайного вектора (ξ, η) равна произведению плотностей координат, $p(x, y) = p_\xi(x)p_\eta(y)$, $x, y \in \mathcal{R}^1$, то ξ и η независимы, ибо в этом случае, очевидно, выполнено условие (8.44). Это свойство доставляет удобный критерий независимости непрерывных случайных величин (см. ниже, пример 1.8.11 в §1.8.6).

Пример 1.8.7. Пусть $\xi \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\eta \sim \mathcal{N}(\mu_2, \sigma_2^2)$ и независимы. Запись $\xi \sim \mathcal{N}(\mu_1, \sigma_1^2)$ обозначает, что ξ — нормальная $\mathcal{N}(\mu_1, \sigma_1^2)$ случайная величина. Плотность случайного вектора (ξ, η) равна произведению маргинальных плотностей:

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{(x - \mu_1)^2}{2\sigma_1^2} - \frac{(y - \mu_2)^2}{2\sigma_2^2} \right\}, \quad (x, y) \in \mathcal{R}^1 \times \mathcal{R}^1.$$

Это равенство — частный случай (8.34), следовательно, (ξ, η) — нормальный случайный вектор,

$$A = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}, \quad \det A = \frac{1}{\sigma_1^2\sigma_2^2}, \quad \text{см. (8.33), } r = 0.$$

1.8.6. Функции случайных величин. В этом пункте рассмотрим распределения функций случайных величин. Ради наглядности ограничимся функциями двух случайных величин. Итак, пусть на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ определены две случайные величины $\xi_1 = \xi_1(\omega)$ и $\xi_2 = \xi_2(\omega)$, $\omega \in \Omega$, и пусть $F(x_1, x_2)$, $x_1, x_2 \in \mathcal{R}^1 \times \mathcal{R}^1$, — функция распределения случайного вектора (ξ_1, ξ_2) . Рассмотрим некоторые функции случайных величин ξ_1 и ξ_2 , т.е. новые случайные величины η_1 , η_2 , связанные функциональными зависимостями с ξ_1 и ξ_2 .

$$\begin{aligned} \eta_1 &= f_1(\xi_1, \xi_2) \triangleq f_1(\xi_1(\omega), \xi_2(\omega)) = \eta_1(\omega), \\ \eta_2 &= f_2(\xi_1, \xi_2) \triangleq f_2(\xi_1(\omega), \xi_2(\omega)) = \eta_2(\omega), \omega \in \Omega. \end{aligned} \quad (8.47)$$

Здесь предполагается, что $f_1(\cdot, \cdot) : \mathcal{R}^1 \times \mathcal{R}^1 \rightarrow \mathcal{R}^1$, $f_2(\cdot, \cdot) : \mathcal{R}^1 \times \mathcal{R}^1 \rightarrow \mathcal{R}^1$ таковы, что η_1 и η_2 вновь являются случайными величинами, определенными на том же вероятностном пространстве $(\Omega, \mathcal{F}, P)$ (количество функций η может быть, разумеется, любым). Тем самым η_1 и η_2 являются сложными функциями $\omega \in \Omega$.

Основная задача, возникающая в этой ситуации, состоит в том, чтобы, зная функцию распределения $F(\cdot, \cdot)$ случайного вектора (ξ_1, ξ_2) и функции $f_1(\cdot, \cdot)$ и $f_2(\cdot, \cdot)$, найти функцию распределения $\Phi(y_1, y_2)$, $(y_1, y_2) \in \mathcal{R}^1 \times \mathcal{R}^1$, случайного вектора (η_1, η_2) . Для рассматриваемых далее дискретных и непрерывных случайных величин

указанная задача решается без труда. В самом деле,

$$\begin{aligned} \Phi(y_1, y_2) &= P(\{\eta_1 < y_1, \eta_2 < y_2\}) = \\ &= P(\{\omega \in \Omega, f_1(\xi_1(\omega), \xi_2(\omega)) < y_1, f_2(\xi_1(\omega), \xi_2(\omega)) < y_2\}), \quad (8.48) \\ &\quad (y_1, y_2) \in \mathcal{R}^1 \times \mathcal{R}^1. \end{aligned}$$

1. Если (ξ_1, ξ_2) — дискретный случайный вектор, $\xi_1 = x_{1i}$, $\xi_2 = x_{2j}$ — значения ξ_1 и ξ_2 , $P_{ij} = P(\{\xi_1 = x_{1i}, \xi_2 = x_{2j}\})$ — вероятность этих значений, $i, j = 1, 2, \dots$, то согласно (8.48)

$$\Phi(y_1, y_2) = \sum_{(i,j) \in Q(y_1, y_2)} p_{ij}, \quad (8.49)$$

где суммирование распространено на множество индексов $(i, j) \in Q(y_1, y_2) = \{(i, j) : f_1(x_{1i}, x_{2j}) < y_1, f_2(x_{1i}, x_{2j}) < y_2\}$, $y_1, y_2 \in \mathcal{R}^1$.

2. Если (ξ_1, ξ_2) — непрерывный случайный вектор, $p(\cdot, \cdot)$ — его плотность, то согласно (8.48) и (8.29)

$$\Phi(y_1, y_2) = \iint_{D(y_1, y_2)} p(x_1, x_2) dx_1 dx_2, \quad y_1, y_2 \in \mathcal{R}^1, \quad (8.50)$$

где область $D \subset \mathcal{R}^1 \times \mathcal{R}^1$ определяется условием: $D(y_1, y_2) = \{(x_1, x_2) : f_1(x_1, x_2) < y_1, f_2(x_1, x_2) < y_2\}$ (считаем, что f_1 и f_2 таковы, что D квадрируема).

Найдем в качестве примера функцию распределения суммы $\eta = \xi_1 + \xi_2$, если задана плотность распределения $p(\cdot, \cdot)$ случайного вектора (ξ_1, ξ_2) . Согласно (8.50)

$$\begin{aligned} \Phi(y) &= P(\{\eta < y\}) = P(\{\xi_1 + \xi_2 < y\}) = \iint_{x_1 + x_2 < y} p(x_1, x_2) dx_1 dx_2 = \\ &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{y-x_1} p(x_1, x_2) dx_2 = \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{y-x_2} p(x_1, x_2) dx_1, \quad y \in \mathcal{R}^1. \end{aligned} \quad (8.51)$$

Сделаем в интеграле замену переменных $x_1 + x_2 = z, x_1 = x_1$, при которой якобиан отображения $(x_1, x_2) \rightarrow (x_1, z)$ равен 1, получим, см. рис 1.14,

$$\Phi(y) = \int_{-\infty}^y dz \int_{-\infty}^{\infty} p(x_1, z - x_1) dx_1, \quad y \in \mathcal{R}^1. \quad (8.52)$$

Из формулы (8.52) следует, что случайная величина η имеет плотность вероятности

$$p_\eta(y) = d\Phi(y)/dy = \int_{-\infty}^{\infty} p(x_1, y - x_1) dx_1, \quad y \in \mathcal{R}^1. \quad (8.53)$$

Таким образом, если двумерное распределение слагаемых ξ_1 и ξ_2 имеет плотность $p(\cdot, \cdot)$, то и их сумма $\eta = \xi_1 + \xi_2$ также имеет плотность, определенную в (8.53).

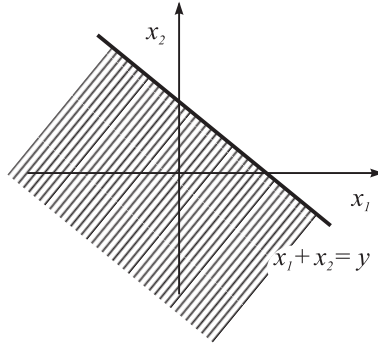


Рис. 1.14. Область интегрирования в двойном интеграле в (8.51)

Если случайные величины ξ_1 и ξ_2 независимы, то $p(x_1, x_2) = p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)$, $x_1, x_2 \in \mathcal{R}^1$, и (8.53) записывается в виде свертки функций $p_1(\cdot)$ и $p_2(\cdot)$:

$$\begin{aligned} p_\eta(y) &= \int_{-\infty}^{\infty} p_{\xi_1}(x) p_{\xi_2}(y-x) dx = \\ &= \int_{-\infty}^{\infty} p_{\xi_2}(x) p_{\xi_1}(y-x) dx = p_{\xi_1} \star p_{\xi_2}(y), \quad y \in \mathcal{R}^1. \end{aligned} \quad (8.54)$$

Докажем теперь теорему, которая будет часто использоваться в дальнейшем.

Теорема 1.8.1. Пусть ξ_1 и ξ_2 независимые случайные величины, $f_1(x)$ и $f_2(x)$, $x \in \mathcal{R}^1$, произвольные функции, такие, что $\eta_1 = f_1(\xi_1)$ и $\eta_2 = f_2(\xi_2)$ также случайные величины. Тогда η_1 и η_2 независимы, т.е. функции независимых случайных величин являются независимыми случайными величинами.

Доказательство. Приведем доказательство для дискретных случайных величин. Пусть ξ_1 принимает значения $x_1, x_2, \dots$, а ξ_2 — значения $y_1, y_2, \dots$. Тогда случайные величины $\eta_1 = f_1(\xi_1)$ и $\eta_2 = f_2(\xi_2)$ также дискретные. Пусть z_1 и z_2 — любые фиксированные значения η_1 и η_2 .

Тогда на основании (8.46) и леммы 1.3.1 о суммировании по блокам

$$\begin{aligned}
& P(\{f_1(\xi_1) = z_1, f_2(\xi_2) = z_2\}) = \\
&= \sum_{(k,l): f_1(x_k)=z_1, f_2(y_l)=z_2} P(\{\xi_1 = x_k, \xi_2 = y_l\}) = \\
&= \sum_{(k,l): f_1(x_k)=z_1, f_2(y_l)=z_2} P(\{\xi_1 = x_k\})P(\{\xi_2 = y_l\}) = \\
&= \sum_{k: f_1(x_k)=z_1} P(\{\xi_1 = x_k\}) \sum_{l: f_2(y_l)=z_2} P(\{\xi_2 = y_l\}) = \\
&= P(\{f_1(\xi_1) = z_1\})P(\{f_2(\xi_2) = z_2\}), \quad z_1, z_2 \in \mathcal{R}^1,
\end{aligned}$$

что ввиду (8.46) и произвольности z_1 и z_2 означает независимость случайных величин $\eta_1 = f_1(\xi_1)$ и $\eta_2 = f_2(\xi_2)$, см. замечание 1.3.4 к лемме о суммировании по блокам.

В заключение этого параграфа приведем несколько примеров использования полученных выше результатов.

Пример 1.8.8. Пусть $\xi = (\xi_1, \dots, \xi_n)$ и $\eta = (\eta_1, \dots, \eta_n)$ — n -мерные случайные величины, причем $\eta = A\xi$, или подробно: $\eta_i = \sum_{j=1}^n a_{ij}\xi_j$, $i = 1, \dots, n$, где $A = \|a_{ij}\|$ — невырожденная квадратная $n \times n$ матрица. Пусть $p_\xi(x) = p_\xi(x_1, \dots, x_n)$, $x = (x_1, \dots, x_n) \in \mathcal{R}^n$, — плотность распределения случайного вектора ξ , и D — произвольная квадрируемая область в n -мерном евклидовом пространстве $\mathcal{R}^n$. Тогда по правилу замены переменных в кратном интеграле и в силу (8.29) получим

$$\begin{aligned}
\int_D p_\eta(y) dy &= P(\{\eta \in D\}) = P(\{A\xi \in D\}) = P(\{\xi \in A^{-1}D\}) = \\
&= \int_{A^{-1}D} p_\xi(x) dx = \int_D p_\xi(A^{-1}y) |\det A^{-1}| dy,
\end{aligned} \tag{8.55}$$

здесь под $A^{-1}D$ понимается множество точек $z = A^{-1}y$, $y \in D$, $A^{-1}D \stackrel{\text{def}}{=} \{z = A^{-1}y, y \in \mathcal{R}^n\}$.

Отсюда ввиду произвольности области D следует, что

$$p_\eta(x) = p_\xi(A^{-1}x) |\det A^{-1}|, \quad x \in \mathcal{R}^n. \tag{8.56}$$

Пример 1.8.9. На отрезок $[0, a]$ случайно и независимо падают две частицы массы m . Найти плотность распределения координаты центра тяжести системы этих двух частиц, если координаты частиц равномерно распределены (математическая формулировка условия, согласно которому частицы падают на $[0, a]$ «случайно») на отрезке $[0, a]$.

Пусть ξ_1, ξ_2 — координаты частиц. Поскольку ξ_1 и ξ_2 равномерно распределены на $[0, a]$, то их плотности

$$p_1(x) = p_2(x) = \begin{cases} 0, & \text{если } x < 0 \text{ или } x > a, \\ 1/a, & \text{если } 0 \leq x \leq a. \end{cases} \quad (8.57)$$

Координата центра тяжести равна $\eta = (\xi_1 + \xi_2)/2$ (массы одинаковы). Поскольку случайные величины ξ_1 и ξ_2 независимы, то $F_\eta(y) = P(\{\eta < y\}) = P(\{\xi_1 + \xi_2 < 2y\}) = \int_{-\infty}^{2y} dz \int_{-\infty}^{\infty} p_1(x)p_2(z-x) dx$, $y \in \mathcal{R}^1$. Дифференцируя это выражение по y , находим плотность $p_\eta(\cdot)$

$$p_\eta(y) = 2 \int_{-\infty}^{\infty} p_1(x)p_2(2y-x) dx, \quad y \in \mathcal{R}^1. \quad (8.58)$$

Так как $p_1(x)$ отлично от нуля только при $0 \leq x \leq a$, то $p_\eta(y) = \frac{2}{a} \int_0^a p_2(2y-x) dx$, $y \in \mathcal{R}^1$. Введем здесь новую переменную $u = 2y - x$, тогда

$$p_\eta(y) = \frac{2}{a} \int_{2y-a}^{2y} p_2(u) du, \quad y \in \mathcal{R}^1. \quad (8.59)$$

Поскольку согласно (8.57) $p_2(u)$ отлично от нуля лишь при $0 \leq u \leq a$, то если в (8.59) $2y - a > a$, т.е. $y > a$, или $2y < 0$, т.е. $y < 0$, то $p_\eta(y) = 0$, $y \notin [0, a]$.

Заметим, что $\{(x_1 + x_2)/2, x_1 \in [0, a], x_2 \in [0, a]\} = [0, a]$. Пусть теперь $0 \leq y \leq a$; если $2y - a < 0$, т.е. $y < a/2$, то интегрируем в (8.59)

от 0 до $2y$, поэтому $p_\eta(y) = \frac{2}{a} \int_0^{2y} \frac{1}{a} du = \frac{4}{a^2} y$ при $0 \leq y < a/2$, если же $2y - a \geq 0$, т.е. $y \geq a/2$, то интегрируем в (8.59) от $2y - a$ до a , поэтому

$$p_\eta(y) = \frac{2}{a} \int_{2y-a}^a \frac{1}{a} du = \frac{4}{a^2} (a - y) \quad \text{при } a/2 \leq y \leq a.$$

Итак, плотность распределения $p_\eta(\cdot)$ центра тяжести равна

$$p_\eta(y) = \begin{cases} 0, & \text{если } y < 0 \text{ или } y > a, \\ \frac{4}{a^2} y, & \text{если } 0 \leq y < \frac{a}{2}, \\ \frac{4}{a^2} (a - y), & \text{если } \frac{a}{2} \leq y \leq a. \end{cases}$$

Пример 1.8.10. Пусть (ξ, η) — непрерывный случайный вектор со значениями в $\mathcal{R}^1 \times \mathcal{R}^1$ и с плотностью $p(\cdot, \cdot)$. Найдем функцию распределения произведения $\zeta = \xi\eta$. Согласно (8.50)

$$\begin{aligned} F_\zeta(z) &= P(\{\xi\eta < z\}) = \iint_{xy < z} p(x, y) dx dy = \\ &= \int_{-\infty}^0 dx \int_{z/x}^{\infty} p(x, y) dy + \int_0^{\infty} dx \int_{-\infty}^{z/x} p(x, y) dy, \quad z \in \mathcal{R}^1. \end{aligned} \quad (8.60)$$

Отсюда (формально) получаем выражение для плотности ζ

$$p_\zeta(z) = F'_\zeta(z) = - \int_{-\infty}^0 \frac{1}{x} p\left(x, \frac{z}{x}\right) dx + \int_0^{\infty} \frac{1}{x} p\left(x, \frac{z}{x}\right) dx, \quad z \in \mathcal{R}^1. \quad (8.61)$$

Пример 1.8.11. Покажем прямым вычислением, что если случайные величины ξ и η независимы и плотности распределений

$$p_\xi(x) = p_\eta(y) = \begin{cases} 0, & \text{если } x < 0, \\ e^{-x}, & \text{если } x \geq 0, \end{cases} \quad (8.62)$$

то случайные величины $\xi + \eta$ и ξ/η независимы.

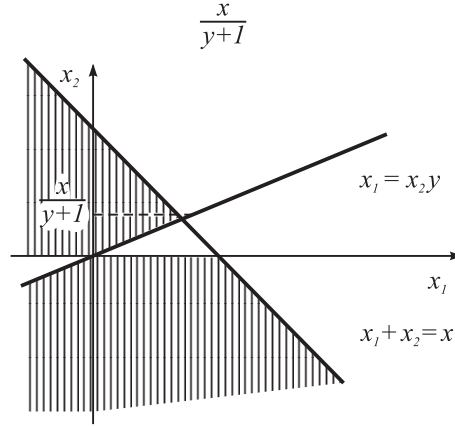
Случайные величины ξ и η независимы, поэтому плотность $p(x_1, x_2)$ случайного вектора (ξ, η) равна $p(x_1, x_2) = p_\xi(x_1)p_\eta(x_2) \equiv p_1(x_1)p_2(x_2)$, $x_1, x_2 \in \mathcal{R}^1$. Поэтому

$$P(\{\xi + \eta < x, \xi/\eta < y\}) = \iint_D p_1(x_1)p_2(x_2) dx_1 dx_2, \quad x, y \in \mathcal{R}^1, \quad (8.63)$$

где область $D = D(x, y) = \{(x_1, x_2) \in \mathcal{R}^2, x_1 + x_2 < x, x_1/x_2 < y\}$. Переходя к формуле (8.63) от двойного интеграла к повторному, найдем, см. рис. 1.15

$$\begin{aligned} P(\{\xi + \eta < x, \xi/\eta < y\}) &= \int_{-\infty}^0 p_2(x_2) dx_2 \int_{x_2 y}^{x-x_2} p_1(x_1) dx_1 + \\ &+ \int_0^{x/(y+1)} p_2(x_2) dx_2 \int_{-\infty}^{x_2 y} p_1(x_1) dx_1 + \int_{x/(y+1)}^{\infty} p_2(x_2) dx_2 \int_{-\infty}^{x-x_2} p_1(x_1) dx_1. \end{aligned}$$

Ввиду (8.62) первый интеграл в правой части равен нулю, два других преобразуются следующим образом, см. рис. 1.15,

Рис. 1.15. Область $D = D(x, y)$.

$$\begin{aligned}
 P(\{\xi + \eta < x, \frac{\xi}{\eta} < y\}) &= \int_0^{x/(y+1)} e^{-x_2} dx_2 \int_0^{x_2 y} e^{-x_1} dx_1 + \\
 &+ \int_{x/(y+1)}^x e^{-x_2} dx_2 \int_0^{x-x_1} e^{-x_1} dx_1 = \\
 &= \int_0^x e^{-x_2} dx_2 - \int_0^{x/(y+1)} e^{-x_2(1+y)} dx_2 - x e^{-x} (1 - \frac{1}{1+y}) = \\
 &= 1 - e^{-x} + \frac{1}{y+1} (e^{-x} - 1) - e^{-x} \frac{xy}{y+1} = \frac{y}{y+1} (1 - e^{-x} - e^{-x} x).
 \end{aligned} \tag{8.64}$$

С другой стороны, согласно (8.52) и (8.62)

$$\begin{aligned}
 P(\{\xi + \eta < x\}) &= \int_{-\infty}^x dz \int_{-\infty}^{\infty} p_1(x_1) p_2(z - x_1) dx_1 = \\
 &= \int_{-\infty}^x dz \int_0^{\infty} e^{-x_1} p_2(z - x_1) dx_1 = \int_0^x dz \int_0^z e^{-x_1} e^{-z+x_1} dx_1 = \\
 &= -x e^{-x} + 1 - e^{-x}, x \geq 0.
 \end{aligned} \tag{8.65}$$

Далее, см. рис. 1.16,

$$\begin{aligned}
 P\left(\left\{\frac{\xi}{\eta} < y\right\}\right) &= \iint_{x_1/x_2 < y} p_1(x_1)p_2(x_2)dx_1dx_2 = \\
 &= \int_{-\infty}^0 p_2(x_2)dx_2 \int_{x_2 y}^{\infty} p_1(x_1)dx_1 + \int_0^{\infty} p_2(x_2)dx_2 \int_{-\infty}^{x_2 y} p_1(x_1)dx_1.
 \end{aligned} \tag{8.66}$$

Преобразовав правую часть с учетом (8.62), найдем

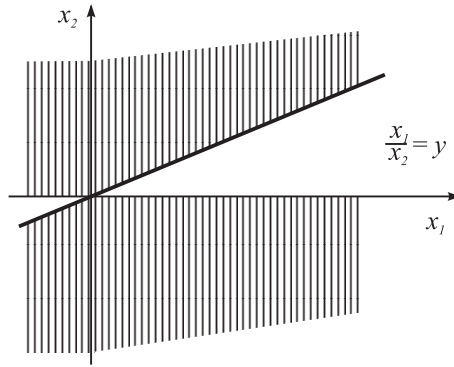


Рис. 1.16. Область интегрирования в (8.66)

$$P\left(\left\{\frac{\xi}{\eta} < y\right\}\right) = \int_0^{\infty} e^{-x_2} dx_2 \int_0^{x_2 y} e^{-x_1} dx_1 = \int_0^{\infty} e^{-x_2} (1 - e^{-x_2 y}) dx_2 = \frac{y}{y+1}. \tag{8.67}$$

Сопоставив (8.64), (8.65), (8.67), получим

$$P\left(\left\{\xi + \eta < x, \frac{\xi}{\eta} < y\right\}\right) = P(\{\xi + \eta < x\})P\left(\left\{\frac{\xi}{\eta} < y\right\}\right), \quad x, y \in \mathcal{R}^1,$$

так что случайные величины $\xi + \eta$ и ξ/η независимы.

Пример 1.8.12. Докажем, что линейная функция нормальных (гауссовых) случайных величин есть нормальная случайная величина. Очевидно, достаточно рассмотреть линейную функцию двух случайных величин. Пусть

$$\zeta = \alpha\xi + \beta\eta + \gamma, \quad |\alpha| + |\beta| > 0, \tag{8.68}$$

где случайный вектор (ξ, η) распределен нормально, α, β, γ — константы. Надо доказать, что ζ имеет нормальное распределение. Заметим, прежде всего, что если $\xi \sim \mathcal{N}(0, 1)$, то $(\sigma\xi + a) \sim \mathcal{N}(a, \sigma^2)$. В самом

деле (пусть для определенности $\sigma > 0$), для $x \in \mathcal{R}^1$

$$F_{\sigma\xi+a}(x) = P(\{\sigma\xi + a < x\}) = P(\{\xi < \frac{x-a}{\sigma}\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a)/\sigma} e^{-\frac{y^2}{2}} dy.$$

Отсюда

$$p_{\sigma\xi+a}(x) = F'_{\sigma\xi+a}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathcal{R}^1,$$

т.е.

$$(\sigma\xi + a) \sim \mathcal{N}(a, \sigma^2).$$

Ввиду этого достаточно доказать высказанное утверждение в случае, когда $\gamma = 0$, а плотность $p(\cdot, \cdot)$ случайного вектора (ξ, η) равна

$$\zeta = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}[x_1^2 - 2rx_1x_2 + x_2^2]\right\}, \quad (x_1, x_2) \in \mathcal{R}^1 \times \mathcal{R}^1, \quad (8.69)$$

($\sigma_1^2 = D\xi = 1, \sigma_2^2 = D\eta = 1, a = b = 0$, см. (8.34)). Используя формулу (8.52) для функции распределения суммы, найдем

$$\begin{aligned} F_\zeta(x) &= P(\{\zeta < x\}) = P(\{\alpha\xi + \beta\eta < x\}) = \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} p(x_1, \frac{1}{\beta}(x - \alpha x_1)) dx_1, \quad x \in \mathcal{R}^1, \end{aligned} \quad (8.70)$$

(считаем для определенности, что $\beta \neq 0$). Отсюда и из (8.69) для плотности $p_\zeta(\cdot)$ случайной величины ζ получим выражение

$$\begin{aligned} p_\zeta(x) &= F'_\zeta(x) = \int_{-\infty}^{\infty} p(x_1, \frac{1}{\beta}(x - \alpha x_1)) dx_1 = \\ &= \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-r^2)} \times \right. \\ &\quad \times [x_1^2 - 2rx_1\frac{1}{\beta}(x - \alpha x_1) + \frac{1}{\beta^2}(x - \alpha x_1)^2]\bigg\} dx_1 = \\ &= \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}[Ax_1^2 - 2Bx_1 + C]} dx_1 = \\ &= \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{C-B^2/A}{2(1-r^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}(\sqrt{A}x_1 - B/\sqrt{A})^2} dx_1 = \\ &= \frac{1}{\sqrt{2\pi A}} e^{-\frac{1}{2(1-r^2)}(C-B^2/A)}, \end{aligned} \quad (8.71)$$

где

$$A = 1 + 2r\frac{\alpha}{\beta} + \frac{\alpha^2}{\beta^2}, \quad B = \frac{x}{\beta^2}(\alpha + r\beta), \quad C = \frac{x^2}{\beta^2}, \quad x \in \mathcal{R}^1;$$

$A > 0$, так как $|r| \leq 1$. Но $C - B^2/A = x^2(1 - r^2)/A$; на основании (8.71) заключаем, что

$$p_\zeta(x) = \frac{1}{\sqrt{2\pi}\sqrt{A}} e^{-\frac{x^2}{2A}}, \quad x \in \mathcal{R}^1, \quad (8.72)$$

и, таким образом, $\zeta \sim \mathcal{N}(0, A)$.

1.9. Числовые характеристики случайных величин

1.9.1. Моменты случайных величин. Случайная величина ξ считается, как известно, заданной, если задано пространство $\tilde{\Omega}$ ее значений, σ -алгебра $\tilde{\mathcal{F}}$ событий (подмножеств) $\tilde{\Omega}$ и вероятность $\tilde{P}$ на $\tilde{\mathcal{F}}$. В рассмотренных ранее случаях $\tilde{\mathcal{F}}$ есть либо σ -алгебра борелевских подмножеств $\tilde{\Omega} = \mathcal{R}^1$, если ξ — непрерывная случайная величина, либо σ -алгебра всех подмножеств $\tilde{\Omega} = \{x_1, x_2, \dots\}$, если ξ дискретная, $x_1, x_2, \dots$ — ее значения.

Но во многих задачах такая полная характеристика случайной величины ξ , с одной стороны, недоступна для исследователя, а с другой стороны, и необязательна, достаточно ограничиться знанием некоторых параметров распределения ξ . Такими параметрами являются моменты случайной величины.

Определение 1.9.1. Моментом (начальным) порядка k дискретной случайной величины ξ , принимающей значение x_i с вероятностью $P(\{\xi = x_i\}) = p_i, i = 1, 2, \dots$, называется число

$$M\xi^k = \sum_{i=1}^{\infty} x_i^k p_i, \quad (9.1)$$

при условии, что ряд (9.1) сходится абсолютно, т.е.

$$M|\xi^k| = \sum_{i=1}^{\infty} |x_i|^k p_i < \infty. \quad (9.2)$$

$M|\xi^k|$ — называется абсолютным моментом порядка $k = 1, 2, \dots$

Моментом (начальным) порядка k непрерывной случайной величины ξ с плотностью вероятности $p(\cdot)$ называется число

$$M\xi^k = \int_{-\infty}^{\infty} x^k p(x) dx \quad (9.3)$$

при условии, что интеграл (9.3) сходится абсолютно, т.е.

$$M|\xi^k| = \int_{-\infty}^{\infty} |x|^k p(x) dx < \infty. \quad (9.4)$$

$M|\xi^k|$ — называется *абсолютным моментом порядка* $k = 1, 2, \dots$

Привлекая интеграл Стильтьеса, можно выражения (9.1) и (9.3) объединить одной записью¹⁾

$$M\xi^k = \int_{-\infty}^{\infty} x^k dF_{\xi}(x) < \infty. \quad (9.5)$$

Если $M|\xi^k|$ не существует, то говорят, что случайная величина ξ не имеет конечного момента порядка k . По определению моменты $M\xi^k$ и $M|\xi^k|$ существуют или не существуют одновременно. Требования абсолютной сходимости гарантирует возможность произвольного порядка суммирования и, следовательно, корректность определения $M\xi^k$ (в (9.1), например, порядок суммирования определяется порядком нумерации значений ξ).

Определение 1.9.2. Момент $M\xi$ первого порядка ($k = 1$ в (9.1), (9.3)) называется *математическим ожиданием случайной величины* ξ .

Рассмотрим несколько примеров.

Пример 1.9.1. *Равномерное распределение.* Случайная величина ξ равномерно распределена в $[a, b]$ с плотностью

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases} \quad M\xi = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2} \quad \text{— середина отрезка } [a, b].$$

Пример 1.9.2. *Распределение Пуассона.* Для пуассоновской случайной величины ξ $p_k = P(\{\xi = k\}) = \frac{\lambda}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$, $\lambda > 0$,

$$M\xi = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda.$$

Пример 1.9.3. *Распределение Коши.* Так называется распределение случайной величины $\eta = \xi_1/\xi_2$, где $\xi_1, \xi_2 \sim \mathcal{N}(0, 1)$ и независимы. Пользуясь формулой (8.66) с $p_1(x) = p_2(x) = \frac{1}{2\pi} e^{-x^2/2}$, $x \in \mathcal{R}^1$, без труда найдем $F_{\eta} = \frac{1}{2} + \frac{1}{\pi} \arctg x$, так что $p_{\eta}(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathcal{R}^1$.

Поэтому $M\eta$ не существует, ибо $\int_{-\infty}^{\infty} \frac{|x|dx}{1+x^2} = \infty$.

Пример 1.9.4. *Нормальное распределение*, $\xi \sim \mathcal{N}(\mu, \sigma^2)$. $M\xi =$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt +$$

¹⁾ Мы далее не пользуемся свойствами интеграла Стильтьеса, и равенство (9.5) означает лишь краткую запись для обоих выражений (9.1) и (9.3).

$+ \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \mu$, так как первый интеграл в правой части равен нулю.

Пример 1.9.5. Некто стоит перед дверью своей квартиры и пытается открыть ее, перебирая ключи из связки и не исключая испробованный ключ при дальнейших попытках. Спрашивается, сколько в среднем придется сделать попыток открывающему дверь, прежде чем он попадет домой?

Пусть ξ — число попыток, понадобившихся для открывания двери, $\xi = 1, 2, \dots$. Если в связке n ключей, из которых один — от данной двери, то событие $\xi = k$ состоит в том, что нужный ключ попадется k -м, а перед этим $k - 1$ раз выбирался неподходящий ключ. Вероятность этого события $p_k = \overline{P}(\{\xi = k\}) = \left(\frac{n-1}{n}\right)^{k-1} \frac{1}{n}$, и, следовательно, $M\xi = \sum_{k=1}^{\infty} k \left(\frac{n-1}{n}\right)^{k-1} \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{n} \frac{d}{dq} \sum_{k=1}^{\infty} q^k = \frac{1}{n} \frac{d}{dq} \left(\frac{q}{1-q}\right) \Big|_{q=\frac{n-1}{n}} = \frac{1}{n} \cdot n^2 = n$.

Предположим теперь, что испробованный ключ устраняется из дальнейших попыток. В этом случае $\xi = 1, 2, \dots, n$ и

$$\overline{P}(\{\xi = k\}) = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-(k-1)}{n-(k-2)} \cdot \frac{1}{n-(k-1)} = \frac{1}{n},$$

$k = 1, \dots, n$.

Поэтому $M\xi = \sum_{k=1}^n \frac{1}{n} k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$.

Пример 1.9.6. *Петербургская игра.* Играющий бросает монету до первого выпадения герба, после чего игра прекращается и бросающий монету получает 2^k коп., где k — число бросаний монеты. Вероятность того, что герб рано или поздно появится, равна единице. Следовательно, с вероятностью единица игрок получит некоторую сумму денег. Для того, чтобы указать осмысленную цену за участие в игре, следует посчитать математическое ожидание выигрыша (средний выигрыш). Обозначим выигрыш ξ . При этом $\overline{\Omega} = \{2, 2^2, \dots, 2^k, \dots\}$, и, очевидно, $p_k = \overline{P}(\{\xi = 2^k\}) = 2^{-k}$, $k = 1, 2, \dots$. Следовательно,

$$M\xi = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + \dots = \infty.$$

Таким образом, играя неопределенно долго, можно рассчитывать получить сколь угодно большую сумму денег. Поэтому априори невозможно указать сумму, которая была бы адекватной платой за участие в игре, цена игры должна быть бесконечно большой.

Однако если речь идет о реальном участии в игре, то следует ограничить максимальный выигрыш. Будем считать, что максимальный

выигрыш определен 2^{25} коп $\approx 335 \cdot 10^3$ руб. Это означает, что такая сумма будет получена, если число бросаний $n \geq 25$. Итак, в этом случае математическое ожидание выигрыша (и цена участия в игре).

$$\begin{aligned} M\xi &= 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + \dots + 2^{24} \cdot \frac{1}{2^{24}} + 2^{25} \left(\frac{1}{2^{25}} + \frac{1}{2^{26}} + \dots \right) = \\ &= 24 + \left(1 + \frac{1}{2} + \dots \right) = 26 \text{ коп.} \end{aligned}$$

Таким образом, если считать реальным, скажем, 2^{20} бросаний, что при условии затраты 3 секунд на бросание занимает свыше месяца непрерывной игры, то средний выигрыш все-таки поразительно мал (< 26 коп.). Этот факт, известный как *петербургский парадокс*, рассмотрен Даниилом Бернулли. Отметим, что испытания Бернулли названы в честь Якова Бернулли.

Переходим теперь к доказательству важной теоремы о математическом ожидании функции от случайной величины.

Теорема 1.9.1. Пусть ξ — дискретная случайная величина (непрерывная случайная величина), принимающая значения $x_1, x_2, \dots$ соответственно с вероятностями $p_1, p_2, \dots$ (имеющая плотность $p(\cdot)$), $f(\cdot)$ — некоторая функция и $\eta = f(\xi)$ — новая случайная величина. Тогда математическое ожидание η существует, если и только если $\sum_{i=1}^{\infty} |f(x_i)|p_i < \infty$ ($\int_{-\infty}^{\infty} |f(x)|p(x)dx < \infty$) и равно

$$M\eta = \sum_{i=1}^{\infty} f(x_i)p_i \quad (M\eta = \int_{-\infty}^{\infty} f(x)p(x)dx). \quad (9.6)$$

Доказательство проведем для дискретной случайной величины ξ . В этом случае случайная величина $\eta = f(\xi)$ также дискретная, ее значениями являются числа $y_1, y_2, \dots$, где множество $y_1, y_2, \dots$ совпадает с множеством всех различных чисел среди $f(x_1), f(x_2), \dots$, а вероятность каждого значения y_s равна

$$q_s = \bar{P}(\eta = y_s) = P(\{\omega \in \Omega, f(\xi(\omega)) = y_s\}) = \sum_{k: f(x_k)=y_s} p_k, \quad s = 1, 2, \dots \quad (9.7)$$

Поэтому согласно определению 1.9.1

$$\begin{aligned} M\eta &= \sum_{s=1}^{\infty} y_s q_s = \sum_{s=1}^{\infty} y_s \sum_{k: f(x_k)=y_s} p_k = \\ &= \sum_{s=1}^{\infty} \left(\sum_{k: f(x_k)=y_s} f(x_k) p_k \right) = \sum_{i=1}^{\infty} f(x_i) p_i, \end{aligned} \quad (9.8)$$

Последнее равенство выполняется ввиду того, что каждое слагаемое $f(x_i)p_i$ участвует в двух последних суммах один и только один раз, поскольку все y_s различны, возможность объединения в одну сумму следует из леммы о суммировании по блокам (§ 3), так как ряд $\sum_{i=1}^{\infty} |f(x_i)|p_i$ по условию сходится. Поскольку как в (9.8) $M|\eta| = \sum_{s=1}^{\infty} |y_s|q_s = \sum_{i=1}^{\infty} |f(x_i)p_i|$, то согласно замечанию 1.3.3 к лемме о суммировании по блокам условие $\sum_{i=1}^{\infty} |f(x_i)p_i| < \infty$ эквивалентно существованию $M\eta$. ■

Следствие 1.9.1. Для любой случайной величины ξ и постоянных $c_k, k = 1, 2, \dots, n$

$$M\left(\sum_{k=0}^n c_k \xi^k\right) = \sum_{k=0}^n c_k M\xi^k, \quad (9.9)$$

если $M|\xi^k| < \infty, k = 1, 2, \dots, n$.

Равенство (9.9) вытекает из теоремы 1.9.1, в которой $f(x) = \sum_{k=0}^n c_k x^k, x \in \mathcal{R}^1$, — полином¹⁾. Из (9.9), в частности, следует, что

$$Mc = c, \quad M(c\xi) = cM\xi, \quad (9.10)$$

где c — любая постоянная (в последнем равенстве предполагается, что $M|\xi| < \infty$).

Замечание 1.9.1. В случае, когда ξ — дискретная случайная величина, принимающая значения $x_1, x_2, \dots$ с вероятностями $p_1, p_2, \dots$, определена на дискретном вероятностном пространстве $(\Omega, \mathcal{F}, P)$, нетрудно дать другое выражение для ее математического ожидания: $M\xi = \sum_{\omega_i \in \Omega} \xi(\omega_i)P(\{\omega_i\})$, если $\sum_{\omega_i \in \Omega} |\xi(\omega_i)|P(\{\omega_i\}) < \infty$. Здесь $P(\{\omega_i\})$ — вероятность элементарного события $\{\omega_i\}$, и сумма распространена на все элементарные события $\omega_i \in \Omega, i = 1, 2, \dots$.

Действительно,

$$\begin{aligned} \sum_{\omega_i \in \Omega} \xi(\omega_i)P(\{\omega_i\}) &= \sum_{k=1}^{\infty} \left(\sum_{i: \xi(\omega_i)=x_k} \xi(\omega_i)P(\{\omega_i\}) \right) = \\ &= \sum_{k=1}^{\infty} x_k \sum_{i: \xi(\omega_i)=x_k} P(\{\omega_i\}) = \sum_{k=1}^{\infty} x_k p_k, \end{aligned} \quad (9.11)$$

если последний ряд сходится абсолютно, ибо тогда применима лемма о суммировании по блокам (см. §1.3). Итак, в случае существования

¹⁾ Мы не останавливаемся на доказательстве того, что непрерывная функция от случайной величины является случайной величиной.

математического ожидания $M\xi$ имеем равенство

$$M\xi = \sum_{\omega_i \in \Omega} \xi(\omega_i)P(\{\omega_i\}). \quad (9.12)$$

Аналог формулы (9.12) справедлив и в общем случае любой случайной величины, но вместо суммы в (9.12) должен быть интеграл Лебега.

Продолжим изучение моментов случайных величин.

1.9.2. Центральные моменты.

Определение 1.9.3. Математическое ожидание

$$M(\xi - M\xi)^k \quad (9.13)$$

называется k -м *центральным моментом*, если существует $M|\xi - M\xi|^k$. Центральный момент второго порядка называется *дисперсией* случайной величины ξ и обозначается $D\xi$

$$D\xi = M(\xi - M\xi)^2. \quad (9.14)$$

Этот момент является удобной характеристикой разброса значений ξ около ее математического ожидания $M\xi$. Так как в согласии со свойством (9.9) $M(\xi - M\xi)^2 = M(\xi - 2\xi M\xi + (M\xi)^2) = M\xi^2 - 2M\xi M\xi + (M\xi)^2$, то справедлива следующая формула для дисперсии:

$$D\xi = M\xi^2 - (M\xi)^2. \quad (9.15)$$

Отсюда следует, в частности, что $M\xi^2 \geq (M\xi)^2$, поскольку $D\xi \geq 0$, а так как $D|\xi| = M\xi^2 - (M|\xi|)^2$, то $M\xi^2 < \infty \Rightarrow M|\xi| < \infty$, т.е. $M\xi$ существует, если $M\xi^2 < \infty$.

Дисперсия имеет размерность квадрата случайной величины; для характеристики разброса (рассеивания) иногда бывает удобнее пользоваться значением, размерность которого совпадает с размерностью случайной величины. Такая величина

$$\sigma\xi = \sqrt{D\xi} \quad (9.16)$$

называется *среднеквадратичным уклонением* ξ .

На основании определения (9.14) и теоремы 1.9.1 можно записать:

$$D\xi = \sum_{k=1}^{\infty} (x_k - M\xi)^2 p_k, \quad (9.17)$$

если ξ — дискретная случайная величина, $x_1, x_2, \dots$ — ее значения, $p_1, p_2, \dots$ — соответствующие вероятности;

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2 p(x) dx, \quad (9.18)$$

если ξ — непрерывная случайная величина и $p(\cdot)$ — ее плотность.

Рассмотрим несколько примеров вычисления дисперсии.

Пример 1.9.7. Нормальное распределение, $\xi \sim \mathcal{N}(\mu, \sigma^2)$. Так как $M\xi = \mu$, см. пример 1.9.4, то в силу (9.18)

$$\begin{aligned} D\xi &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \\ &= -\frac{\sigma^2}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sigma^2. \end{aligned}$$

Таким образом, параметры нормального распределения $\mathcal{N}(\mu, \sigma^2)$: μ — математическое ожидание, σ^2 — дисперсия. Нормальное распределение полностью определяется этими двумя моментами.

Пример 1.9.8. Распределение Пуассона, $\xi = k$, $\tilde{P}(\{\xi = k\}) = \lambda^k e^{-\lambda}/k!$, $\lambda > 0$, $k = 0, 1, 2, \dots$. В примере 1.9.2 было показано, что $M\xi = \lambda$. Ввиду (9.15)

$$\begin{aligned} D\xi &= \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} - \lambda^2 = e^{-\lambda} \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} + e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} - \lambda^2 = \\ &= e^{-\lambda} \lambda^2 \frac{d^2}{d\lambda^2} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) + e^{-\lambda} \lambda \frac{d}{d\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) - \lambda^2 = \\ &= e^{-\lambda} (\lambda^2 e^{\lambda} + \lambda e^{\lambda}) - \lambda^2 = \lambda. \end{aligned}$$

Таким образом, единственный параметр распределения Пуассона задает как математическое ожидание, так и дисперсию.

Пример 1.9.9. Математическое ожидание и дисперсия случайной величины ξ смешанного типа (комбинации дискретного и непрерывного распределений, см. §1.8):

$$\begin{aligned} M\xi &= q_1 \sum_{k=1}^{\infty} x_k p_k + q_2 \int_{-\infty}^{\infty} x p(x) dx, \\ D\xi &= q_1 \sum_{k=1}^{\infty} (x_k - M\xi)^2 p_k + q_2 \int_{-\infty}^{\infty} (x - M\xi)^2 p(x) dx, \quad (9.19) \\ q_1, q_2 &\geq 0, q_1 + q_2 = 1. \end{aligned}$$

1.9.3. Свойства математического ожидания и дисперсии.

Лемма 1.9.1. Математическое ожидание суммы случайных величин равно сумме математических ожиданий

$$M(\xi + \eta) = M\xi + M\eta, \quad (9.20)$$

при условии, что $M|\xi|$ и $M|\eta|$ конечны.

Доказательство. Пусть (ξ, η) — дискретная случайная величина, $p_{i,j} = P(\{\xi = x_i, \eta = y_j\})$, $i, j = 1, 2, \dots$, — ее распределение. Пусть ξ

принимает значения $x_1, x_2, \dots$, а η — значения $y_1, y_2, \dots$. Тогда $\xi + \eta$ принимает значения $z_1, z_2, \dots$, где все z_s различны, $z_s = x_i + y_j$, а

$$r_s = \bar{P}(\{\xi + \eta = z_s\}) = \sum_{i,j: x_i + y_j = z_s} p_{i,j}, \quad s = 1, 2, \dots \quad (9.21)$$

Поэтому

$$\begin{aligned} M(\xi + \eta) &= \sum_{s=1}^{\infty} z_s r_s = \sum_{s=1}^{\infty} z_s \left(\sum_{i,j: x_i + y_j = z_s} p_{i,j} \right) = \\ &= \sum_{s=1}^{\infty} \left(\sum_{i,j: x_i + y_j = z_s} (x_i + y_j) p_{i,j} \right) = \sum_{i,j=1}^{\infty} (x_i + y_j) p_{i,j} = \\ &= \sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} p_{i,j} + \sum_{j=1}^{\infty} y_j \sum_{i=1}^{\infty} p_{i,j} = \sum_{i=1}^{\infty} x_i p_i + \sum_{j=1}^{\infty} y_j q_j = M\xi + M\eta. \end{aligned}$$

Здесь $\sum_{j=1}^{\infty} p_{i,j} = \bar{P}(\{\xi = x_i\}) = p_i$ и $\sum_{i=1}^{\infty} p_{i,j} = \bar{P}(\{\eta = y_j\}) = q_j$, $i, j = 1, 2, \dots$; группировать ряды можно в силу леммы и обратной леммы о суммировании по блокам, поскольку ряды по условию сходятся абсолютно.

Эти выкладки повторяют доказательства теоремы 1.9.1 для рассматриваемого случая. Но теорему 1.9.1 можно применить непосредственно, заметив, что $|\xi + \eta| \leq |\xi| + |\eta|$ и, следовательно, $M|\xi + \eta| \leq M|\xi| + M|\eta| < \infty$, см. (9.24). Согласно теореме 1.9.1 $M(\xi + \eta) = \sum_{i,j} (x_i + y_j) p_{i,j} = \sum_{i,j} x_i p_{i,j} + \sum_{i,j} y_j p_{i,j} = \sum_i x_i p_i + \sum_j y_j q_j = M\xi + M\eta$.

Пусть (ξ, η) — непрерывная случайная величина, $p(\cdot, \cdot)$ — ее плотность. Согласно (8.53) плотность суммы $\zeta = \xi + \eta$ имеет вид $p_\zeta(z) = \int_{-\infty}^{\infty} p(x, z - x) dx$. Поэтому

$$\begin{aligned} M(\xi + \eta) &= \int_{-\infty}^{\infty} z p_\zeta(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z p(x, z - x) dx dz = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) p(x, y) dx dy = \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} p(x, y) dy + \\ &+ \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} p(x, y) dx = \int_{-\infty}^{\infty} x p_\xi(x) dx + \int_{-\infty}^{\infty} y p_\eta(y) dy = M\xi + M\eta, \end{aligned}$$

порядок интегрирования можно изменить в силу их абсолютной сходимости. Свойство 1 доказано. ■

Следствие 1.9.2. Из следствия 1.9.1 и леммы 1.9.1 получим *свойство линейности математического ожидания*:

$$M(c_1\xi + c_2\eta) = c_1M\xi + c_2M\eta, \quad (9.22)$$

c_1, c_2 — любые постоянные, если $M|\xi|$ и $M|\eta|$ конечны.

Замечание 1.9.2. Если воспользоваться формулой (9.12), то равенства (9.20) и (9.22) можно доказать еще проще:

$$\begin{aligned} M(c_1\xi + c_2\eta) &= \sum_{\omega_i \in \Omega} (c_1\xi(\omega_i) + c_2\eta(\omega_i))P(\{\omega_i\}) = \\ &= c_1 \sum_{\omega_i \in \Omega} \xi(\omega_i)P(\{\omega_i\}) + c_2 \sum_{\omega_i \in \Omega} \eta(\omega_i)P(\{\omega_i\}) = c_1M\xi + c_2M\eta. \end{aligned}$$

Лемма 1.9.2. Если случайные величины ξ и η независимы, то

$$M\xi\eta = M\xi M\eta, \quad (9.23)$$

при условии, что $M|\xi|$ и $M|\eta|$ конечны.

Доказательство. Пусть $\xi\eta$ — дискретная случайная величина со значениями $t_1, t_2, \dots$, где все t_s , $s = 1, 2, \dots$, различны, $t_s = x_i y_j$ и $r_s = \tilde{P}(\{\xi\eta = t_s\}) = \sum_{i,j: x_i y_j = t_s} p_{ij}$, $s = 1, 2, \dots$, причем $p_{ij} = \tilde{P}(\{\xi = x_i, \eta = y_j\}) = p_i q_j$, так как ξ и η независимы. Поэтому $M\xi\eta = \sum_{s=1}^{\infty} t_s r_s = \sum_{s=1}^{\infty} t_s \left(\sum_{i,j: x_i y_j = t_s} p_i q_j \right) = \sum_{s=1}^{\infty} \left\{ \sum_{i,j: x_i y_j = t_s} x_i y_j p_i q_j \right\} = \sum_{i,j=1}^{\infty} x_i y_j p_i q_j = \sum_{i=1}^{\infty} x_i p_i \sum_{j=1}^{\infty} y_j q_j = M\xi \cdot M\eta$. Группировка рядов законна в силу их абсолютной сходимости и леммы о суммировании по блокам.

Если $\xi\eta$ — непрерывная случайная величина, то согласно формуле (8.61) плотность произведения $\zeta = \xi\eta$ имеет следующий вид (с учетом того, что $p(x, y) = p_i(x)p_j(y)$, $x, y \in \mathcal{R}^1$, в силу независимости ξ и η): $p_\zeta(z) = - \int_{-\infty}^0 \frac{1}{x} p_\xi(x) p_\eta\left(\frac{z}{x}\right) dx + \int_0^{\infty} \frac{1}{x} p_\xi(x) p_\eta\left(\frac{z}{x}\right) dx$. Отсюда $M\xi\eta = \int_{-\infty}^{\infty} z p_\zeta(z) dz = - \int_{-\infty}^0 \frac{1}{x} p_\xi(x) dx \int_{-\infty}^{\infty} z p_\eta\left(\frac{z}{x}\right) dz + \int_0^{\infty} \frac{1}{x} p_\xi(x) dx \int_{-\infty}^{\infty} z p_\eta\left(\frac{z}{x}\right) dz = \int_{-\infty}^0 \frac{1}{x} p_\xi(x) dx \int_{-\infty}^{\infty} t x^2 p_\eta(t) dt + \int_0^{\infty} \frac{1}{x} p_\xi(x) dx \int_{-\infty}^{\infty} t x^2 p_\eta(t) dt = \int_{-\infty}^{\infty} x p_\xi(x) dx \int_{-\infty}^{\infty} t p_\eta(t) dt = M\xi M\eta$. Перестановки интегралов законны в силу их абсолютной сходимости.

Рассмотрим некоторые неравенства.

1. Если с вероятностью единица $\xi \geq \eta$, то

$$M\xi \geq M\eta. \quad (9.24)$$

В самом деле, случайная величина $\zeta = \xi - \eta$ с вероятностью единица принимает неотрицательные значения, поэтому согласно определению 1.9.1 (с $k = 1$) справедливо (9.24).

2. *Неравенство Коши-Буняковского:*

$$M|\xi\eta| \leq \sqrt{M\xi^2 M\eta^2}, \quad (9.25)$$

если математические ожидания в правой части конечны.

Действительно, $|\xi\eta| \leq 1/2(\xi^2 + \eta^2)$. Отсюда из (9.24) следует, что если $M\xi^2$ и $M\eta^2$ конечны, то конечно и $M|\xi\eta|$. Далее, при любом λ $0 \leq M(\lambda|\xi| + |\eta|)^2 = \lambda^2 M\xi^2 + 2\lambda M|\xi\eta| + M\eta^2$. Квадратный трехчлен относительно λ неотрицателен при всех $\lambda \in (-\infty, \infty)$, стало быть, его дискриминант неположителен, т.е. $(M|\xi\eta|)^2 - M\xi^2 \cdot M\eta^2 \leq 0$, что эквивалентно (9.25). ■

3. *Неравенство Чебышева.* Для любого $\varepsilon > 0$ выполнено неравенство

$$P(\{|\xi| > \varepsilon\}) \leq M|\xi|^2/\varepsilon^2, \quad (9.26)$$

если $M|\xi|^2$ конечно.

Действительно, определим случайную величину η по формуле

$$\eta = \begin{cases} 0, & \text{если } |\xi| \leq \varepsilon, \\ \varepsilon, & \text{если } |\xi| > \varepsilon, \end{cases} \quad \varepsilon > 0. \quad (9.27)$$

Таким образом, η — дискретная случайная величина, принимающая два значения: 0 с вероятностью $p_1 = P(\{|\xi| \leq \varepsilon\})$ и ε с вероятностью $P(\{|\xi| > \varepsilon\})$. Из определения η следует, что $\eta^2 \leq |\xi|^2$, и в силу (9.24) имеем $M|\xi|^2 \geq M\eta^2 = \varepsilon^2 P(\{|\xi| > \varepsilon\})$, что совпадает с (9.26). ■

Взяв в (9.26) вместо ξ случайную величину $\xi - M\xi$ и учитывая, что $M(\xi - M\xi)^2 = D\xi$, запишем (9.26) в виде

$$P(\{|\xi - M\xi| > \varepsilon\}) \leq D\xi/\varepsilon^2, \quad (9.28)$$

именно это неравенство называют неравенством Чебышева.

То же рассуждение с использованием случайной величины η из (9.27) и неравенства (9.24) приводит к неравенству: для любого $\varepsilon > 0$ $P(\{|\xi| > \varepsilon\}) \leq (1/\varepsilon)M|\xi|$, если $M|\xi|$ конечно и, в частности, если случайная величина ξ неотрицательна, то

$$P(\{\xi > \varepsilon\}) \leq \frac{1}{\varepsilon}M\xi.$$

Лемма 1.9.3. Если $\eta = c\xi$, то $D\eta = c^2 D\xi$, c — любая постоянная; дисперсия постоянной равна нулю, $Dc = 0$, и наоборот, если $D\xi = 0$, то с вероятностью единица ξ равна константе, $\xi = M\xi$.

В самом деле, $Dc = M(c - M\xi)^2 = M(c - c)^2 = 0$. В силу (9.28) при любом $\varepsilon > 0$ $P(\{|\xi - M\xi| > \varepsilon\}) = 0$. Поэтому на основании полной аддитивности вероятности получим

$$P(\{|\xi - M\xi| > 0\}) = P(\{|\xi - M\xi| > 1\}) + P(\{1/2 < |\xi - M\xi| \leq 1\}) + \dots + P(\{1/2^k < |\xi - M\xi| \leq 1/2^{k-1}\}) + \dots = 0,$$

и, таким образом, $P(\{\xi - M\xi = 0\}) = 1$.

Лемма 1.9.4. Дисперсия суммы конечного числа попарно независимых случайных величин равна сумме дисперсий слагаемых

$$D\left(\sum_{i=1}^n \xi_i\right) = \sum_{i=1}^n D\xi_i. \quad (9.29)$$

В самом деле,

$$\begin{aligned} D\left(\sum_{i=1}^n \xi_i\right) &= M\left(\sum_{i=1}^n \xi_i - M\sum_{i=1}^n \xi_i\right)^2 = M\left(\sum_{i=1}^n (\xi_i - M\xi_i)\right)^2 = \\ &= M\sum_{i,k=1}^n (\xi_i - M\xi_i)(\xi_k - M\xi_k) = \sum_{i,k=1}^n M[(\xi_i - M\xi_i)(\xi_k - M\xi_k)] = \\ &= \sum_{i=1}^n M(\xi_i - M\xi_i)^2 + \sum_{i,k=1, i \neq k}^n M(\xi_i - M\xi_i)M(\xi_k - M\xi_k) = \sum_{i=1}^n D\xi_i. \end{aligned}$$

Здесь несколько раз использована лемма 1.9.1, в предпоследнем равенстве учтена независимость ξ_i и ξ_k при $i \neq k$ и лемма 1.9.3, а последнее равенство основано на том, что $M(\xi_i - M\xi_i) = M\xi_i - M\xi_i = 0$. ■

Равенство (9.29) иногда называют равенством *Бьенеме*.

Пример 1.9.10. Пусть случайная величина ξ распределена по биномиальному закону. Найдём $M\xi$ и $D\xi$. Введём случайные величины ξ_k , равные числу успехов при k -м испытании в серии из n испытаний Бернулли. Если вероятность успеха при каждом испытании равна p , то ξ_k принимает два значения 0 и 1 с вероятностями: $P(\{\xi_k = 1\}) = p$ и $P(\{\xi_k = 0\}) = q = 1 - p$, $k = 1, 2, \dots, n$. Поэтому $M\xi_k = 1 \cdot p + 0 \cdot q = p$, $D\xi_k = M\xi_k^2 - (M\xi_k)^2 = 1 \cdot p + 0 \cdot q - p^2 = p - p^2 = pq$. Число успехов ξ в серии из n испытаний равно, очевидно, сумме $\xi = \xi_1 + \xi_2 + \dots + \xi_n$. Отсюда ввиду независимости слагаемых и свойств 1) и 7) получим

$$M\xi = \sum_{k=1}^n M\xi_k = pn, \quad \sum_{k=1}^n D\xi_k = npq. \quad (9.30)$$

1.9.4. Условное математическое ожидание. Наилучший в среднеквадратичном прогноз. В §1.8.3 была определена условная функция распределения¹⁾ $F(x|B) = P(\{\xi < x|B\})$ случайной величины ξ при условии B , если $P(B) > 0$. Математическое ожидание (или среднее значение) ξ по этому условному распределению называется *условным*

¹⁾ См. замечание 1.8.1

математическим ожиданием. Таким образом

$$M(\xi|B) = \int_{-\infty}^{\infty} x dF_{\xi}(x|B), \quad (9.31)$$

или подробно

$$M(\xi|B) = \sum_{k=1}^n x_k p_k(B), \quad (9.32)$$

при условии, что $\sum_{k=1}^n |x_k| p_k(B) < \infty$, если ξ дискретная случайная величина, x_k , $k = 1, 2, \dots$, — ее значения, а $p_k(B) = \tilde{P}(\{\xi = x_k|B\})$, $k = 1, 2, \dots$, — соответствующие условные вероятности. Если ξ непрерывная случайная величина, то

$$M(\xi|B) = \int_{-\infty}^{\infty} x p(x|B) dx \quad (9.33)$$

при условии, что $\int_{-\infty}^{\infty} |x| p(x|B) dx < \infty$, где $p(x|B)$ — условная плотность вероятности (см. §1.8.4). В частности, если событие B состоит в том, что случайная величина η принимает некоторое значение y , $\eta = y$, то согласно формулам (8.36) и (8.41) получим

$$M(\xi|y_j) \equiv M(\xi|\eta = y_j) = \sum_{k=1}^n x_k P_{(\xi=k|\eta=y_j)}, \quad j = 1, 2, \dots, \quad (9.34)$$

для дискретных случайных величин, и

$$M(\xi|y) \equiv M(\xi|\eta = y) = \int_{-\infty}^{\infty} x p_{\xi}(x|y) dx, \quad y \in \mathcal{R}^1, \quad (9.35)$$

для непрерывных случайных величин.

Понятно, что так определенные условные математические ожидания обладают всеми свойствами обычных математических ожиданий, рассмотренными в §1.9.3, однако у них имеются и некоторые специфические свойства, связанные с возможностью применения к ним различных вариантов формулы полной вероятности. Так, если имеется полная группа попарно несовместных событий B_k , $k = 1, 2, \dots, n$, и $F_{\xi}(\cdot|B_k)$, $k = 1, 2, \dots, n$, — соответствующие условные функции распределения, то поскольку $F_{\xi}(x) = \sum_{k=1}^n F_{\xi}(x|B_k) P(B_k)$, $x \in \mathcal{R}^1$, немедленно получим

$$M\xi = \sum_{k=1}^n P(B_k) M(\xi|B_k). \quad (9.36)$$

Так как правая часть равенства (9.36) имеет вид математического ожидания новой дискретной случайной величины η , принимающей значение $\eta = M(\xi|B_k)$ с вероятностью $P(B_k)$, $k = 1, 2, \dots$, то естественно записать (9.36) в виде

$$M\xi = M\{M(\xi|\eta)\}. \quad (9.37)$$

Точно так же на основании формул (9.34) и (9.35) получаются равенства:

$$M\xi = \sum_{j=1}^{\infty} M(\xi|y_j)q_j, \quad q_j = \tilde{P}(\eta = y_j), \quad j = 1, 2, \dots, \quad (9.38)$$

в дискретном случае и

$$M\xi = \int_{-\infty}^{\infty} M(\xi|y)p_{\eta}(y)dy \quad (9.38^*)$$

в непрерывном.

В самом деле, в первом случае умножим равенство (9.34) на q_j и просуммируем по всем j , получим, пользуясь формулой (8.41),

$$\sum_{j=1}^{\infty} M(\xi|y_j)q_j = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} x_k p_{kj} = \sum_{k=1}^{\infty} x_k p_k = M\xi.$$

Во втором случае умножим (9.35) на $p_{\eta}(y)$ и, пользуясь равенством (8.37), проинтегрируем от $-\infty$ до ∞

$$\begin{aligned} \int_{-\infty}^{\infty} M(\xi|y)p_{\eta}(y)dy &= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} xp_{\eta}(y)p_{\xi}(x|y)dx = \\ &= \int_{-\infty}^{\infty} xdx \int_{-\infty}^{\infty} p(x,y)dy = \int_{-\infty}^{\infty} xp_{\xi}(x)dx = M\xi. \end{aligned}$$

Проведенные выкладки оправданы при условии, что ряд (9.34) или интеграл (9.35) сходятся абсолютно.

Рассмотрим теперь условные математические ожидания $M(\xi|y_j)$ или $M(\xi|y)$, определенные формулами (9.34) или (9.35), как функции аргумента y_j , $j = 1, 2, \dots$, или соответственно $y \in \mathcal{R}^1$. Этот аргумент — значение случайной величины η , и поэтому мы можем рассматривать $M(\xi|y_j)$ или $M(\xi|y)$ как значения случайной величины, зависящей от η , $M(\xi|y) = M(\xi|\eta)|_{\eta=y}$, $y \in \mathcal{R}^1$. При этом для любой случайно величины $\varphi = \varphi(\eta)$

$$M(\xi \cdot \varphi(\eta)|\eta) = \varphi(\eta) \cdot M(\xi|\eta), \quad (9.39)$$

поскольку в силу линейности математического ожидания $M(\xi \times \varphi(\eta)|\eta)|_{\eta=y} = M(\xi \cdot \varphi(y)|y) = \varphi(y)M(\xi|y)$, $y \in \mathcal{R}^1$. При таком подходе

правые части равенств (9.38) и (9.38*) определяют в силу теоремы 1.9.1 математическое ожидание функции $M(\xi|\eta)$ от случайной величины η и, следовательно, могут быть записаны в виде

$$M\xi = M\{M(\xi|\eta)\}. \quad (9.40)$$

Пример 1.9.11. Рассмотрим случайную величину χ_A , являющуюся индикатором события A :

$$\chi_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \text{ т.е. если } A \text{ происходит,} \\ 0, & \text{если } \omega \notin A, \text{ т.е. если } A \text{ не происходит,} \end{cases}$$

χ_A — дискретная случайная величина, $M\chi_A = 1 \cdot P(A) + 0 \cdot P(\bar{A}) = P(A)$.

Поэтому, $M\chi_A = P(A)$ и $M(\chi_A|\eta) = P(A|\eta)$, где $P(A|\eta)$ — условная вероятность события A — функция случайной величины η . Равенство (9.40) в этом случае выглядит так:

$$P(A) = M\{P(A|\eta)\}. \quad (9.41)$$

Формулы (9.37) и (9.40) находят многочисленные применения в теории вероятностей и в математической статистике. Условное математическое ожидание $M(\xi|\eta)$, рассматриваемое как функция η , в статистике называется *функцией регрессии* ξ на η . Если, например, $M(\xi|\eta) = a_1\eta + a_2$, то говорят о линейной регрессии ξ на η , а a_1 и a_2 называют *коэффициентами регрессии*.

В качестве примера, в котором естественно возникает конструкция условного математического ожидания, рассмотрим задачу о наилучшем в среднем квадратичном (с.к.) прогнозе (или с.к. оценивании), в которой задано совместное распределение случайных величин ξ и η , значения первой из которых наблюдаемы, а второй — нет, и требуется определить функцию $f(\xi)$ случайной величины ξ из условия $M(\eta - f(\xi))^2 \sim \min_{f(\xi)}$. Если минимум достигается на $\hat{f}(\cdot)$, то случайная величина $\hat{\eta} = \hat{f}(\xi)$ определяет наиболее точный в с.к. прогноз значения η , основанный на наблюдаемом значении ξ .

Теорема 1.9.2. Если $M\eta^2 < \infty$, то $\hat{\eta} = \hat{f}(\xi) = M(\eta|\xi)$.

Доказательство. Согласно равенствам (9.39) и (9.40)

$$\begin{aligned} M(\eta - f(\xi))^2 &= M(M((\eta - f(\xi))^2|\xi)) = \\ &= M(M((\eta - M(\eta|\xi) + M(\eta|\xi) - f(\xi))^2|\xi)) = \\ &= M(M((\eta - M(\eta|\xi))^2|\xi) + \\ &\quad + 2M((\eta - M(\eta|\xi))(M(\eta|\xi) - f(\xi))|\xi) + \\ &\quad + M((M(\eta|\xi) - f(\xi))^2|\xi)) = \\ &= M(\eta - M(\eta|\xi))^2 + M(M(\eta|\xi) - f(\xi))^2 \geq M(\eta - M(\eta|\xi))^2, \end{aligned}$$

откуда следует, что $\min_{f(\cdot)}$ достигается при $\hat{f}(\xi) = M(\eta|\xi)$. При получении четвертого равенства использована формула (9.39): $2M((\eta - M(\eta|\xi))(M(\eta|\xi) - f(\xi))|\xi) = 2(M(\eta|\xi) - f(\xi))M(\eta - M(\eta|\xi)|\xi) = 2(M(\eta|\xi) - f(\xi))(M(\eta|\xi) - M(\eta|\xi)) = 0$. Первое равенство основано на формуле (9.40), которая использована и при получении четвертого равенства: $M(M(\eta - M(\eta|\xi))^2|\xi) = M(\eta - M(\eta|\xi))^2$, $M(M(M(\eta|\xi) - f(\xi))^2|\xi) = M(M(\eta|\xi) - f(\xi))^2$. ■

Замечание 1.9.3. Если при прогнозировании значений η наблюдения невозможны, то наилучший в с.к. прогноз должен быть основан на значении const , определенном из условия $M(\eta - c)^2 \sim \min_c$. Так как $M(\eta - c)^2 = M\eta^2 - 2cM\eta + c^2$, то минимум достигается при $c = M\eta$. Согласно теореме 1.9.2 учет наблюдения ξ сводится к замене $M\eta$ на $M(\eta|\xi)$.

Если η и ξ независимы, то наблюдения ξ при прогнозировании значения η бесполезны, так как в этом случае согласно лемме 1.9.2, замечанию 1.9.3 и теореме 1.9.2

$$\begin{aligned} M(\eta - f(\xi))^2 &= M\eta^2 - 2M\eta Mf(\xi) + Mf^2(\xi) \geq \\ &\geq M\eta^2 - 2M\eta Mf(\xi) + (Mf(\xi))^2 = M(\eta - Mf(\xi))^2 \geq M(\eta - M\eta)^2 \end{aligned}$$

и при $f(\xi) = M\eta$ все неравенства обращаются в равенства. В этом случае, как и следовало ожидать, $M(\eta|\xi) = M\eta$.

Следствие 1.9.3. Теорема 1.9.2 верна и для случайных векторов ξ и η со значениями в $\mathcal{R}^m$ и $\mathcal{R}^n$, если речь идет о задаче $M\|\eta - f(\xi)\|_n^2 \sim \min_{f(\cdot)}$, где $f(\cdot)$ — функция, определенная на $\mathcal{R}^m$, принимающая значения в $\mathcal{R}^n$, $\|\eta - f(\xi)\|_n^2 = (\eta - f(\xi), \eta - f(\xi))_n$ и $\|\cdot\|_n$, $(\cdot, \cdot)_n$ — символы нормы и скалярного произведения в $\mathcal{R}^n$. И в этом случае $\min_{f(\cdot)}$ достигается на $\hat{f}(\xi) = M(\eta|\xi)$. Читателю предлагается проверить этот факт.

В заключение приведем *решение задачи о наилучшем в среднем квадратичном линейном прогнозе*, в которой требуется найти линейную функцию $\hat{f}(\xi) = \hat{a}_1\xi + \hat{a}_2$, удовлетворяющую условию

$$M(\eta - \hat{a}_1\xi - \hat{a}_2)^2 = \min_{a_1, a_2} M(\eta - a_1\xi - a_2)^2. \quad (9.42)$$

Согласно замечанию 1.9.3 искомое значение $\hat{a}_2$ должно удовлетворять условию $\hat{a}_2 = M(\eta - a_1\xi) = M\eta - a_1M\xi$, и задача отыскания $\hat{f}(\xi)$ сводится к следующей задаче на минимум:

$$\begin{aligned} M(\eta - \hat{a}_1\xi - \hat{a}_2)^2 &= \min_{a_1} M(\eta - M\eta - a_1(\xi - M\xi))^2 = \\ &= \min_{a_1} (D\eta - 2a_1 \text{cov}(\eta, \xi) + a_1^2 D\xi) = D\eta - \frac{(\text{cov}(\eta, \xi))^2}{D\xi}, \end{aligned} \quad (9.43)$$

где $\text{cov}(\eta, \xi) = M((\eta - M\eta)(\xi - M\xi))$, см. следующий §1.9.5, $\hat{a}_1 = (\text{cov}(\eta, \xi))/D\xi$ и *искомый линейный прогноз*

$$\hat{f}(\xi) = \hat{a}_1\xi + \hat{a}_2 = M\eta + \frac{\text{cov}(\eta, \xi)}{D\xi}(\xi - M\xi), \quad (9.44)$$

где правая часть называется *линейной регрессией η на ξ* . Заметим, что для получения линейного прогноза нет необходимости знать совместное распределение η, ξ , но при этом, разумеется, $M(\eta - M(\eta|\xi))^2 \leq D\eta - \frac{(\text{cov}(\eta, \xi))^2}{D\xi}$.

1.9.5. Моменты векторных случайных величин. Этот параграф посвящен изучению моментов многомерных случайных величин. Специфические свойства моментов векторных случайных величин обусловлены взаимной зависимостью координат случайного вектора.

Определение 1.9.4. Математическим ожиданием случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется вектор

$$M\xi = (M\xi_1, \dots, M\xi_n), \quad (9.45)$$

составленный из математических ожиданий его координат.

Дисперсией вектора $\xi = (\xi_1, \dots, \xi_n)$ называется вектор

$$D\xi = (D\xi_1, \dots, D\xi_n), \quad (9.46)$$

составленный из дисперсий координат.

Для многомерных случайных величин справедлив аналог теоремы 1.9.1, так что, например, если $f(x) = f(x_1, \dots, x_n)$, $x = (x_1, \dots, x_n) \in \mathcal{R}^n$, — произвольная функция такая, что $\eta = f(\xi)$ — новая случайная величина, то

$$M\eta = Mf(\xi) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \dots dx_n \quad (9.47)$$

при условии, что интеграл (9.47) сходится абсолютно (здесь $p(x_1, \dots, x_n)$ — плотность случайного вектора $\xi = (\xi_1, \dots, \xi_n)$). В частности, $M\left(\sum_{k=1}^n a_k \xi_k\right) = M(a, \xi)_n = \sum_{k=1}^n a_k M\xi_k = (a, M\xi)_n$, где $a = (a_1, \dots, a_n)$ и $(\cdot, \cdot)_n$ — символ скалярного произведения в $\mathcal{R}^n$, а если координаты ξ независимы в совокупности, то $M(\xi_1 \dots \xi_n) = M\xi_1 \dots M\xi_n$.

Важной характеристикой n -мерной случайной величины ξ является так называемая *матрица ковариаций* или *дисперсионная матрица*, матричные элементы которой суть

$$\begin{aligned} a_{ij} &= \text{cov } \xi_i \xi_j \equiv M[(\xi_i - M\xi_i)(\xi_j - M\xi_j)] = \\ &= M\xi_i \xi_j - M\xi_i M\xi_j, \quad i, j = 1, \dots, n. \end{aligned} \quad (9.48)$$

На главной диагонали в матрице ковариаций стоят дисперсии $a_{ij}|_{j=i} = D\xi_i$, $i = 1, \dots, n$, $a_{ij} = a_{ji}$ называется также *ковариационным*

моментом случайных величин ξ_i и ξ_j , $i \neq j$, $i, j = 1, \dots, n$. В силу определения (9.48), если ξ_i и ξ_j независимы, то $a_{ij} \equiv \text{cov } \xi_i \xi_j = 0$. Таким образом, условие $a_{ij} \neq 0$ является достаточным признаком зависимости ξ_i и ξ_j . Обратное утверждение неверно: из равенства нулю $\text{cov}(\xi_i \xi_j)$ не следует независимость ξ_i и ξ_j .

Поскольку при любом c $D(\xi_i + c\xi_j) = D\xi_i + c^2 D\xi_j + 2c(\text{cov } \xi_i, \xi_j) \geq 0$, то, полагая $c = -(\text{cov}(\xi_i, \xi_j))/D\xi_j$, найдем $D\xi_i - (\text{cov}(\xi_i, \xi_j))^2/D\xi_j \geq 0$, т.е. $|\text{cov}(\xi_i, \xi_j)| \leq \sqrt{D\xi_i D\xi_j}$. Отсюда следует существование матрицы ковариаций в случае, когда $D\xi_j < \infty$, $j = 1, 2, \dots, n$.

Вспомнив доказательство равенства Бьенеме (9.29), получим для любого вектора $\xi = (\xi_1, \dots, \xi_n)$ с конечной дисперсией:

$$D\left(\sum_{i=1}^n \xi_i\right) = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j) = \sum_{i=1}^n D\xi_i + \sum_{i,j=1, i \neq j}^n \text{cov}(\xi_i, \xi_j). \quad (9.49)$$

Из определения (9.48) видно, что корреляционный момент характеризует не только зависимость величин ξ_i и ξ_j , но и их рассеивание.

Как характеристику линейной связи между ξ_i и ξ_j вводят так называемый коэффициент корреляции

$$r_{ij} = \frac{\text{cov}(\xi_i, \xi_j)}{\sqrt{D\xi_i D\xi_j}}. \quad (9.50)$$

Мы видели, что $|\text{cov}(\xi_i, \xi_j)| \leq \sqrt{D\xi_i D\xi_j}$, так что $|r_{ij}| \leq 1$, $r_{ii} = 1$, $i, j = 1, \dots, n$. На основании равенства (9.49) находим, что $(\sigma_i = \sqrt{D\xi_i}, \sigma_j = \sqrt{D\xi_j})$ $D\left(\frac{\xi_i}{\sigma_i} \pm \frac{\xi_j}{\sigma_j}\right) = 2(1 \pm r_{ij}) \geq 0$, откуда следует, что $r_{ij} = \pm 1$ тогда и только тогда, когда $D(\xi_i/\sigma_i \pm \xi_j/\sigma_j) = 0$, а это в свою очередь, в силу свойства дисперсии, полученного в лемме 1.9.3, возможно лишь тогда, когда случайная величина $\eta = \xi_i/\sigma_i \pm \xi_j/\sigma_j$ равна постоянной с вероятностью единица. Таким образом, $\xi_i = \alpha\xi_j + \beta$ (линейно связаны), если и только если $r_{ij} = \pm 1$.

Если $r_{ij} > 0$, то говорят, что между ξ_i и ξ_j *положительная корреляция*, и это означает, что ξ_i и ξ_j имеют тенденцию возрастать и убывать одновременно. При $r_{ij} < 0$ ситуация обратная. Если $r_{ij} = 0$, то говорят, что случайные величины ξ_i и ξ_j *некоррелированы*, и этому свойству можно придать определенный геометрический смысл.

С этой целью рассмотрим множество случайных величин с конечным моментом второго порядка. Из $M\xi^2 < \infty$, $M\eta^2 < \infty$ следует, что при любых постоянных a и b $M(a\xi + b\eta)^2 < \infty$, поэтому множество таких случайных величин относительно операций сложения и умножения на число образует *линейное пространство* $\mathcal{L}$. Определим в $\mathcal{L}$ *квазискальное произведение*

$$(\xi, \eta) = M\xi\eta. \quad (9.51)$$

$M\xi\eta$ обладает всеми свойствами скалярного произведения, за исключением одного: из $M\xi^2 = 0$ не следует $\xi = 0$, а следует лишь, что $\xi = 0$ с вероятностью единица. Это обстоятельство не мешает, однако, дать следующую геометрическую интерпретацию: матрица ковариаций (9.48) является матрицей квазискалярных произведений случайных величин $(\xi_i - M\xi_i), i = 1, \dots, n$. При этом некоррелированность означает ортогональность в смысле квазискалярного произведения (9.51).

Заметим, что $(\xi_i - M\xi_i)$ является «истинно случайной» частью случайной величины ξ_i и матрица ковариаций характеризует связь «вероятностную» именно этих «случайных» частей. Равенство $|r_{ij}| = 1$ означает, что $(\xi_i - M\xi_i)$ и $(\xi_j - M\xi_j)$ линейно зависимы в построенном линейном пространстве $\mathcal{L}$.

Замечание 1.9.4. Вернемся к задаче (9.42) о линейном прогнозе и обозначим $\tilde{\eta} = \eta - M\eta$, $\tilde{\xi} = \xi - M\xi$. Тогда согласно (9.44), (9.51) $\hat{\eta} = \hat{f}(\xi) - M\eta = \frac{(\tilde{\eta}, \tilde{\xi})}{\|\tilde{\xi}\|^2} \tilde{\xi}$ — ортогональная проекция $\tilde{\eta}$ на линейное подпространство $\mathcal{L}_{\tilde{\xi}} = \{a\tilde{\xi}, -\infty < a < \infty\} \subset \mathcal{L}$, а в (9.43) $D\eta - \frac{(\text{cov}(\eta, \xi))^2}{D\xi} = \|\tilde{\eta}\|^2 - \frac{(\tilde{\eta}, \tilde{\xi})^2}{\|\tilde{\xi}\|^2}$ — квадрат длины составляющей $\tilde{\eta}$, принадлежащей $\mathcal{L}_{\tilde{\eta}}^\perp$.

В заключение рассмотрим задачу наилучшего в среднем квадратичном линейного прогнозирования для векторных случайных величин, в которой требуется найти $m \times n$ матрицу R_* и вектор-столбец r_* $m \times 1$, удовлетворяющие условию

$$M\|\eta - R_*\xi - r_*\|_m^2 = \min_{R, r} M\|\eta - R\xi - r\|_m^2, \quad (9.52)$$

в котором $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix}$ и $\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}$ — пара случайных векторов, охарактеризованных математическими ожиданиями $M\eta$ и $M\xi$ и ковариационными матрицами

$$\begin{aligned} S_{\xi\xi} &= M(\xi - M\xi)(\xi - M\xi)^* = \\ &= M \begin{pmatrix} \xi_1 - M\xi_1 \\ \vdots \\ \xi_n - M\xi_n \end{pmatrix} (\xi_1 - M\xi_1 \dots \xi_n - M\xi_n) = \\ &= \begin{pmatrix} \text{cov}(\xi_1, \xi_1) & \dots & \text{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(\xi_n, \xi_1) & \dots & \text{cov}(\xi_n, \xi_n) \end{pmatrix}, \end{aligned} \quad (9.53)$$

$$S_{\eta\xi} = M(\eta - M\eta)(\xi - M\xi)^* = \begin{pmatrix} \text{cov}(\eta_1, \xi_1) & \dots & \text{cov}(\eta_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(\eta_m, \xi_1) & \dots & \text{cov}(\eta_m, \xi_n) \end{pmatrix},$$

$$S_{\eta\eta} = M(\eta - M\eta)(\eta - M\eta)^* = \begin{pmatrix} \text{cov}(\eta_1, \eta_1) & \dots & \text{cov}(\eta_1, \eta_m) \\ \vdots & \ddots & \vdots \\ \text{cov}(\eta_m, \eta_1) & \dots & \text{cov}(\eta_m, \eta_m) \end{pmatrix}.$$

Прежде чем приступить к решению задачи (9.52), заметим, что для

$$\text{любого случайного вектора } \gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix}$$

$$\|\gamma\|_k^2 \stackrel{\text{def}}{=} \sum_{i=1}^k \gamma_i^2 = \text{tr}(\gamma\gamma^*) \equiv \text{tr} \left(\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} (\gamma_1, \dots, \gamma_k) \right) = \text{tr}(\gamma^*\gamma) \equiv$$

$$= \text{tr} \left((\gamma_1, \dots, \gamma_k) \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} \right), \quad M\|\gamma - M\gamma\|_k^2 = \text{tr } S_{\gamma\gamma}.$$

(9.54)

Заметим также, что $\min_c M\|\gamma - c\|_k^2 = \text{tr } S_{\gamma\gamma}$, где $c = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$ —

неслучайный вектор, и достигается при $c = M\gamma$, ибо $M\|\gamma - c\|_k^2 = M\|\gamma - M\gamma + M\gamma - c\|_k^2 = M\|\gamma - M\gamma\|_k^2 + 2M(\gamma - M\gamma, M\gamma - c)_k + \|M\gamma - c\|_k^2 = \text{tr } S_{\gamma\gamma} + \|M\gamma - c\|_k^2 \geq \text{tr } S_{\gamma\gamma}$, где $M(\gamma - M\gamma, M\gamma - c)_k = (M\gamma - M\gamma, M\gamma - c)_k = 0$. Согласно последнему замечанию $\min_r M\|\eta - R\xi - r\|_m^2 = M\|\eta - R\xi - M\eta + RM\xi\|_m^2 = M\|\eta - M\eta - R(\xi - M\xi)\|_m^2 = M\|\tilde{\eta} - R\tilde{\xi}\|_m^2$, где $\tilde{\eta} = \eta - M\eta$, $\tilde{\xi} = \xi - M\xi$. Поэтому задача (9.52) сведена к задаче

$$M\|\tilde{\eta} - R_*\tilde{\xi}\|_m^2 = \min_R M\|\tilde{\eta} - R\tilde{\xi}\|_m^2, \quad (9.55)$$

в которой $M\|\tilde{\eta} - R\tilde{\xi}\|_m^2 = M \text{tr}((\tilde{\eta} - R\tilde{\xi})(\tilde{\eta} - R\tilde{\xi})^*) = M \text{tr}(\tilde{\eta}\tilde{\eta}^* - R\tilde{\xi}\tilde{\eta}^* - \tilde{\eta}\tilde{\xi}^*R^* + R\tilde{\xi}\tilde{\xi}^*R^*) = \text{tr}(S_{\eta\eta} - RS_{\xi\eta} - S_{\eta\xi}R^* + RS_{\xi\xi}R^*) = \text{tr}(S_{\eta\eta} - 2S_{\eta\xi}R^* + RS_{\xi\xi}R^*)$, где использованы тождества $(R\tilde{\xi})^* = \tilde{\xi}^*R^*$, $\text{tr } RS_{\xi\eta} = \text{tr}(RS_{\xi\eta})^* = \text{tr } S_{\eta\xi}R^*$ и $S_{\xi\eta}^* = S_{\eta\xi}$. Пусть матрица $S_{\xi\xi}$ невырождена. Тогда как нетрудно проверить,

$$\min_R M\|\tilde{\eta} - R\tilde{\xi}\|_m^2 =$$

$$= \min_R \text{tr}(S_{\eta\eta} - 2S_{\eta\xi}R^* + RS_{\xi\xi}R^*) = \text{tr}(S_{\eta\eta} - S_{\eta\xi}S_{\xi\xi}^{-1}S_{\xi\eta}) \quad (9.56)$$

и достигается при ¹⁾

$$R_* = S_{\eta\xi} S_{\xi\xi}^{-1}. \quad (9.57)$$

Соответственно искомый прогноз η дается равенством $\hat{\eta} = R_*\xi + r_* = M\eta + S_{\eta\xi} S_{\xi\xi}^{-1}(\xi - M\xi)$.

1.10. Законы больших чисел

Стандартное теоретико-вероятностное заключение («вероятность события A равна p »), как правило, не позволяет предсказать, произойдет событие A или нет. Исключения составляют лишь те случаи, когда p либо «очень мало» ²⁾, либо «очень близко» к единице, и при этом можно утверждать, что событие A «практически невозможно» или соответственно «практически достоверно». Так, например, если подбрасывать монету 1000 раз, то событие, состоящее в выпадении герба все 1000 раз, можно считать практически невозможным, а событие, состоящее в том, что герб выпадет хотя бы один раз, — практически достоверным.

На практике заключение о малости p определяется контекстом. Например, значение $p = 0.001$ мало, если речь идет о вероятности простудиться, и немало, если 0.001 — вероятность утонуть.

Мы рассмотрим результаты теории вероятностей, известные под названием «законов больших чисел» и позволяющие делать подобные предсказания. С этой целью введем понятия сходимости последовательности случайных величин: сходимости по вероятности и с вероятностью единица (или почти наверное, п.н.). Разумеется, если речь идет о последовательности случайных величин, то считается, что все случайные величины определены на одном и том же вероятностном пространстве.

Определение 1.10.1. Последовательность случайных величин $\{\xi_n\}$ *сходится по вероятности* к случайной величине ξ , если для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{|\xi_n - \xi| > \varepsilon\}) = 0; \quad (10.1)$$

обозначения: $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$ или $\xi_n \xrightarrow{P} \xi, n \rightarrow \infty$.

Этот тип сходимости означает, что каково бы ни было $\varepsilon > 0$, найдется число N , такое, что для всех $n \geq N$ вероятность неравенства $|\xi_n - \xi| > \varepsilon$ будет сколь угодно малой, или, иначе, событие $|\xi_n - \xi| > \varepsilon$ будет практически невозможным. Точнее: $\forall \varepsilon > 0 \forall \delta > 0 \exists N = N(\varepsilon, \delta) \forall n \geq N P(\{|\xi_n - \xi| > \varepsilon\}) < \delta$ (или $P(\{|\xi_n - \xi| \leq \varepsilon\}) \geq 1 - \delta$). При этом

¹⁾ $\text{tr}(S_{\eta\eta} - 2S_{\eta\xi}R^* + RS_{\xi\xi}R^*)$ — квадратичная форма как функция матричных элементов матрицы R .

²⁾ Утверждение о малости требует уточнения — малости по сравнению с чем? Предельный переход позволяет говорить о малости по сравнению с чем угодно.

тот или иной критерий «практической невозможности» обуславливает выбор соответствующего достаточно большого числа N .

Следующая теорема, характерная для стохастического анализа, будет использована в §2.2 главы 2.

Теорема 1.10.1. Пусть $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$, $g(x)$, $x \in R_1$, — непрерывная функция, так что $\eta = g(\xi)$ и $\eta_n = g(\xi_n)$, $n = 1, 2, \dots$, — случайные величины. Тогда $\eta_n \xrightarrow{P} \eta$ при $n \rightarrow \infty$.

Доказательство. Пусть $\varkappa > 0$ — любое, а $I \subset \mathcal{R}^1$ — конечный интервал, такой, что $P(\{\xi \in \mathcal{R}^1 \setminus I\}) \leq \varkappa/2$. Функция $g(\cdot)$ равномерно непрерывна на I , поэтому для любого $\varepsilon > 0$ найдется такое $\delta = \delta(\varepsilon) > 0$, что для всех x и y , таких, что $x \in I$ и $|x - y| < \delta$, $|g(y) - g(x)| < \varepsilon$. Таким образом,

$$|g(\xi_n) - g(\xi)| < \varepsilon, \text{ если } \xi \in I \text{ и } |\xi_n - \xi| < \delta. \quad (10.2)$$

Поскольку $\xi_n \xrightarrow{P} \xi$ при $n \rightarrow \infty$, то для указанного δ

$$P(\{|\xi_n - \xi| < \delta\}) \geq 1 - \varkappa/2, \quad (10.3)$$

если $n \geq n_0 = n_0(\delta, \varkappa)$. Из (10.2) и (10.3) следует, что при $n \geq n_0$

$$\begin{aligned} P(\{|g(\xi_n) - g(\xi)| < \varepsilon\}) &\geq P(\{|\xi_n - \xi| < \delta, \xi \in I\}) \geq \\ &\geq P(\{|\xi_n - \xi| < \delta\}) - P(\{\xi \in \mathcal{R}^1 \setminus I\}) \geq 1 - \varkappa, \end{aligned} \quad (10.3')$$

где предпоследнее неравенство основано на соотношении

$$\begin{aligned} P(\{|\xi_n - \xi| < \delta\}) &= \\ &= P(\{|\xi_n - \xi| < \delta, \xi \in I\}) + P(\{|\xi_n - \xi| < \delta, \xi \in \mathcal{R}^1 \setminus I\}) \leq \\ &\leq P(\{|\xi_n - \xi| < \delta, \xi \in I\}) + P(\{\xi \in \mathcal{R}^1 \setminus I\}). \end{aligned}$$

Неравенство (10.3') ввиду произвольности ε и $\varkappa$ доказывает теорему. ■

В этом параграфе мы будем постоянно использовать неравенство Чебышева (см. § 1.9): для любой случайной величины ξ с конечной дисперсией $D\xi$ и любого $\varepsilon > 0$ справедливо неравенство

$$P(\{|\xi - M\xi| > \varepsilon\}) \leq D\xi/\varepsilon^2. \quad (10.4)$$

Основываясь на неравенстве Чебышева, сформулируем достаточное условие (признак) сходимости по вероятности.

Лемма 1.10.1. Если для последовательности случайных величин $\{\xi_n\}$ $M\xi_n = 0$, $D\xi_n \rightarrow 0$ при $n \rightarrow \infty$, то $\xi_n \xrightarrow[n \rightarrow \infty]{P} 0$.

Доказательство. В силу (10.4) и того, что $M\xi_n = 0$, для любого фиксированного $\varepsilon > 0$

$$0 \leq P(\{|\xi_n| > \varepsilon\}) \leq D\xi_n/\varepsilon^2 \rightarrow 0, \quad n \rightarrow \infty,$$

т.е. $\xi_n \xrightarrow[n \rightarrow \infty]{P} 0$.

Следствием леммы 1.10.1 является закон больших чисел в форме Чебышева.

Теорема 1.10.2. (П. Л. Чебышев). Пусть $\xi_1, \xi_2, \dots$ — последовательность попарно независимых случайных величин, дисперсии которых ограничены в совокупности: $D\xi_i \leq c, i = 1, 2, \dots$. Тогда последовательность случайных величин $\eta_n = \frac{1}{n} \sum_{i=1}^n (\xi_i - M\xi_i), n = 1, 2, \dots$, сходится по вероятности к нулю при $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| > \varepsilon \right) = 0, \varepsilon > 0 \text{ — любое.} \quad (10.5)$$

Доказательство. При $n \rightarrow \infty$

$$M\eta_n = \frac{1}{n} \sum_{i=1}^n (M\xi_i - M\xi_i) = 0, \quad D\eta_n = \frac{1}{n^2} \sum_{i=1}^n D\xi_i \leq \frac{cn}{n^2} \rightarrow 0,$$

ибо в силу попарной независимости $\xi_1, \xi_2, \dots$ $D \left(\sum_{i=1}^n (\xi_i - M\xi_i) \right) = \sum_{i=1}^n D(\xi_i - M\xi_i) = \sum_{i=1}^n D\xi_i$. Отсюда на основании леммы 1.10.1 следует утверждение теоремы. ■

Теорему Чебышева можно записать в виде

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| \leq \varepsilon \right) = 1, \varepsilon > 0 \text{ — любое.} \quad (10.6)$$

Следствие 1.10.1. Пусть в условиях теоремы Чебышева случайные величины $\xi_i, i = 1, 2, \dots$, имеют одинаковые математические ожидания: $M\xi_1 = M\xi_2 = \dots = \mu$. Тогда последовательность $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ при $n \rightarrow \infty$ сходится по вероятности к μ :

$$\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{P} \mu. \quad (10.7)$$

Это следствие теоремы Чебышева служит обоснованием правила среднего арифметического, применяемого в теории измерений, которое сводится к тому, что, повторив n раз измерение величины μ и получив в качестве результатов значения случайных величин $\xi_1, \xi_2, \dots, \xi_n$, за приближенное значение μ принимают среднее арифметическое наблюдаемых значений, т.е. значение случайной величины $\hat{\mu} = \frac{1}{n}(\xi_1 + \dots + \xi_n)$. Если при измерениях отсутствует систематическая ошибка (т.е. все $M\xi_i = \mu, i = 1, 2, \dots, n$), то согласно закону больших чисел при достаточно больших n с вероятностью, сколь угодно близкой к единице, будет получен результат $\hat{\mu}$, сколь угодно мало отличающийся от истинного значения μ .

Следствием закона больших чисел в форме Чебышева является **Теорема 1.10.3 (Бернулли)**. Пусть η_n — число успехов в серии n испытаний Бернулли и p — вероятность успеха при каждом испытании. Тогда последовательность частот успехов $\{\eta_n/n\}$ при $n \rightarrow \infty$ сходится по вероятности к p .

Доказательство. Введем случайную величину ξ_k , равную числу успехов при k -м испытании, $k = 1, 2, \dots$. Тогда $\eta_n = \xi_1 + \dots + \xi_n$, $M\xi_k = p$, $D\xi_k = pq$ (см. § 9). Поэтому согласно теореме Чебышева (условия которой, очевидно, выполнены для ξ_i , $i = 1, 2, \dots$) при любом $\varepsilon > 0$

$$\begin{aligned} & \lim_{n \rightarrow \infty} P\left(\left|\frac{\eta_n}{n} - p\right| > \varepsilon\right) \equiv \\ & \equiv \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k\right| > \varepsilon\right) = 0. \end{aligned} \quad (10.8)$$

Таким образом $\forall \varepsilon > 0 \quad \sum_{k: |\frac{k}{n} - p| \leq \varepsilon} C_n^k p^k q^{n-k} \rightarrow 1, \quad n \rightarrow \infty$.

Приведем *примеры* применений закона больших чисел.

Пример 1.10.1. Рассмотрим следующую модель независимых испытаний: при каждом испытании полная группа событий состоит из $A_1, A_2, \dots, A_r$, и вероятность наступления при каждом испытании события A_i равна p_i , $p_i \geq 0$, $p_1 + \dots + p_r = 1$. Пусть $\eta_n^{(i)}$, $i = 1, 2, \dots, r$, случайная величина, равная числу наступлений события A_i в серии из n испытаний, тогда последовательность частот $\eta_n^{(i)}/n$ появлений события A_i при $n \rightarrow \infty$ сходится по вероятности к p_i , $i = 1, 2, \dots, r$. В самом деле, пусть ξ_k — число наступлений A_i в k -м испытании, $\xi_k = 0$ или 1 с вероятностями соответственно $1 - p_i$ и p_i . Поэтому $M\xi_k = p_i$, $D\xi_k = M\xi_k^2 - (M\xi_k)^2 = p_i - p_i^2 = p_i(1 - p_i)$, а $\eta_n^{(i)} = \sum_{k=1}^n \xi_k$.

Применяя теорему Чебышева, получим $\frac{\eta_n^{(i)}}{n} \xrightarrow{P} p_i, \quad n \rightarrow \infty, \quad i = 1, 2, \dots, r$.

Пример 1.10.2. Метод Монте-Карло вычисления интеграла. Пусть $f(\cdot)$ — непрерывная функция, определенная на интервале $[0, 1]$, принимающая значения в $[0, 1]$, и $\xi_1, \eta_1, \xi_2, \eta_2, \dots$ — последовательность взаимно независимых равномерно распределенных на $[0, 1]$ случайных величин. Определим последовательность случайных величин $\rho_i = \begin{cases} 1, & \text{если } f(\xi_i) \geq \eta_i, \\ 0, & \text{если } f(\xi_i) < \eta_i, \end{cases} \quad i = 1, 2, \dots$. Так как

$$M\rho_i = P(f(\xi_i) \geq \eta_i) = \int_0^1 f(x) dx, \quad i = 1, 2, \dots, \text{ то в силу закона больших}$$

$$\text{чисел } \frac{1}{n} \sum_{i=1}^n \rho_i \xrightarrow[P]{n \rightarrow \infty} M\rho_i = \int_0^1 f(x) dx.$$

Пример 1.10.3.

Теорема 1.10.4. (С.Н. Бернштейн). Пусть $f(x)$, $x \in [0, 1]$, — непрерывная функция, η_n — число успехов в последовательности n испытаний Бернулли с вероятностью успеха p . Тогда при $n \rightarrow \infty$

$$b_n(p) = Mf(\eta_n/n) = \sum_{k=0}^n f(k/n) C_n^k p^k q^{n-k} \rightarrow f(p)$$

равномерно для $p \in [0, 1]$.

Доказательство. При достаточно большом n и любом $\varepsilon > 0$

$$\begin{aligned} |f(p) - b_n(p)| &= \left| \sum_{k=0}^n (f(p) - f(k/n)) C_n^k p^k q^{n-k} \right| \leq \\ &\leq \sum_{k: |\frac{k}{n} - p| \leq \delta} |f(p) - f(k/n)| C_n^k p^k q^{n-k} + \\ &+ \sum_{k: |\frac{k}{n} - p| > \delta} |f(p) - f(k/n)| C_n^k p^k q^{n-k} \leq \frac{\varepsilon}{2} + 2m \sum_{k: |\frac{k}{n} - p| > \delta} C_n^k p^k q^{n-k} = \\ &= \frac{\varepsilon}{2} + 2qP(|\frac{\eta_n}{n} - p| > \delta) < \varepsilon, \end{aligned} \quad (10.9)$$

где δ выбрано так, чтобы $|f(x) - f(y)| \leq \varepsilon/2$, если $|x - y| \leq \delta$, $m = \max_{0 \leq p \leq 1} |f(p)|$. Согласно (10.9) $\lim_{n \rightarrow \infty} \max_{0 \leq p \leq 1} |f(p) - b_n(p)| = 0$, т.е. всякая непрерывная на $[0, 1]$ функция $f(\cdot)$ есть равномерный при $n \rightarrow \infty$ предел последовательности полиномов $\sum_{k=0}^n f(k/n) C_n^k p^k (1-p)^{n-k}$, $p \in [0, 1]$. ■

В определенном смысле теорема 1.10.3 может служить «аксиомой измерения», доставляя способ практического определения тех вероятностей, о которых идет речь в аксиоматической теории вероятностей. Закон больших чисел Бернулли утверждает, что для фиксированного достаточно большого n с вероятностью, сколь угодно близкой к единице, частота η_n/n будет уклоняться от вероятности p меньше, чем на ε . Отсюда, однако, не следует, что разность $|\eta_n/n - p|$ останется малой для всех достаточно больших n . Теорема 1.10.3 гарантирует лишь, что большие отклонения η_n/n от p могут появляться весьма редко. Для полного обоснования частотной интерпретации вероятности (см. § 1) желательно иметь теорему, обеспечивающую сходимости последовательности частот к вероятности. Мы сейчас введем некоторые новые понятия и дадим усиленный вариант теоремы Бернулли, согласно которому частота η_n/n с увеличением n приближается (с вероятностью единица) и остается близкой к p . этому требованию.

Определение 1.10.2. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ с вероятностью единица (или

почти наверное, п. н.), если ¹⁾

$$P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\}) = 1, \quad (10.10)$$

т. е. при $n \rightarrow \infty$ $\xi_n(\omega) \rightarrow \xi(\omega)$ для всех $\omega \in \Omega$, за исключением, быть может, множества $C \subset \Omega$ нулевой вероятности, $P(C) = 0$. Эта сходимость обозначается так: $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} \xi$.

Согласно этому определению для каждого $\omega \in \Omega \setminus C$ и любого $\varepsilon > 0$ $|\xi_n(\omega) - \xi(\omega)| \leq \varepsilon$ для всех достаточно больших n . Поэтому если событие $A_{n,\varepsilon} = \{\omega \in \Omega : |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}$, $n = 1, 2, \dots$, то для любого $\varepsilon > 0$ с вероятностью 1 происходит лишь конечное число событий $A_{n,\varepsilon}$. Оказывается, что это условие является и достаточным для сходимости с вероятностью 1. В самом деле, возьмем $\varepsilon = 1/k$ и обозначим B_k событие, состоящее в том, что происходит лишь конечное число событий из $A_{n,1/k} = \{\omega \in \Omega : |\xi_n(\omega) - \xi(\omega)| > 1/k\}$, $n = 1, 2, \dots$. По условию $P(B_k) = 1$, $k = 1, 2, \dots$. Очевидно, что события B_k , $k = 1, 2, \dots$, образуют монотонно убывающую последовательность: $B_1 \supset B_2 \supset B_3 \supset \dots$. Обозначим $B = \bigcap_{k=1}^{\infty} B_k$. В силу непрерывности вероятности $P(B) = \lim_{k \rightarrow \infty} P(B_k) = 1$, так как все $P(B_k) = 1$. Из определения события B следует, что B состоит из всех таких $\omega \in \Omega$, для которых $\xi_n(\omega) \rightarrow \xi(\omega)$ при $n \rightarrow \infty$. Итак, $P(B) = 1$, и высказанное выше утверждение доказано. Таким образом, $\xi_n \xrightarrow{\text{п. н.}} \xi$ при $n \rightarrow \infty$ тогда и только тогда, когда для любого $\varepsilon > 0$ вероятность того, что осуществляется лишь конечное число событий

$$|\xi_n - \xi| > \varepsilon, \quad n = 1, 2, \dots, \quad (10.11)$$

равна 1.

¹⁾ Множество тех $\omega \in \Omega$, при которых $\{\xi_n(\omega)\}$ сходятся к $\xi(\omega)$ — событие, ибо оно равно $\bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |\xi_n(\omega) - \xi(\omega)| < \frac{1}{k}\} = \{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\}$. В случае сходимости ξ_n к ξ с вероятностью единица $P(\bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |\xi_n(\omega) - \xi(\omega)| < \frac{1}{k}\}) = 1$. Тем более: $\forall \varepsilon > 0 \quad P(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |\xi_n(\omega) - \xi(\omega)| < \varepsilon\}) = 1$ и следовательно, $P(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\}) = 0$, т. е. $\forall \varepsilon > 0$ с вероятностью единица происходит не более конечного числа событий $|\xi_n(\omega) - \xi(\omega)| \geq \varepsilon$. В иных обозначениях: $\forall \varepsilon > 0 \quad P(\varliminf_{n \rightarrow \infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| < \varepsilon\}) = 1$, $P(\varlimsup_{n \rightarrow \infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\}) = 0$.

Лемма 1.10.2 (Бореля–Кантелли). Если для последовательности $\{A_n\}$ произвольных событий A_n , $n = 1, 2, \dots$, выполнено условие

$$\sum_{n=1}^{\infty} P(A_n) < \infty, \quad (10.12)$$

то в последовательности $\{A_n\}$ происходит п. н. конечное число событий.

Доказательство. Событие $A = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$ происходит тогда и только тогда, когда происходит бесконечно много событий в последовательности $A_1, A_2, \dots$, см. § 1.3.3. Действительно, A происходит, если и только если происходят все события $B_n = \bigcup_{k \geq n} A_k$, $n = 1, 2, \dots$, а каждое B_n происходит, если происходит хотя бы одно A_k при $k \geq n$, $n = 1, 2, \dots$. Поскольку $B_1 \supset B_2 \supset \dots$, то в силу непрерывности вероятности и сходимости ряда $\sum_{n=1}^{\infty} P(A_n)$

$$P(A) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} P(A_k) = 0. \quad (10.13)$$

Равенство (10.13) означает, что $P(\{\text{происходит конечное число событий в последовательности } A_1, A_2, \dots\}) = 1$. ■

Используя лемму Бореля–Кантелли, установим следующий *усиленный вариант закона больших чисел*.

Теорема 1.10.5 (усиленный вариант закона больших чисел). Пусть $\xi_1, \xi_2, \dots$ — последовательность попарно независимых случайных величин, для которых $M\xi_i = \mu$, $D\xi_i = \sigma^2$. Тогда при $n \rightarrow \infty$

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{\text{п. н.}} \mu. \quad (10.14)$$

Доказательство. Вводя в случае необходимости новые случайные величины $\xi'_i = \xi_i - \mu$, можем считать, что $\mu = 0$. Определим случайную величину

$$\eta_k = \sum_{i=1}^k \xi_i. \quad (10.15)$$

Докажем, что при $n \rightarrow \infty$ $(1/n)\eta_n \xrightarrow{\text{п. н.}} 0$. Для каждого натурального n возьмем натуральное число m так, чтобы

$$m^2 \leq n \leq (m+1)^2. \quad (10.16)$$

Так как $M\eta_k = 0$, то неравенство Чебышева (10.4) дает

$$P\left(\left|\frac{\eta_{m^2}}{m^2}\right| > \varepsilon\right) \leq \frac{D\eta_{m^2}}{\varepsilon^2 m^4} = \frac{\sigma^2}{\varepsilon^2 m^2}. \quad (10.17)$$

Положим

$$\widehat{\eta}_{m^2} = \max_{m^2+1 \leq k \leq (m+1)^2} |\xi_{m^2} + \dots + \xi_k|. \quad (10.18)$$

Снова применяя неравенство Чебышева (10.4), получим

$$\begin{aligned} P\left(\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right| > \varepsilon\right\}\right) &\leq \sum_{k=m^2+1}^{(m+1)^2} P\left(\left\{\left|\frac{\xi_{m^2} + \dots + \xi_k}{m^2}\right| > \varepsilon\right\}\right) \leq \\ &\leq \sum_{k=m^2+1}^{(m+1)^2} \frac{(k - m^2)\sigma^2}{\varepsilon^2 m^4} \leq 2m \frac{(2m+1)\sigma^2}{\varepsilon^2 m^4} \leq \frac{5\sigma^2}{\varepsilon^2 m^2} \end{aligned} \quad (10.19)$$

(здесь в сумме $2m$ слагаемых и $(k - m^2) \leq (2m + 1)$). В силу оценок (10.17) и (10.19) числовые ряды $\sum_{m=1}^{\infty} P\left(\left\{\left|\frac{\eta_{m^2}}{m^2}\right| > \varepsilon\right\}\right)$ и $\sum_{m=1}^{\infty} P\left(\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right| > \varepsilon\right\}\right)$ сходятся, а тогда на основании леммы Бореля—Кантелли заключаем, что с вероятностью 1 может произойти только конечное число событий $\left\{\left|\frac{\eta_{m^2}}{m^2}\right| > \varepsilon\right\}$ и $\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right| > \varepsilon\right\}$, т. е. согласно критерию (10.12)

$$\frac{\eta_{m^2}}{m^2} \xrightarrow{\text{п.н.}} 0 \text{ и } \frac{\widehat{\eta}_{m^2}}{m^2} \xrightarrow{\text{п.н.}} 0 \text{ при } m \rightarrow \infty. \quad (10.20)$$

Поскольку для любого n из интервала (10.16) $\left|\frac{\eta_n}{n}\right| \leq \left|\frac{\eta_{m^2}}{m^2}\right| + \left|\frac{\widehat{\eta}_{m^2}}{m^2}\right|$, то из (10.20) следует, что при $n \rightarrow \infty$ $\eta_n/n \rightarrow 0$ с вероятностью единица. ■

Простым следствием доказанной теоремы является усиленный закон больших чисел Бернулли.

Теорема 1.10.6 (Борель). Пусть η_n — число успехов в серии из n независимых испытаний Бернулли, p — вероятность успеха при каждом испытании. Тогда последовательность частот $\{\eta_n/n\}$ при $n \rightarrow \infty$ почти наверное сходится к вероятности p .

Доказательство. Достаточно ввести случайную величину ξ_i , равную числу успехов в i -м испытании, $\xi_i = 1, 0$, $M\xi_i = p$, $D\xi_i = pq$, $i = 1, \dots, n$, $\eta_n = \sum_{i=1}^n \xi_i$ и применить теорему 1.10.5 к η_n (10.15). ■

Рассмотрим теперь один важный вариант закона больших чисел, принадлежащий Хинчину. В этом варианте не требуется существования (а тем более ограниченности в совокупности) дисперсий случайных величин.

Теорема 1.10.7 (Хинчин А.Я.). Пусть одинаково распределенные случайные величины $\xi_1, \xi_2, \dots$ попарно независимы и имеют конечное математическое ожидание $M\xi_i = \mu$, $i = 1, 2, \dots$. Тогда при $n \rightarrow \infty$

$$\eta_n = \frac{1}{n} \sum_{k=1}^n \xi_k \text{ сходится по вероятности к } \mu, \quad \eta_n \xrightarrow[n \rightarrow \infty]{P} \mu.$$

Доказательство. Проведем доказательство с помощью широко применяемого в теории вероятностей метода урезания: введем новые случайные величины $\bar{\xi}_k$ по формуле:

$$\bar{\xi}_k = \begin{cases} \xi_k, & \text{если } |\xi_k| \leq y, \\ 0, & \text{если } |\xi_k| > y, \end{cases} \quad k = 1, 2, \dots, \quad (10.21)$$

число $y > 0$ будет выбрано позже. Помимо $\eta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ рассмотрим $\bar{\eta}_n$ — среднее «урезанных» случайных величин, $\bar{\eta}_n = \frac{1}{n} \sum_{k=1}^n \bar{\xi}_k$. Для любого $\varepsilon > 0$ выполняется неравенство, в котором $\mu = M\xi_k$:

$$P(\{|\eta_n - \mu| > \varepsilon\}) \leq P(\{|\bar{\eta}_n - \bar{\mu}| > \varepsilon\}) + P(\{\eta_n \neq \bar{\eta}_n\}), \quad (10.22)$$

поскольку если наступает событие в левой части этого неравенства, то наступает хотя бы одно из событий в его правой части. Далее ¹⁾

$$\bar{\mu} = M\bar{\xi}_k = \int_{-y}^y x dF(x), \quad (10.23)$$

$F(\cdot)$ — функция распределения случайных величин ξ_k , $k = 1, 2, \dots$, поэтому

$$|\bar{\mu} - \mu| < \varepsilon, \quad (10.24)$$

если $y \geq y_0 = y_0(\varepsilon)$.

Ввиду (10.22) и (10.24) при $y \geq y_0$

$$P(\{|\eta_n - \mu| > 2\varepsilon\}) \leq P(\{|\bar{\eta}_n - \bar{\mu}| > \varepsilon\}) + P(\{\eta_n \neq \bar{\eta}_n\}). \quad (10.25)$$

Оценим правую часть в (10.25). По неравенству Чебышева (10.4):

$$\begin{aligned} P(\{|\bar{\eta}_n - \bar{\mu}| > \varepsilon\}) &\leq \frac{D\bar{\eta}_n}{\varepsilon^2} \leq \frac{1}{\varepsilon^2} M\bar{\eta}_n^2 \leq \frac{1}{\varepsilon^2 n^2} \sum_{k=1}^n M\bar{\xi}_k^2 = \\ &= \frac{n}{\varepsilon^2 n^2} \int_{-y}^y x^2 dF(x) \leq \frac{y}{\varepsilon^2 n} \int_{-y}^y |x| dF(x) \leq \frac{y}{\varepsilon^2 n} \int_{-\infty}^{\infty} |x| dF(x) = \frac{Ay}{\varepsilon^2 n}, \end{aligned} \quad (10.26)$$

где A обозначена существующая по условию величина $A = \int_{-\infty}^{\infty} |x| dF(x)$. Событие $\{\eta_n \neq \bar{\eta}_n\}$ влечет событие, состоящее в том,

¹⁾ Здесь и ниже интеграл $\int x dF(x)$, как обычно, есть сокращенная запись для $\sum x_k p_k$ в дискретном случае и $\int x p(x) dx$ — в непрерывном.

что хотя бы одна случайная величина $|\xi_k| > y$, $k = 1, 2, \dots, n$. Поэтому для любого $\varepsilon_1 > 0$

$$\begin{aligned} P(\{\eta_n \neq \bar{\eta}_n\}) &\leq \sum_{k=1}^n P(\{|\xi_k| > y\}) = \\ &= n \int_{|x|>y} dF(x) < \frac{n}{y} \int_{|x|>y} |x| dF(x) < \frac{n}{y} \varepsilon_1, \end{aligned} \quad (10.27)$$

если не только $y \geq y_0$, но сверх того $y \geq y_1 = y_1(\varepsilon_1)$.

Подставляя (10.26) и (10.27) в (10.25), получим, что при $y \geq \bar{y} = \max\{y_0, y_1\}$

$$P(\{|\eta_n - \mu| > 2\varepsilon\}) \leq \frac{Ay}{\varepsilon^2 n} + \frac{n}{y} \varepsilon_1. \quad (10.28)$$

Пусть теперь $\delta > 0$ — любое, положим $y = (\varepsilon^2/A)\delta n$, а $\varepsilon_1 = \delta^2/\varepsilon^2 A$. Тогда из (10.28) при всех достаточно больших n , $n \geq n_0(\delta)$, находим

$$P(\{|\eta_n - \mu| > 2\varepsilon\}) < 2\delta, \quad (10.29)$$

что и доказывает теорему. ■

В заключение докажем теорему, в которой не требуется попарной независимости случайных величин.

Теорема 1.10.8. (А.А. Марков). Если последовательность случайных величин $\xi_1, \xi_2, \dots$ такова, что $M\xi_i = \mu$ и $\frac{1}{n^2} D\left(\sum_{i=1}^n \xi_i\right) \rightarrow 0$ при

$n \rightarrow \infty$, то при $n \rightarrow \infty$ $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i \rightarrow \mu$ по вероятности.

Доказательство. Положим $\zeta_n = \frac{1}{n} \sum_{i=1}^n \xi_i - \mu$. Тогда $M\zeta_n = 0$,

$D\zeta_n = \frac{1}{n^2} D\left(\sum_{i=1}^n \xi_i\right) \rightarrow 0$ при $n \rightarrow \infty$ и согласно лемме 1.10.1 получаем утверждение теоремы. ■

В заключение определим и вкратце охарактеризуем еще два типа сходимости, с которыми будем встречаться далее.

Определение 1.10.3. Сходимость по распределению: при $n \rightarrow \infty$ $\xi_n \xrightarrow{d} \xi$, если $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности $F(x)$, $-\infty < x < \infty$. Эта сходимость называется слабой.

Определение 1.10.4. Сходимость в среднем квадратичном (с.к.): $\xi_n \xrightarrow{\text{с.к.}} \xi$, или $\text{l.i.m.}_{n \rightarrow \infty} \xi_n = \xi$, если при $n \rightarrow \infty$ $M(\xi_n - \xi)^2 \rightarrow 0$.

Зависимость между различными типами сходимости выражается схемой:

$$\begin{aligned} \xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi &\implies \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{d} \xi. \\ \xi_n \xrightarrow[n \rightarrow \infty]{\text{с.к.}} \xi &\implies \end{aligned}$$

Теорема 1.10.9. $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{d} \xi$, а если $\xi = \text{const}$, то $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi \iff \xi_n \xrightarrow[n \rightarrow \infty]{d} \xi$.

Доказательство. Покажем, что $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$. Действительно, если $\xi_n \not\xrightarrow[n \rightarrow \infty]{P} \xi$, то $\exists \varepsilon > 0, \exists \delta > 0 \forall N \exists n \geq N : P(|\xi_n - \xi| \geq \varepsilon) \geq \delta$. Это означает, что $\xi_n \not\xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi$, ибо для $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi$ необходимо и достаточно, чтобы выполнялось не более конечного числа событий $|\xi_n - \xi| \geq \varepsilon$.

Докажем теперь, что $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{d} \xi$. Рассмотрим событие $\xi_n - x = \xi_n - (\xi - \varepsilon) + (\xi - \varepsilon) - x < 0$. Оно влечет по крайней мере одно из событий $\xi_n - (\xi - \varepsilon) \leq 0, (\xi - \varepsilon) - x < 0$. Следовательно, $P(\{\xi_n < x\}) \leq P(\{(\xi_n - \xi \leq -\varepsilon)\} \cup \{(\xi - \varepsilon) < x\}) \leq P(\{\xi_n - \xi \leq -\varepsilon\}) + P(\{(\xi - \varepsilon) < x\})$ и далее, в силу $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$,

$$\overline{\lim}_{n \rightarrow \infty} P(\{\xi_n < x\}) \leq F_\xi(x + \varepsilon). \quad (10.30)$$

Аналогично, $\xi_n - x = \xi_n - (\xi + \varepsilon) + (\xi + \varepsilon) - x \geq 0$ и, следовательно, $\xi_n - x \geq 0$ влечет по крайней мере одно из событий $\xi_n - (\xi + \varepsilon) \geq 0, (\xi + \varepsilon) - x \geq 0$. Поэтому $P(\{\xi_n \geq x\}) \leq P(\{\xi_n - \xi \geq -\varepsilon\}) + P(\{\xi \geq x - \varepsilon\})$ и $\overline{\lim}_{n \rightarrow \infty} (1 - F_{\xi_n}(x)) \leq -F_\xi(x - \varepsilon) + 1$. Поскольку $\overline{\lim}_{n \rightarrow \infty} (1 - t_n) = \lim_{n \rightarrow \infty} \sup_{k \geq n} (-t_k) = \lim_{n \rightarrow \infty} [-\inf_{k \geq n} t_k] = -\lim_{n \rightarrow \infty} t_n$, то

$$\underline{\lim}_{n \rightarrow \infty} F_{\xi_n}(x) \geq F_\xi(x - \varepsilon). \quad (10.31)$$

Неравенства (10.30) и (10.31) верны для любого $\varepsilon > 0$. Поэтому если точка x — точка непрерывности $F_\xi(\cdot)$, то при $\varepsilon \rightarrow 0$ найдем $\underline{\lim}_{n \rightarrow \infty} F_{\xi_n}(x) \geq F_\xi(x) \geq \overline{\lim}_{n \rightarrow \infty} F_{\xi_n}(x)$, то есть $F_\xi(x) = \lim_{n \rightarrow \infty} F_{\xi_n}(x)$.

Покажем, что $\xi_n \xrightarrow[n \rightarrow \infty]{d} \xi \not\Rightarrow \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$. Пусть $\xi, \xi_1, \xi_2, \dots$ — попарно независимые одинаково распределенные случайные величины, отличные от const. Поскольку их функции распределения совпадают, то $\xi_n \xrightarrow[n \rightarrow \infty]{d} \xi$. С другой стороны, $\{\xi_n < -\varepsilon, \xi \geq \varepsilon\} \subset \{\xi - \xi_n \geq 2\varepsilon\} \Rightarrow P(\{\xi - \xi_n \geq 2\varepsilon\}) \geq F(-\varepsilon)(1 - F(\varepsilon)) \not\xrightarrow[n \rightarrow \infty]{} 0$, т.е. $\xi_n \not\xrightarrow[n \rightarrow \infty]{P} \xi$. Однако, если $\xi = \text{const}$, то $\xi_n \xrightarrow[n \rightarrow \infty]{d} \xi \Rightarrow \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$. Действительно, пусть $P(\{\xi = a\}) = 1$ и $F_{\xi_n} \xrightarrow[n \rightarrow \infty]{} F_\xi(x)$ при любом $x \neq a$ (в точке a — разрыв). Тогда поскольку $\begin{cases} F_\xi(x) = 0, & x \leq a, \\ F_\xi(x) = 1, & x > a, \end{cases} \lim_{n \rightarrow \infty} F_{\xi_n}(x) = 0$ при $x < a$, $\lim_{n \rightarrow \infty} F_{\xi_n}(x) = 1$ при

$x > a$, то при $n \rightarrow \infty$. $P(\{\xi_n > a - \varepsilon\}) \rightarrow 0$, $P(\{\xi_n < a + \varepsilon\}) \rightarrow 0$
 $\Rightarrow P(\{|\xi_n - a| < \varepsilon\}) \rightarrow 0$, т.е. $\xi_n \xrightarrow{P} \xi$. ■

1.11. Характеристические функции. Центральные предельные теоремы

Утверждения, полученные в форме законов больших чисел, представляют собой заключения о сходимости последовательности случайных величин ξ_n , $n = 1, 2, \dots$, к некоторой случайной (или неслучайной) величине ξ . Эти утверждения не содержат информации о том, как аппроксимировать распределение случайной величины ξ_n при больших n . Ответ на этот вопрос доставляют так называемые *центральные предельные теоремы*, в которых речь идет о сходимости по распределению последовательности случайных величин. Основным аппаратом, используемым при изучении центральных предельных теорем, является аппарат характеристических функций, играющий важную роль и в других разделах теории вероятностей.

1.11.1. Характеристические функции. Хорошо известна большая роль теории преобразования Фурье в математическом анализе и в теории дифференциальных уравнений. Характеристические функции — инструмент для использования преобразования Фурье в теории вероятностей.

Определение 1.11.1. *Характеристической функцией (х.ф.)* случайной величины ξ называется функция $f_\xi(t)$, $t \in \mathcal{R}^1$, определенная равенством

$$f_\xi(t) = M e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF_\xi(x), \quad -\infty < t < \infty. \quad (11.1)$$

Математическое ожидание комплексной случайной величины $\zeta = \xi + i\eta$, где ξ и η — действительные случайные величины¹⁾, по определению

$$M\zeta = M\xi + iM\eta.$$

В определении (11.1) интеграл понимается либо как сумма абсолютно сходящегося ряда

$$f_\xi(t) = \sum_{k=1}^{\infty} e^{itx_k} p_k, \quad (11.2)$$

¹⁾ Относительно совместного распределения случайных величин ξ и η не делается никаких предположений.

если ξ — дискретная случайная величина, x_k — ее значения, а p_k , $k = 1, 2, \dots$, — соответствующие вероятности, либо как абсолютно сходящийся интеграл

$$f_\xi(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx, \quad (11.3)$$

если ξ — абсолютно непрерывная случайная величина с плотностью $p(\cdot)$. Хотя интеграл (11.1) представляет собой обычный интеграл Стильтьеса, мы не опираемся на его специфические свойства и будем рассматривать (11.1) как краткую запись для выражений (11.2) и (11.3). Все дальнейшие рассуждения этого параграфа проводятся таким образом, что они одинаково применимы для случаев (11.2) и (11.3), ввиду чего мы будем использовать лишь обозначение (11.1).

Характеристическая функция существует для любой случайной величины, поскольку ввиду равенства $|e^{itx}| = 1$ ряд (11.2) и интеграл (11.3) сходятся абсолютно.

Очевидно, $f_\xi(0) = 1$, и $|f_\xi(t)| \leq 1$, $t \in \mathcal{R}^1$.

Лемма 1.11.1. Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые в совокупности случайные величины, то

$$\begin{aligned} f_{\xi_1 + \dots + \xi_n}(t) &= f_{\xi_1}(t) \dots f_{\xi_n}(t), \quad t \in \mathcal{R}^1; \\ f_{\sigma\xi + \mu}(t) &= e^{it\mu} f_\xi(\sigma t), \quad t \in \mathcal{R}^1, \text{ где } \sigma, \mu = \text{const.} \end{aligned} \quad (11.4)$$

Доказательство. $f_{\xi_1 + \dots + \xi_n}(t) \equiv M e^{it(\xi_1 + \dots + \xi_n)} = M(e^{it\xi_1} \dots e^{it\xi_n}) = M e^{it\xi_1} M e^{it\xi_2} \dots M e^{it\xi_n} = f_{\xi_1}(t) f_{\xi_2}(t) \dots f_{\xi_n}(t)$. Здесь мы воспользовались теоремой о математическом ожидании произведения независимых случайных величин. Независимость $e^{it\xi_1}, \dots, e^{it\xi_n}$ следует из независимости $\xi_1, \dots, \xi_n$ (см. §8) (тот факт, что $\eta_k = e^{it\xi_k}$ — комплексные случайные величины, очевидно, не имеет значения). Что касается второго равенства в (11.4), то

$$f_{\sigma\xi + \mu}(t) = e^{it\mu} M e^{it\sigma\xi} = e^{it\mu} f_\xi(\sigma t), \quad t \in \mathcal{R}^1. \blacksquare$$

В лемме 1.11.2 сформулированы основные свойства характеристических функций, которые используются при доказательстве центральных предельных теорем.

Лемма 1.11.2. Если существует момент $M\xi^k$, то k -я производная $f_\xi^{(k)}(\cdot)$ х. ф. $f_\xi(\cdot)$ существует, равномерно непрерывна на $\mathcal{R}^1$ и

$$f_\xi^{(k)}(0) = i^k M\xi^k. \quad (11.5)$$

Доказательство. Дифференцируя левую и правую части равенства (1.11.1) и полагая $t = 0$, получим равенство

$$f_{\xi}^{(k)}(0) = \int_{-\infty}^{\infty} e^{itx} (ix)^k dF_{\xi}(x)|_{t=0} = i^k M_{\xi}^k. \quad (11.6)$$

Дифференцирование под знаком интеграла законно, так как из существования M_{ξ}^k следует, что $\int_{-\infty}^{\infty} x^k e^{itx} dF_{\xi}(x)$ сходится равномерно по $t \in \mathcal{R}^1$ (и абсолютно). Докажем равномерную непрерывность на $\mathcal{R}^1$ k -й производной $f_{\xi}^{(k)}(\cdot)$

$$\begin{aligned} |f_{\xi}^{(k)}(t+h) - f_{\xi}^{(k)}(t)| &\leq \int_{-\infty}^{\infty} |x|^k |e^{ihx} - 1| dF_{\xi}(x) = \\ &= \int_{|x|>A} |x|^k |e^{ihx} - 1| dF_{\xi}(x) + \int_{-A}^A |x|^k |e^{ihx} - 1| dF_{\xi}(x) \leq \\ &\leq 2 \int_{|x|>A} |x|^k dF_{\xi}(x) + \int_{-A}^A |x|^k |e^{ihx} - 1| dF_{\xi}(x). \end{aligned} \quad (11.7)$$

Пусть $\varepsilon > 0$ — любое. В силу существования момента k -го порядка выбором достаточно большого $A > 0$ первый интеграл в правой части (11.7) может быть сделан меньше $\varepsilon/2$. Зафиксировав это A , возьмем $\delta > 0$ столь малым, чтобы для всех h , $|h| \leq \delta$, второй интеграл в правой части (11.7) также был меньше $\varepsilon/2$ (что возможно ввиду оценки $\max_{x \in [-A, A]} |e^{ihx} - 1| = \max_{x \in [-A, A]} \left| \int_0^{hx} e^{it} dt \right| \leq |hA|$). Таким образом, $|f_{\xi}^{(k)}(t+h) - f_{\xi}^{(k)}(t)| < \varepsilon$ для всех $t \in \mathcal{R}^1$, если $|h| \leq \delta$. ■

Из доказанной леммы следует, в частности, что х.ф. $f_{\xi}(\cdot)$ равномерно непрерывна на $\mathcal{R}^1$.

Теорема 1.11.1. Если характеристическая функция $f_{\xi}(\cdot)$ абсолютно интегрируема на $\mathcal{R}^1$, то случайная величина ξ абсолютно непрерывна, а ее плотность распределения

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} f_{\xi}(t) dt, \quad x \in \mathcal{R}^1, \quad (11.8)$$

и равномерно непрерывна на $\mathcal{R}^1$.

Доказательство. Пусть x и y , $y > x$, любые два числа. Рассмотрим интеграл

$$j(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} f_{\xi}(t) dt = \frac{1}{2\pi} \lim_{A \rightarrow \infty} \int_{-A}^A \frac{e^{-itx} - e^{-ity}}{it} f_{\xi}(t) dt. \quad (11.9)$$

Интеграл (11.9) сходится абсолютно ввиду абсолютной интегрируемости $f_{\xi}(\cdot)$ и оценки

$$|e^{-itx} - e^{-ity}| = \left| \int_{ty}^{tx} e^{-i\alpha} d\alpha \right| \leq |t(x - y)|. \quad (11.10)$$

Оценка (11.10) показывает также, что если $y = x + h$, $h > 0$, то

$$j(x, x + h) \rightarrow 0 \text{ при } h \rightarrow 0. \quad (11.11)$$

Подставив в (11.9) вместо $f_{\xi}(t)$ выражение (1.11.1) и изменив порядок интегрирования, что возможно в силу абсолютной сходимости двойного интеграла, получим

$$\begin{aligned} j(x, y) &= \frac{1}{2\pi} \lim_{A \rightarrow \infty} \int_{-\infty}^{\infty} \left(\int_{-A}^A \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dt \right) dF_{\xi}(z) = \\ &= \frac{1}{\pi} \lim_{A \rightarrow \infty} \int_{-\infty}^{\infty} dF_{\xi}(z) \int_0^A \left\{ \frac{e^{it(z-x)} - e^{-it(z-x)}}{2i} - \frac{e^{it(z-y)} - e^{-it(z-y)}}{2i} \right\} \frac{dt}{t} = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} dF_{\xi}(z) \left[\int_0^{\infty} \frac{\sin(z-x)t}{t} dt - \int_0^{\infty} \frac{\sin(z-y)t}{t} dt \right]. \end{aligned} \quad (11.12)$$

Последнее равенство в (11.12) есть следствие формулы

$$\lim_{A \rightarrow \infty} j_0 \equiv \lim_{A \rightarrow \infty} \int_{-\infty}^{\infty} dF_{\xi}(z) \int_A^{\infty} \frac{\sin \alpha t}{t} dt = 0, \text{ где } \alpha = z - x \text{ или } \alpha = z - y.$$

Для доказательства этой формулы будем считать, что x и y — точки непрерывности $F_{\xi}(\cdot)$ (см. §1.8.3) и воспользуемся элементарной оценкой:

$$\left| \int_A^{\infty} \frac{\sin \alpha t}{t} dt \right| \leq \begin{cases} c_0, & \text{если } A \geq 0, \alpha - \text{любое;} \\ 2/(\delta A), & \text{если } A > 0, |\alpha| \geq \delta > 0. \end{cases}$$

Пусть $\alpha = z - x$ ($\alpha = z - y$ — аналогично). Тогда

$$|j_0| = \left| \int_{|z-x| \leq \delta} dF_\xi(z) \int_A^\infty \frac{\sin \alpha t}{t} dt + \int_{|z-x| > \delta} dF_\xi(z) \int_A^\infty \frac{\sin \alpha t}{t} dt \right| \leq \\ \leq c_0[F_\xi(x + \delta) - F_\xi(x - \delta)] + \frac{2}{\delta A} \int_{-\infty}^\infty dF_\xi(z) < \varepsilon,$$

ибо первое слагаемое $< \varepsilon/2$, если $\delta \leq \delta_0(\varepsilon)$, ввиду непрерывности $F_\xi(\times \times)$ в точке x , а второе слагаемое $< \varepsilon/2$, если A достаточно велико ($A \geq 4\varepsilon^{-1}\delta_0^{-1}(\varepsilon)$, $\int_{-\infty}^\infty dF_\xi(z) = 1$). Как известно,

$$\int_0^\infty \frac{\sin \alpha t}{t} dt = \frac{\pi}{2} \operatorname{sgn} \alpha. \quad (11.13)$$

Поэтому квадратная скобка в правой части (11.12) равна нулю, если $z < x$ или $z > y$, и равна π , если $x < z < y$. Таким образом,

$$j(x, y) = \int_x^y dF_\xi(z). \quad (11.14)$$

Отсюда в силу (11.11) заключаем, что $f_\xi(\cdot)$ — непрерывная функция, и (11.14) дает

$$j(x, y) = F_\xi(y) - F_\xi(x), \quad x, y, y > x \text{ — любые}. \quad (11.15)$$

Из формулы (11.15) вытекает равенство

$$\frac{F_\xi(x + h) - F_\xi(x - h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^\infty \frac{\sin th}{th} e^{-itx} f_\xi(t) dt. \quad (11.16)$$

Покажем, что предел правой части (11.16) при $h \rightarrow 0$ равен

$$\frac{1}{2\pi} \int_{-\infty}^\infty e^{-itx} f_\xi(t) dt. \quad (11.17)$$

В самом деле, пусть $\varepsilon > 0$ — любое, тогда

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} \frac{\sin th}{th} e^{-itx} f_{\xi}(t) dt - \int_{-\infty}^{\infty} e^{-itx} f_{\xi}(t) dt \right| \leq \\ & \leq 2 \int_{|t| > A} |f_{\xi}(t)| dt + \int_{-A}^A \left| \frac{\sin th}{th} - 1 \right| |f_{\xi}(t)| dt < \varepsilon \end{aligned} \quad (11.18)$$

при $|h| \leq \delta(\varepsilon)$, поскольку первый интеграл в правой части (11.18) $< \varepsilon/4$, если $A > 0$ достаточно велико (в силу абсолютной интегрируемости $f_{\xi}(\cdot)$), а второй интеграл $< \varepsilon/2$, поскольку при фиксированном A $|(\sin th)/th - 1| \rightarrow 0$ при $h \rightarrow 0$ равномерно по $t \in [-A, A]$. Поэтому существует и предел левой части в (11.16), и мы получаем

$$p(x) = F'_{\xi}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f_{\xi}(t) dt, \quad x \in \mathcal{R}^1. \quad (11.19)$$

Из представления (11.19) ввиду абсолютной интегрируемости $f_{\xi}(\cdot)$ следует равномерная непрерывность $p(\cdot)$ на $\mathcal{R}^1$ (ср. с леммой 1.11.1). ■

Замечание 1.11.1. Из проведенного доказательства теоремы 1.11.1 вытекает следующее утверждение: если $F_{\xi}(\cdot)$ — функция распределения, а $f_{\xi}(\cdot)$ — х.ф. случайной величины ξ , то для любых точек непрерывности x и y функции $F_{\xi}(\cdot)$ справедливо равенство

$$F_{\xi}(y) - F_{\xi}(x) = \frac{1}{2\pi} \lim_{A \rightarrow \infty} \int_{-A}^A \frac{e^{-itx} - e^{-ity}}{it} f_{\xi}(t) dt. \quad (11.20)$$

Равенство (11.20) называется *формулой обращения*, оно позволяет находить функцию распределения по известной х.ф.

Рассмотрим несколько *примеров* х.ф.

Пример 1.11.1. Пусть ξ — случайная величина, $P(\xi = m) = 1$, тогда $f_{\xi}(t) = M e^{itm} = e^{itm}$, $-\infty < t < \infty$.

Пример 1.11.2. Нормальное распределение $N(0, 1)$ имеет х.ф.

$$f_{\xi}(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(it-x)^2}{2}} dx = e^{-\frac{t^2}{2}}. \quad (11.21)$$

Если $\xi \sim N(0, 1)$, то $(\sigma\xi + \mu) \sim N(\mu, \sigma^2)$ и согласно лемме 1.11.1

$$f_{\sigma\xi+\mu}(t) = e^{it\mu - \sigma^2 t^2/2}.$$

Так как $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$, то для $\xi \sim N(0, 1)$, как нетрудно проверить, $f_{\xi^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx^2} e^{-x^2/2} dx = (1 - 2it)^{-1/2} = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y^{-1/2} e^{-y/2} e^{ity} dy$. Поэтому для $\chi_n^2 = \xi_1^2 + \dots + \xi_n^2$, где $\xi_k \sim N(0, 1)$, $k = 1, \dots, n$, и $\xi_1, \dots, \xi_n$ взаимно независимы, $f_{\chi_n^2}(t) = (1 - 2it)^{-n/2}$, $t \in \mathcal{R}^1$.

Пример 1.11.3. Х. ф. распределения Пуассона

$$f_{\xi}(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} (\lambda e^{it})^k \frac{1}{k!} = e^{-\lambda} e^{\lambda e^{it}}, \quad -\infty < t < \infty. \quad (11.22)$$

Пример 1.11.4. Х. ф. биномиального распределения

$$f_{\xi}(t) = \sum_{k=0}^n e^{itk} C_n^k p^k q^{n-k} = \sum_{k=0}^n C_n^k (pe^{it})^k q^{n-k} = (pe^{it} + q)^n, \quad -\infty < t < \infty. \quad (11.23)$$

Пример 1.11.5. Характеристическая функция суммы случайного числа случайных величин. Пусть ν — дискретная случайная величина, принимающая значения $0, 1, \dots$ $P(\nu = k) = p_k$, $k = 0, 1, 2, \dots$, $\sum_{k=0}^{\infty} p_k = 1$, $\nu, \xi_1, \xi_2, \dots$ — независимые в совокупности случайные величины и $\eta = \xi_1 + \xi_2 + \dots + \xi_{\nu}$. Тогда $f_{\eta}(t) = M \exp(it(\xi_1 + \dots + \xi_{\nu})) = p_0 + \sum_{k=1}^{\infty} M(\exp(it(\xi_1 + \dots + \xi_k | \nu = k))) p_k = p_0 + \sum_{k=1}^{\infty} p_k \prod_{j=1}^k f_{\xi_j}(t)$, $-\infty < t < \infty$. Если $\xi_1, \xi_2, \dots$ — одинаково распределены, т.е. являются взаимно независимыми копиями случайной величины ξ , то $f_{\eta}(t) = \sum_{k=0}^{\infty} (f_{\xi}(t))^k p_k$, $-\infty < t < \infty$, а если ν — пуассоновская случайная величина, то $f_{\eta}(t) = \sum_{k=0}^{\infty} [(f_{\xi}(t))^k \lambda^k / k!] e^{-\lambda} = e^{-\lambda} e^{\lambda f_{\xi}(t)}$, $-\infty < t < \infty$.

В заключение приведем основные свойства многомерных характеристических функций.

Определение 1.11.2. Характеристической функцией n -мерного случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется функция

$$\begin{aligned} f_{\xi}(t_1, \dots, t_n) &= M e^{i(t_1 \xi_1 + \dots + t_n \xi_n)} = \\ &= \int \dots \int_{-\infty}^{\infty} e^{i \sum_{k=1}^n t_k x_k} dF_{\xi}(x_1, \dots, x_n), \quad (t_1, \dots, t_n) \in \mathcal{R}^n. \end{aligned} \quad (11.24)$$

Из этого определения непосредственно вытекают следующие свойства х. ф. $f_{\xi}(\cdot, \dots, \cdot)$:

- 1) $f_{\xi}(0, \dots, 0) = 1$, $|f_{\xi}(t_1, \dots, t_n)| \leq 1$, $(t_1, \dots, t_n) \in \mathcal{R}^n$;

2) если координаты $\xi_1, \dots, \xi_n$ случайного вектора ξ взаимно независимы, то

$f_\xi(t_1, \dots, t_n) = f_{\xi_1}(t_1) \dots f_{\xi_n}(t_n)$, ибо в этом случае

$$Me^{i(t_1\xi_1 + \dots + t_n\xi_n)} = Me^{it_1\xi_1} \dots Me^{it_n\xi_n}, \quad (t_1, \dots, t_n) \in \mathcal{R}^n;$$

3) х. ф. случайного вектора $\eta = (\sigma_1\xi_1 + \mu_1, \dots, \sigma_n\xi_n + \mu_n)$ равна

$$f_\eta(t_1, \dots, t_n) = e^{i \sum_{k=1}^n \mu_k t_k} f_\xi(\sigma_1 t_1, \dots, \sigma_n t_n), \quad (t_1, \dots, t_n) \in \mathcal{R}^n.$$

4) Х. ф. суммы координат случайного вектора ξ равна

$$f_{\xi_1 + \dots + \xi_n}(t) = f_\xi(t, \dots, t), \quad \text{ибо } f_{\xi_1 + \dots + \xi_n}(t) = Me^{it(\xi_1 + \dots + \xi_n)}, \quad t \in \mathcal{R}^1.$$

В полной аналогии с одномерным случаем для х. ф. многомерных случайных величин могут быть доказаны формула обращения и теорема о непрерывности х. ф. (см. ниже).

1.11.2. Теорема о непрерывности. Хорошо известно из анализа, что для финитных ¹⁾ функций справедливы формулы прямого и обратного преобразования Фурье: если $\varphi(x) \in C_0^\infty(\mathcal{R}^1)$, то

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \tilde{\varphi}(t) dt, \quad x \in \mathcal{R}^1, \quad (11.25)$$

где $\tilde{\varphi}(\cdot)$ — преобразование Фурье функции $\varphi(\cdot)$

$$\tilde{\varphi}(t) = \int_{-\infty}^{\infty} e^{-itx} \varphi(x) dx, \quad t \in \mathcal{R}^1; \quad (11.26)$$

$\tilde{\varphi}(\cdot)$ убывает на бесконечности быстрее любой степени и поэтому абсолютно интегрируема на $\mathcal{R}^1$.

Лемма 1.11.3. Пусть $\varphi(\cdot)$ — произвольная финитная функция, $\varphi(\cdot) \in C_0^\infty(\mathcal{R}^1)$, ξ — случайная величина, а $f_\xi(\cdot)$ — ее х. ф. Тогда

$$M\varphi(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\xi(t) \tilde{\varphi}(t) dt. \quad (11.27)$$

Доказательство. Используя представление (11.25), найдем $M\varphi(\xi) = \int_{-\infty}^{\infty} \varphi(x) dF_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dF_\xi(x) \int_{-\infty}^{\infty} e^{itx} \tilde{\varphi}(t) dt =$

¹⁾ Финитной называется бесконечно дифференцируемая функция с компактным носителем. Множество всех финитных на $\mathcal{R}^1$ функций обозначается $C_0^\infty(\mathcal{R}^1)$.

$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\varphi}(t) dt \int_{-\infty}^{\infty} e^{itx} dF_{\xi}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\varphi}(t) f_{\xi}(t) dt$, перестановка интегралов законна в силу абсолютной сходимости двойного интеграла. ■

Лемма 1.11.4. Пусть последовательность $\{f_{\xi_1}(\cdot), f_{\xi_2}(\cdot), \dots\}$ х.ф. последовательности случайных величин $\{\xi_1, \xi_2, \dots\}$ сходится при $n \rightarrow \infty$ к х.ф. случайной величины ξ равномерно по t в каждом конечном интервале $|t| \leq T$. Тогда для любой финитной функции $\varphi(\cdot) \in C_0^{\infty}(\mathcal{R}^1)$

$$M\varphi(\xi_n) \rightarrow M\varphi(\xi) \text{ при } n \rightarrow \infty. \quad (11.28)$$

Доказательство. С помощью формулы (11.27) получим

$$\begin{aligned} |M\varphi(\xi_n) - M\varphi(\xi)| &= \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} [f_{\xi_n}(t) - f_{\xi}(t)] \tilde{\varphi}(t) dt \right| \leq \\ &\leq \frac{1}{2\pi} \int_{-T}^T |f_{\xi_n}(t) - f_{\xi}(t)| |\tilde{\varphi}(t)| dt + \frac{1}{2\pi} \int_{|t|>T} |f_{\xi_n}(t)| |\tilde{\varphi}(t)| dt + \\ &\quad + \frac{1}{2\pi} \int_{|t|>T} |f_{\xi}(t)| |\tilde{\varphi}(t)| dt. \end{aligned} \quad (11.29)$$

Второй и третий интегралы меньше $\varepsilon/3$, если $T \geq T_0(\varepsilon) > 0$, ибо $|f_{\xi_n}(t)| \leq 1$, $t \in \mathcal{R}^1$, а $\int_{|t|>T} |\tilde{\varphi}(t)| dt < \frac{\varepsilon}{3}$, так как $\tilde{\varphi}(\cdot)$ абсолютно интегрируема.

Фиксировав такое T , видим, что первый интеграл также меньше $\varepsilon/3$, если $n \geq n_0(\varepsilon)$, так как $f_{\xi_n}(t) \rightarrow f_{\xi}(t)$ при $n \rightarrow \infty$ равномерно по t при $|t| \leq T$ согласно условию леммы. ■

Теорема 1.11.2. (о непрерывности). Пусть выполнены условия леммы 1.11.4, и пусть, кроме того, функция распределения $F_{\xi}(\cdot)$ случайной величины ξ непрерывна. Тогда при $n \rightarrow \infty$ равномерно по¹⁾ $x \in \mathcal{R}^1$

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x). \quad (11.30)$$

Доказательство. Возьмем любой сегмент $[a, b]$ и любое $\varepsilon > 0$. Существует финитная функция $\varphi(x)$, $x \in \mathcal{R}^1$, равная нулю вне $[a, b]$, равная 1 на $[a + \varepsilon, b - \varepsilon]$ и $0 \leq \varphi(x) \leq 1$ в остальных точках, $x \in \mathcal{R}^1$.

При всех n $M\varphi(\xi_n) = \int_a^b \varphi(x) dF_{\xi_n}(x) \leq \int_a^b dF_{\xi_n}(x) = F_{\xi_n}(b) - F_{\xi_n}(a)$.

Отсюда, поскольку на основании леммы 1.11.4 $M\varphi(\xi_n) \rightarrow M\varphi(\xi)$ при

¹⁾ В этом случае говорят, что последовательность $\xi_1, \xi_2, \dots$ сходится к ξ по распределению, или слабо, см. определение 1.10.3.

$n \rightarrow \infty$, находим

$$\begin{aligned} \lim_{n \rightarrow \infty} [F_{\xi_n}(b) - F_{\xi_n}(a)] &\geq \lim_{n \rightarrow \infty} M\varphi(\xi_n) = M\varphi(\xi) = \\ &= \int_a^b \varphi(x) dF_\xi(x) \geq \int_{a+\varepsilon}^{b-\varepsilon} dF_\xi(x) = F_\xi(b-\varepsilon) - F_\xi(a+\varepsilon). \end{aligned} \quad (11.31)$$

Аналогично, взяв финитную функцию $\bar{\varphi}(\cdot)$, равную 1 на $[a, b]$, равную нулю вне $[a-\varepsilon, b+\varepsilon]$ и $0 \leq \bar{\varphi}(x) \leq 1$ — в остальных точках $x \in \mathcal{R}^1$, получим: $F_{\xi_n}(b) - F_{\xi_n}(a) = \int_a^b dF_{\xi_n}(x) \leq \int_{a-\varepsilon}^{b+\varepsilon} \bar{\varphi}(x) dF_{\xi_n}(x) = M\bar{\varphi}(\xi_n)$. Отсюда, вновь используя лемму 1.11.4, найдем

$$\begin{aligned} \overline{\lim}_{n \rightarrow \infty} [F_{\xi_n}(b) - F_{\xi_n}(a)] &\leq \lim_{n \rightarrow \infty} M\bar{\varphi}(\xi_n) = M\bar{\varphi}(\xi) = \\ &= \int_{a-\varepsilon}^{b+\varepsilon} \bar{\varphi}(x) dF_\xi(x) \leq \int_{a-\varepsilon}^{b+\varepsilon} dF_\xi(x) = F_\xi(b+\varepsilon) - F_\xi(a-\varepsilon). \end{aligned} \quad (11.32)$$

Поскольку $F_\xi(\cdot)$ непрерывна в каждой точке $x \in \mathcal{R}^1$, то, устремив $\varepsilon \rightarrow 0$ в (11.31) и (11.32), получим

$$\lim_{n \rightarrow \infty} [F_{\xi_n}(b) - F_{\xi_n}(a)] = \overline{\lim}_{n \rightarrow \infty} [F_{\xi_n}(b) - F_{\xi_n}(a)] = F_\xi(b) - F_\xi(a). \quad (11.33)$$

Из совпадения верхнего и нижнего пределов в (11.33) следует существование предела и равенство

$$\lim_{n \rightarrow \infty} [F_{\xi_n}(b) - F_{\xi_n}(a)] = F_\xi(b) - F_\xi(a). \quad (11.34)$$

Из (11.34), в свою очередь, вытекает, что в каждой точке $x \in \mathcal{R}^1$ $F_{\xi_n} \rightarrow F_\xi(x)$ при $n \rightarrow \infty$. В самом деле, пусть $\delta > 0$ — любое. Возьмем числа A и B так, чтобы

$$F_\xi(A) - F_\xi(B) \geq 1 - \delta \quad (11.35)$$

(числа A и B найдутся, ибо $F_\xi(+\infty) = 1$, $F_\xi(-\infty) = 0$ и $F_\xi(\cdot)$ непрерывна). В силу (11.34) найдется такой номер $n_0(\delta)$, что при $n \geq n_0(\delta)$

$$F_{\xi_n}(A) - F_{\xi_n}(B) \geq 1 - 2\delta. \quad (11.36)$$

Из (11.35) и (11.36) ввиду неравенств $0 \leq F_\xi(A)$, $F_{\xi_n}(A) \leq 1$ следует, что

$$F_\xi(B) \leq \delta, \quad F_{\xi_n}(B) \leq 2\delta. \quad (11.37)$$

Теперь

$$\begin{aligned} |F_\xi(x) - F_{\xi_n}(x)| &\leq |[F_\xi(x) - F_\xi(B)] - \\ &- [F_{\xi_n}(x) - F_{\xi_n}(B)]| + F_\xi(B) + F_{\xi_n}(B) \leq 4\delta, \end{aligned} \quad (11.38)$$

так как первое слагаемое справа $< \delta$ в силу (11.34), если $n \geq n_1(\delta)$, а для двух других слагаемых выполнено (11.37). В (11.38) $\delta > 0$ — любое, так что мы доказали, что при каждом $x \in \mathcal{R}^1$

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x). \quad (11.39)$$

Докажем, наконец, что сходимость в (11.39) равномерная по x на всей прямой $-\infty < x < \infty$. Пусть $\varepsilon > 0$ — любое. Выберем точки $x_1 < x_2 < \dots < x_N$, $N = [1/\varepsilon] + 1$, так, чтобы

$$F_{\xi}(x_{i+1}) - F_{\xi}(x_i) \leq \varepsilon, \quad 1 - F_{\xi}(x_N) \leq \varepsilon, \quad i = 1, 2, \dots, N. \quad (11.40)$$

Точек $x_1, \dots, x_N$ конечное число, поэтому в силу (11.39) найдется такое $n_2 = n_2(\varepsilon)$, что при $n \geq n_2$

$$|F_{\xi}(x_j) - F_{\xi_n}(x_j)| \leq \varepsilon, \quad j = 1, 2, \dots, N. \quad (11.41)$$

Пусть x — произвольная точка $\mathcal{R}^1$, тогда либо $x < x_1$, либо $x > x_N$, либо при некотором $j \in \{1, \dots, N-1\}$ $x_j \leq x \leq x_{j+1}$. В последнем случае

$$|F_{\xi}(x) - F_{\xi_n}(x)| \leq |F_{\xi}(x) - F_{\xi}(x_j)| + |F_{\xi}(x_j) - F_{\xi_n}(x_j)| + |F_{\xi_n}(x_j) - F_{\xi_n}(x)| \leq 2\varepsilon + |F_{\xi_n}(x) - F_{\xi_n}(x_j)| \quad (11.42)$$

в силу (11.40), (11.41) и монотонности $F_{\xi}(x)$. Далее,

$$|F_{\xi_n}(x) - F_{\xi_n}(x_j)| \leq |F_{\xi_n}(x_{j+1}) - F_{\xi_n}(x_j)| \leq |F_{\xi_n}(x_{j+1}) - F_{\xi}(x_{j+1})| + |F_{\xi}(x_{j+1}) - F_{\xi}(x_j)| + |F_{\xi}(x_j) - F_{\xi_n}(x_j)| \leq 3\varepsilon \quad (11.43)$$

Из (11.42) и (11.43)

$$|F_{\xi}(x) - F_{\xi_n}(x)| \leq 5\varepsilon.$$

Аналогичная оценка верна, очевидно, и в случае $x < x_1$ или $x > x_N$ ввиду (11.39) и (11.40). Тем самым доказана равномерность сходимости в (11.39) на всей прямой $\mathcal{R}^1$. ■

Справедливо и обратное утверждение: если в каждой точке $x \in \mathcal{R}^1$ непрерывности $F_{\xi}(\cdot)$ выполнено условие $F_{\xi}(x) = \lim_{n \rightarrow \infty} F_{\xi_n}(x)$, то равномерно по $t \in \mathcal{R}^1$ $f_{\xi}(t) = \lim_{n \rightarrow \infty} f_{\xi_n}(t)$. Доказательства не приводим, так как эта теорема в дальнейшем не используется.

Пример 1.11.6. Пусть $\xi, \xi_1, \xi_2, \dots$ — последовательность взаимно независимых одинаково распределенных случайных величин (копий случайной величины ξ), $M\xi = m$ и $f_{\xi}(t) = Me^{it\xi}$, $-\infty < t < \infty$, — х. ф. ξ . Если $\eta_n = \sum_{j=1}^n \xi_j$, то при $n \rightarrow \infty$ $f_{\eta_n/n}(t) \rightarrow e^{itm}$, $-\infty < t < \infty$.

Действительно, при $n \rightarrow \infty$ $f_{\eta_n/n}(t) = Me^{it\eta_n/n} = (f_{\xi}(t/n))^n = (1 + f'_{\xi}(0)(t/n) + o(t/n))^n = (1 + itm/n + o(t/n))^n \rightarrow e^{itm}$, $-\infty < t < \infty$. Здесь использованы леммы 1.11.1, 1.11.2, причем согласно лемме 1.11.2 $f'(\cdot)$ — непрерывная функция, и это позволило в разложении $f_{\xi}(t/n)$ в окрестности $t = 0$ использовать остаточный член в форме Пеано.

Согласно теореме 1.11.2 последовательность функций распределения $\eta_n/n = (\xi_1 + \dots + \xi_n)/n$ $F_{\eta_n/n}(x) \xrightarrow{n \rightarrow \infty} F_\mu(x)$, $-\infty < x < \infty$, где $F_\mu(\cdot)$ — функция распределения случайной величины μ , равной m с вероятностью единица, $P(\mu = m) = 1$, т. е. в терминах слабой сходимости, см. определение 1.10.3, при $n \rightarrow \infty$ последовательность $(\xi_1 + \dots + \xi_n)/n$ по распределению (слабо) сходится к $\text{const} = m = M\xi = M\xi_j$, $j = 1, 2, \dots$, $(\xi_1 + \dots + \xi_n)/n \xrightarrow{d} M\xi = m$. Этот факт слабее, чем закон больших чисел¹⁾, согласно которому при $n \rightarrow \infty$ $(\xi_1 + \dots + \xi_n)/n \xrightarrow{P} m$, но в рассматриваемом случае слабой сходимости к const слабая сходимость эквивалентна сходимости по вероятности, см. теорему 1.10.9.

Заметим, что из доказанной теоремы о непрерывности следует, что если у двух законов распределения совпадают характеристические функции, то совпадают и сами законы распределения (надо взять $f_{\xi_n}(t) \equiv f_\xi(t)$ при всех n). Приведем несколько важных следствий этого замечания.

Следствие 1.11.1. Сумма $\eta = \xi_1 + \xi_2$, где случайные величины $\xi_1 \sim N(\mu_1, \sigma_1^2)$, $\xi_2 \sim N(\mu_2, \sigma_2^2)$ и независимы, имеет нормальное распределение $\eta \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Действительно, $f_{\xi_1}(t) = e^{i\mu_1 t - \frac{\sigma_1^2 t^2}{2}}$, $f_{\xi_2}(t) = e^{i\mu_2 t - \frac{\sigma_2^2 t^2}{2}}$, $t \in \mathcal{R}^1$, и так как ξ_1 и ξ_2 независимы, то $f_\eta(t) = f_{\xi_1}(t)f_{\xi_2}(t) = e^{i(\mu_1 + \mu_2)t - \frac{\sigma_1^2 + \sigma_2^2}{2}t^2}$, $t \in \mathcal{R}^1$, поэтому $\eta \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Следствие 1.11.2. Сумма $\eta = \xi_1 + \xi_2$, где случайные величины ξ_1 и ξ_2 распределены по закону Пуассона с параметрами λ_1 и λ_2 и независимы, также распределена по закону Пуассона с параметром $\lambda_1 + \lambda_2$.

В самом деле, $f_{\xi_1}(t) = e^{\lambda_1(e^{it} - 1)}$, $f_{\xi_2}(t) = e^{\lambda_2(e^{it} - 1)}$, $t \in \mathcal{R}^1$, и так как ξ_1 и ξ_2 независимы, то $f_\eta(t) = f_{\xi_1}(t)f_{\xi_2}(t) = e^{(\lambda_1 + \lambda_2)(e^{it} - 1)}$, $t \in \mathcal{R}^1$. Отсюда и следует высказанное утверждение.

Следствие 1.11.3. Для распределения $\chi_n^2 = \xi_1^2 + \dots + \xi_n^2$, где $\xi_k \sim N(0, 1)$, $k = 1, \dots, n$, и слагаемые взаимно независимы, согласно примеру 1.11.2 $f_{\chi_n^2}(t) = (1 - 2it)^{-n/2}$, $t \in \mathcal{R}^1$. Поскольку $f_{\xi_k^2}(t) = (1 -$

$$- 2it)^{-1/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx^2} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{y} e^{-y/2} e^{ity} dy, \quad t \in \mathcal{R}^1, \text{ то,}$$

судя по подынтегральному выражению, плотность распределения χ_n^2 имеет вид: $p_{\chi_n^2}(y) = cy^{n/2-1}e^{-y/2}$, $0 \leq y < \infty$. Множитель c определя-

¹⁾ Из сходимости по вероятности следует сходимость по распределению, но в общем случае обратное неверно

ется из условия нормировки $\int_0^\infty p_{\chi_n^2}(y)dy = 1$ и равен $= 2^{-\frac{n}{2}} \frac{1}{\Gamma(\frac{n}{2})}$ (см.

Приложение).

1.11.3. Центральные предельные теоремы. Интуитивное понимание содержания центральных предельных теорем может быть получено следующим образом. Рассмотрим суммы

$$\eta_n = \xi_1 + \dots + \xi_n, \quad n = 1, 2, \dots, \quad (11.44)$$

взаимно независимых случайных величин, которые принимают целочисленные значения и все имеют распределение $P\{\xi_i = m\} = p_m$, $m = 0, \pm 1, \dots$, для всех $i = 1, 2, \dots$. Это распределение изобразим прямоугольниками, рис. 1.17, основание каждого из которых равно единице, высота — p_m , его площадь равна p_m , $m = 0, \pm 1, \pm 2, \dots$, и $\sum_{-\infty}^\infty p_m = 1$. В общем случае получим практически произвольный набор прямоугольников.

Рассмотрим теперь вместо η_n нормированные случайные величины

$$\eta_n^* = \frac{\eta_n - M\eta_n}{\sqrt{D\eta_n}} = \frac{\eta_n - \mu n}{\sigma\sqrt{n}}, \quad n = 1, 2, \dots, \quad (11.45)$$

где $\mu = M\xi_i$ и $\sigma^2 = D\xi_i$, $i = 1, 2, \dots$. Значениями случайной величины η_n^* являются числа $x_n(m) = (m - n\mu)/\sigma\sqrt{n}$, причем $P(\{\eta_n^* = x_n(m)\}) = P(\{\eta_n = m\})$.

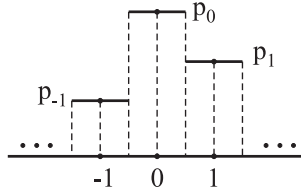


Рис. 1.17. Распределение $P(\xi_i = m) = p_m$, $m = 0, \pm 1, \dots$, $i = 1, 2, \dots$

Построим теперь аналогичный «график» распределения η_n^* . По оси абсцисс отложим значения $x_n(m)$, $m = 0, \pm 1, \dots$, и, как и на рис. 1.18, построим прямоугольники, площади которых равны $P(\{\eta_n^* = x_n(m)\})$, $m = 1, 2, \dots$. Поскольку длина основания этих прямоугольников теперь равна $1/\sigma\sqrt{n}$, то их высоты должны быть равны $P(\{\eta_n^* = x_n(m)\})\sigma\sqrt{n} = P(\{\eta_n = m\})\sigma\sqrt{n}$, $m = 0, \pm 1, \dots$. При достаточно большом n окажется ("чудо" Лапласа), что верхние основания прямоугольников почти точно лягут на фиксированную кривую

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (\text{рис. 1.18}), \quad \text{т. е. для } m = 1, 2, \dots$$

$$\sigma\sqrt{n} P(\{\eta_n^* = x_n(m)\}) \approx \frac{1}{\sqrt{2\pi}} e^{-x_n^2(m)/2} \quad \text{при достаточно больших } n.$$

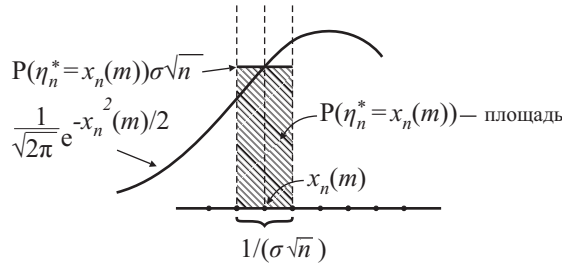


Рис. 1.18. Распределение высот $P(\eta_n^* = x_n(m))\sigma\sqrt{n}$, $m = 0, \pm 1, \dots$, заштрихованных прямоугольников.

При этом естественно, что при достаточно больших n

$$\begin{aligned} P(\{a \leq \eta_n^* \leq b\}) &= \sum_{a \leq x_n \leq b} P(\{\eta_n^* = x_n(m)\}) \approx \\ &\approx \sum_{a \leq x_n \leq b} \frac{1}{\sqrt{2\pi}} e^{-x_n^2(m)/2} \frac{1}{\sigma\sqrt{n}} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx, \end{aligned} \quad (11.46)$$

где при замене суммы на интеграл мы считали, что $\Delta x = 1/\sigma\sqrt{n}$. Этот факт и составляет, по существу, содержание центральных предельных теорем (которые отличаются друг от друга формулировками различных математических условий, обеспечивающих указанное выше утверждение).

Рассмотрим вначале центральную предельную теорему в ее простейшей форме для случая одинаково распределенных случайных величин, а затем приведем достаточно общую центральную предельную теорему Ляпунова.

Определение 1.11.3. Последовательность случайных величин ξ_k , $k = 1, 2, \dots$, сходится к случайной величине ξ_0 по распределению или слабо, $\xi_k \xrightarrow{d} \xi_0$, $n \rightarrow \infty$, если

$$\lim_{k \rightarrow \infty} F_k(x) = F_0(x) \quad (11.47)$$

в каждой точке $x \in \mathcal{R}^1$ непрерывности $F_0(x)$, где $F_k(x)$ — функция распределения случайной величины ξ_k , $k = 0, 1, 2, \dots$

С понятием сходимости по распределению мы уже встречались в теореме 1.11.2, см. определение 1.10.3, теорему 1.10.8 и пример 1.11.6. **Теорема 1.11.3 (центральная предельная теорема).** Пусть $\{\xi_1, \xi_2, \dots\}$ — последовательность взаимно независимых и одинаково распределенных случайных величин, $M\xi_n = \mu$, $D\xi_n = \sigma^2$, $\sigma > 0$, $n = 1, 2, \dots$. Тогда последовательность

$$\eta_n = \frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma\sqrt{n}}, \quad n = 1, 2, \dots, \quad (11.48)$$

сходится по распределению к $N(0, 1)$, т. е.

$$\lim_{n \rightarrow \infty} P(\{\eta_n < x\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad x \in \mathcal{R}^1, \quad (11.49)$$

причем стремление к пределу в (11.49) равномерно по $x \in \mathcal{R}^1$.

Доказательство. Нам надо доказать, что равномерно по $x \in \mathcal{R}^1$ $\lim_{n \rightarrow \infty} F_n(x) = \Phi_0(x)$, где $F_n(\cdot)$ — функция распределения случайной

величины η_n , $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz$, $x \in \mathcal{R}^1$, — функция распределения нормальной $N(0, 1)$ случайной величины. В силу теоремы 1.11.2 достаточно доказать, что равномерно на каждом конечном интервале $|t| \leq T$, $T > 0$, $f_n(t) \rightarrow e^{-t^2/2}$ при $n \rightarrow \infty$, где $f_n(\cdot)$ — х.ф. случайной величины η_n , $e^{-t^2/2}$, $t \in \mathcal{R}^1$, — х.ф. нормального распределения. Поскольку η_n представляет собой сумму взаимно независимых случайных величин $(\xi_k - \mu)/\sigma\sqrt{n}$, $k = 1, \dots, n$, то в силу леммы 1.11.1 $f_n(t) = (\varphi(t/\sigma\sqrt{n}))^n$, где $\varphi(\cdot)$ — х.ф. случайной величины $\xi_k - \mu$, $k = 1, 2, \dots$. По условию ξ_k имеет конечный момент 2-го порядка, и вследствие леммы 1.11.2 производная $\varphi''(\cdot)$ непрерывна. Записав для $\varphi(\cdot)$ разложение по формуле Тейлора с остаточным членом в форме Пеано, получим

$$\begin{aligned} \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) &= \varphi(0) + \varphi'(0)\frac{t}{\sigma\sqrt{n}} + \varphi''(0)\frac{t^2}{2\sigma^2 n} + o\left(\frac{t^2}{n\sigma^2}\right) = \\ &= 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right), \quad t \in \mathcal{R}^1, \end{aligned} \quad (11.50)$$

так как $\varphi(0) = 1$, $\varphi'(0) = iM(\xi_k - \mu) = 0$, $\varphi''(0) = -M(\xi_k - \mu)^2 = -\sigma^2$. Отсюда при $n \rightarrow \infty$

$$f_n(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right)\right]^n \rightarrow e^{-t^2/2} \quad (11.51)$$

равномерно по t , $|t| \leq T$, где $T > 0$ — любое. ■

Следствие 1.11.4. (Интегральная предельная теорема Муавра–Лапласа). Пусть ξ — число успехов в серии из n взаимно независимых испытаний, p — вероятность успеха при каждом испытании. Тогда при $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} P\left\{\frac{\xi - np}{\sqrt{npq}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad (11.52)$$

причем стремление к пределу равномерно по $x \in \mathcal{R}^1$.

Доказательство. Так как $\xi = \xi_1 + \dots + \xi_n$, где ξ_k — число успехов при k -м испытании, и ξ_k , $k = 1, \dots, n$, независимы и одинаково распределены, $P\{\xi_k = 1\} = p$, $P\{\xi_k = 0\} = q$, $M\xi_k = p$, $D\xi_k = pq$, то

подставив эти значения в (11.48), получим на основании теоремы 1.11.3 равенство (11.52). ■

Пример 1.11.7. Проиллюстрируем применения закона больших чисел и центральной предельной теоремы в задачах эмпирического определения вероятности и вычисления интегралов. Пусть $\xi_1, \dots, \xi_n$ — взаимно независимые копии случайной величины ξ , принимающей значения в $\mathcal{R}^m$, $A \subset \mathcal{R}^m$, $P(\xi \in A) \equiv P(A)$, где $P(A) = \int_A p(x)dx$, если, в частности, ξ — абсолютно непрерывная случайная величина, и $\chi_A(x) = \begin{cases} 1, & \text{если } x \in A, \\ 0, & \text{если } x \in \mathcal{R}^1 \setminus A, \end{cases}$ — индикаторная функция множества A . Определим "эмпирическую вероятность" события $\xi \in A$

$$P_n(A) = \nu_n(A)/n = (1/n) \sum_{j=1}^n \chi_A(\xi_j), \quad (*)$$

где $\nu_n(A)$ — число тех из $\xi_1, \dots, \xi_n$, которые попадают в A , $P_n(A)$ — частота событий $\xi_j \in A$, $j = 1, \dots, n$.

Поскольку $\chi_A(\xi_1), \dots, \chi_A(\xi_n)$ суть взаимно независимые копии случайной величины $\chi_A(\xi)$, $M\chi_A(\xi) = 1 \cdot P(A) + 0 \cdot P(\mathcal{R}^m \setminus A) = P(A) = M\chi_A(\xi)$, и $D(\chi_A(\xi)) = M\chi_A^2(\xi) - (M\chi_A(\xi))^2 = P(A)(1 - P(A))$, то согласно закону больших чисел $P_n(A) \xrightarrow[n \rightarrow \infty]{P} P(A)$, откуда следует, в частности, способ вычисления интегралов, ибо $P_n(A) \xrightarrow[n \rightarrow \infty]{P} \int_A p(x)dx$, если ξ — абсолютно непрерывна (метод Монте-Карло), см. пример 1.10.2.

Центральная предельная теорема позволяет уточнить характер сходимости $P_n(A) \xrightarrow[n \rightarrow \infty]{} P(A)$, поскольку согласно теореме Муавра-Лапласа

$$\frac{\sum_{k=1}^n (\chi_A(\xi_k) - P(A))}{\sqrt{nP(A)(1 - P(A))}} = \sqrt{n} \frac{P_n(A) - P(A)}{\sqrt{P(A)(1 - P(A))}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1),$$

и, следовательно, для любого $\varepsilon > 0$ $P\left(\frac{|P_n(A) - P(A)|}{\sqrt{P(A)(1 - P(A))}} \leq \varepsilon\right) \approx$

$$\frac{1}{\sqrt{2\pi}} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} e^{-u^2/2} du \xrightarrow[n \rightarrow \infty]{} 1, \text{ а так как для любого (из- меримого) } A \subset \mathcal{R}^m \quad P(A)(1 - P(A)) \leq 1/4, \text{ то } P(|P_n(A) - P(A)| \leq \delta) \geq P\left(\frac{|P_n(A) - P(A)|}{\sqrt{P(A)(1 - P(A))}} \leq 2\delta\right) \approx$$

$$\frac{1}{\sqrt{2\pi}} \int_{-2\delta\sqrt{n}}^{2\delta\sqrt{n}} e^{-u^2/2} du \xrightarrow[n \rightarrow \infty]{} 1.$$

Естественно ожидать, что если в условиях теоремы 1.11.3 функция распределения $F(\cdot)$ случайных величин ξ_k , $k = 1, 2, \dots$, имеет плотность $p(\cdot)$, то плотность $p_n(\cdot)$ случайной величины η_n должна сходиться при $n \rightarrow \infty$ к плотности $p_0(\cdot)$ нормального распределения. Вообще говоря, это неверно, но во всех практически интересных случаях высказанное утверждение имеет место, точнее, справедлива

Теорема 1.11.4. Пусть выполнены условия теоремы 1.11.3 и, кроме того, х.ф. $\varphi(\cdot)$ случайных величин ξ_k , $k = 1, 2, \dots$, абсолютно интегрируема на $\mathcal{R}^1$. Тогда плотность $p_n(\cdot)$ случайной величины $\eta_n = \sum_{k=1}^n (\xi_k - \mu)/\sigma\sqrt{n}$ при $n \rightarrow \infty$ сходится к плотности $p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ равномерно по $x \in \mathcal{R}^1$.

Доказательство. Мы видели в теореме 1.11.3, что х.ф. $f_n(\cdot)$ случайной величины η_n равна $f_n(t) = [\varphi(t/\sigma\sqrt{n})]^n$, $t \in \mathcal{R}^1$, где $\varphi(\cdot)$ — х.ф. случайной величины $\xi_k - \mu$, и, следовательно, также абсолютно интегрируема на $\mathcal{R}^1$. Применяя теорему 1.11.1 к $f_n(\cdot)$ и к х.ф. $e^{-t^2/2}$, $t \in \mathcal{R}^1$, нормального распределения, получим

$$p_n(x) - p_0(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} [f_n(t) - e^{-t^2/2}] dt, \quad x \in \mathcal{R}^1, \quad (11.53)$$

где $p_n(\cdot)$ — плотность распределения случайной величины η_n .

Отсюда сразу для всех $x \in \mathcal{R}^1$ получим

$$|p_n(x) - p_0(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi^n(t/\sigma\sqrt{n}) - e^{-t^2/2}| dt. \quad (11.54)$$

Докажем, что правая часть здесь стремится к нулю при $n \rightarrow \infty$. Пусть $\varepsilon > 0$ — любое. При доказательстве теоремы 1.11.3 было показано, что $|\varphi^n(t/\sigma\sqrt{n}) - e^{-t^2/2}| \rightarrow 0$ при $n \rightarrow \infty$ равномерно в любом интервале $|t| \leq T$, $T > 0$. Таким образом, для любого $T > 0$ найдется такой номер $n_0 = n_0(T)$, что при $n \geq n_0$

$$\frac{1}{2\pi} \int_{-T}^T |\varphi^n(t/\sigma\sqrt{n}) - e^{-t^2/2}| dt < \frac{\varepsilon}{3}. \quad (11.55)$$

Поскольку $\varphi(0) = 1$, $\varphi'(0) = 0$ и $\varphi''(0) = -\sigma^2$ (см. теорему 1.11.3), а для функции $\psi(t) = e^{-\sigma^2 t^2/4}$, $t \in \mathcal{R}^1$, $\psi(0) = 1$, $\psi'(0) = 0$ и $\psi''(0) = -\sigma^2/2$, то существует такое $\delta > 0$, что $|\varphi(t)| \leq e^{-\sigma^2 t^2/4}$ при $|t| < \delta$. Таким образом,

$$|\varphi^n(t/\sigma\sqrt{n})| \leq e^{-t^2/4} \text{ при } |t| < \delta\sigma\sqrt{n}. \quad (11.56)$$

Поэтому, для части интеграла (11.54) по области $T \leq |t| \leq \delta\sigma\sqrt{n}$ имеет место оценка

$$\begin{aligned} \frac{1}{2\pi} \left[\int_T^{\delta\sigma\sqrt{n}} \left| \varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - e^{-\frac{t^2}{2}} \right| dt + \int_{-\delta\sigma\sqrt{n}}^{-T} \left| \varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - e^{-\frac{t^2}{2}} \right| dt \right] &\leq \\ &\leq \frac{4}{2\pi} \int_T^{\delta\sigma\sqrt{n}} e^{-\frac{t^2}{4}} dt \leq \frac{2}{\pi} \int_T^{\infty} e^{-\frac{t^2}{4}} dt < \frac{\varepsilon}{3}, \end{aligned} \quad (11.57)$$

если T достаточно велико.

Наконец, рассмотрим область $|t| \geq \delta\sigma\sqrt{n}$. В этой области аргумент функции $\varphi^n(t/\sigma\sqrt{n})$ не меньше $\delta > 0$. Покажем, что существует число β , $0 \leq \beta < 1$, такое, что при $|\tau| \geq \delta$

$$|\varphi(\tau)| \leq \beta. \quad (11.58)$$

Действительно, в силу теоремы 1.11.1 существует непрерывная на $\mathcal{R}^1$ плотность $p(\cdot)$, такая, что

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx, \quad t \in \mathcal{R}^1. \quad (11.59)$$

Если в некоторой точке $t \in \mathcal{R}^1$ $|\varphi(t)| = 1$, т.е. $\varphi(t) = e^{i\alpha t}$, где α — некоторое действительное число, то

$$1 = \int_{-\infty}^{\infty} e^{itx - i\alpha t} p(x) dx = \int_{-\infty}^{\infty} \cos(x - \alpha)t p(x) dx = \int_{-\infty}^{\infty} \cos(yt) p(y + \alpha) dy. \quad (11.60)$$

Поскольку $\int_{-\infty}^{\infty} p(y + \alpha) dy = 1$, то равенство (11.60) можно переписать в виде

$$\int_{-\infty}^{\infty} (1 - \cos yt) p(y + \alpha) dy = 0. \quad (11.61)$$

Поскольку подынтегральная функция в (11.61) неотрицательна, то это равенство возможно лишь при $t = 0$ (если $t \neq 0$, то $p(y + \alpha) = 0$ во всех точках $y \in \mathcal{R}^1$, кроме, быть может, точек вида $y = 2\pi k/t$, $k = 0, \pm 1, \dots$, а тогда в силу непрерывности $p(y + \alpha) \equiv 0$, что невозможно).

Итак, $|q(t)| < 1$ для всех $t \in \mathcal{R}^1$, кроме $t = 0$. Поскольку в силу известной леммы Римана–Лебега

$$\lim_{t \rightarrow \infty} \varphi(t) = \lim_{t \rightarrow \infty} \int_{-\infty}^{\infty} e^{itx} p(x) dx = 0,$$

то отсюда следует утверждение (11.58).

Оценим теперь интеграл (11.54) по оставшейся части $|t| \geq \delta\sigma\sqrt{n}$. В силу (11.58)

$$\begin{aligned} \frac{1}{2\pi} \int_{|t| \geq \delta\sigma\sqrt{n}} \left| \varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - e^{-\frac{t^2}{2}} \right| dt &\leq \frac{\sigma\sqrt{n}}{2\pi} \int_{|\tau| \geq \delta} |\varphi^n(\tau)| d\tau + \\ &+ \frac{1}{2\pi} \int_{|t| \geq \delta\sigma\sqrt{n}} e^{-\frac{t^2}{2}} dt \leq \frac{\sigma\sqrt{n}}{2\pi} \beta^{n-1} \int_{|\tau| \geq \delta} |\varphi(\tau)| d\tau + \\ &+ \frac{1}{2\pi} \int_{|t| \geq \delta\sigma\sqrt{n}} e^{-\frac{t^2}{2}} dt < \frac{\varepsilon}{3}, \end{aligned} \quad (11.62)$$

если n достаточно велико. Сопоставляя оценки (11.55), (11.57), (11.62), завершим доказательство теоремы. ■

Теперь мы откажемся от нежелательного в ряде вопросов предположения о том, что случайные величины ξ_k , $k = 1, 2, \dots$, одинаково распределены, и докажем центральную предельную теорему в форме Ляпунова.

Теорема 1.11.5. (Ляпунов А.М.). Пусть $\{\xi_n\}$ — последовательность взаимно независимых случайных величин с $M\xi_n = \mu_n$, $D\xi_n = \sigma_n^2$ и $M|\xi_n - \mu_n|^3 < \infty$, $n = 1, 2, \dots$. Тогда, если выполнено условие Ляпунова

$$\frac{1}{B_n^3} \sum_{k=1}^n M|\xi_k - \mu_k|^3 \rightarrow 0 \text{ при } n \rightarrow \infty, \quad (11.63)$$

где $B_n = \sqrt{\sum_{k=1}^n D\xi_k} \equiv \sqrt{\sum_{k=1}^n \sigma_k^2}$, то последовательность

$$\eta_n = \frac{\sum_{k=1}^n (\xi_k - \mu_k)}{B_n}, \quad n = 1, 2, \dots, \quad (11.64)$$

сходится по распределению к $N(0, 1)$ равномерно на $\mathcal{R}^1$.

Доказательство. Обозначим

$$\zeta_{nk} = \frac{1}{B_n}(\xi_k - \mu_k), \quad f_{nk}(\cdot) — \text{х. ф. } \zeta_{nk}, \quad k = 1, 2, \dots, n. \quad (11.65)$$

Тогда в (11.64)

$$\begin{aligned} \eta_n &= \sum_{k=1}^n \zeta_{nk}, \quad M\zeta_{nk} = \frac{1}{B_n} M(\xi_k - \mu_k) = 0, \quad D\zeta_{nk} = \frac{1}{B_n^2} D\xi_k, \\ M|\zeta_{nk}|^3 &= \frac{1}{B_n^3} M|\xi_k - \mu_k|^3. \end{aligned} \quad (11.66)$$

Если $f_n(\cdot)$ — х. ф. η_n , то, поскольку

$$f_n(t) = \prod_{k=1}^n f_{nk}(t), \quad t \in \mathcal{R}^1, \quad (11.67)$$

достаточно в силу теоремы (1.11.2) доказать, что при $n \rightarrow \infty$

$$\prod_{k=1}^n f_{nk}(t) \rightarrow e^{-\frac{t^2}{2}} \quad (11.68)$$

равномерно по t , $|t| \leq T$, $T > 0$ — любое фиксированное. Запишем для $f_{nk}(\cdot)$ формулу Тейлора при $|t| \leq T$ с остаточным членом Лагранжа:

$$f_{nk}(t) = 1 - \frac{D\xi_k}{B_n^2} \frac{t^2}{2} + R_{nk}(t), \quad (11.69)$$

ибо $f_{nk}(0) = 1$, $f'_{nk}(0) = iM\zeta_{nk} = 0$, $f''_{nk}(0) = -M\zeta_{nk}^2 = -D\xi_k/B_n^2$, а для остатка $R_{nk}(t)$ при $|t| \leq T$ при $k = 1, \dots, n$

$$\begin{aligned} |R_{nk}(t)| &\leq \frac{T^3}{6} \max_{|t| \leq T} |f'''_{nk}(t)| \leq \frac{T^3}{6} M|\zeta_{nk}|^3 = \\ &= \frac{T^3}{6} \frac{1}{B_n^3} M|\xi_k - \mu_k|^3 \leq \frac{T^3}{6} \frac{1}{B_n^3} \sum_{k=1}^n M|\xi_k - \mu_k|^3 \rightarrow 0, \quad n \rightarrow \infty \end{aligned} \quad (11.70)$$

в силу условия (11.63).

Далее, при любом $\varepsilon > 0$ и $k = 1, \dots, n$

$$\begin{aligned} \frac{D\xi_k}{B_n^2} &= D\zeta_{nk} = \int_{-\infty}^{\infty} x^2 dF_{nk}(x) \leq \\ &\leq \varepsilon^2 \int_{|x| \leq \varepsilon} dF_{nk}(x) + \frac{1}{\varepsilon} \int_{|x| \geq \varepsilon} |x|^3 dF_{nk} \leq \varepsilon^2 + \frac{1}{\varepsilon} M|\zeta_{nk}|^3 \leq 2\varepsilon^2, \end{aligned} \quad (11.71)$$

если n — достаточно велико, так как $M|\zeta_{nk}|^3 = \frac{1}{B_n^3} M|\xi_k - \mu_k|^3 \rightarrow 0$ при $n \rightarrow \infty$. Обозначим

$$z_{nk} = -\frac{D\xi_k}{B_n^2} \frac{t^2}{2} + R_{nk}(t). \quad (11.72)$$

Из оценок (11.70) и (11.71) следует, что

$$|z_{nk}| \rightarrow 0 \text{ при } n \rightarrow \infty \text{ и } k = 1, \dots, n, \quad |t| \leq T. \quad (11.73)$$

Равенство (11.69) перепишем в виде

$$f_{nk}(t) = 1 + z_{nk}$$

и

$$\ln f_{nk}(t) = z_{nk} + (\ln f_{nk}(t) - z_{nk}), \quad (11.74)$$

где взято главное значение логарифма. Из (11.72) и (11.74) находим

$$\begin{aligned} \sum_{k=1}^n \ln f_{nk}(t) &= \sum_{k=1}^n z_{nk} + \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}) = \\ &= -\frac{t^2}{2} \frac{1}{B_n^2} \sum_{k=1}^n D\xi_k + \sum_{k=1}^n R_{nk}(t) + \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}) = \\ &= -\frac{t^2}{2} + \sum_{k=1}^n R_{nk}(t) + \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}). \end{aligned} \quad (11.75)$$

В силу предпоследнего неравенства в (11.70) и условия (11.63) при $|t| \leq T$

$$\left| \sum_{k=1}^n R_{nk}(t) \right| \leq \frac{T^3}{6} \frac{1}{B_n^3} \sum_{k=1}^n M|\xi_k - \mu_k|^3 \rightarrow 0, \quad n \rightarrow \infty. \quad (11.76)$$

Для любого комплексного β с $|\beta| \leq 1/2$ выполнено неравенство

$$|\ln(1 + \beta) - \beta| = \left| \int_0^\beta \frac{zdz}{1+z} \right| \leq 2 \left| \int_0^\beta |z| |dz| \right| \leq |\beta|^2, \quad (11.77)$$

поэтому при $n \rightarrow \infty$

$$\begin{aligned} \left| \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}) \right| &\leq \sum_{k=1}^n |z_{nk}|^2 \leq \max_{k=1, \dots, n} |z_{nk}| \sum_{k=1}^n |z_{nk}| \leq \\ &\leq \max_k |z_{nk}| \sum_{k=1}^n \left(\frac{t^2}{2} \frac{D\xi_k}{B_n^2} + |R_{nk}(t)| \right) \leq \\ &\leq \max_k |z_{nk}| \left(\frac{T^2}{2} + \sum_{k=1}^n |R_{nk}(t)| \right) \rightarrow 0, \end{aligned} \quad (11.78)$$

в силу (11.73) и (11.76).

Сопоставляя (11.75), (11.76) и (11.78), заключаем, что $\sum_{k=1}^n \ln f_{nk}(t) \rightarrow -\frac{t^2}{2}$ при $n \rightarrow \infty$ равномерно по t при $|t| \leq T$, что равносильно (11.68). ■

1.11.4. Применения центральных предельных теорем. Приме-

ры. Отметим прежде всего случай, когда нельзя применять центральные предельные теоремы. Пусть мы хотим оценить вероятность $P\{\eta_n < x\}$, причем речь идет о тех $x \in \mathcal{R}^1$, при которых эта вероятность близка к нулю или к единице. Если заменить $P\{\eta_n < x\}$ на $\Phi_0(x)$, то ошибка может оказаться очень большой (порядка сотен процентов) ввиду того, что хотя разность $P(\{\eta_n < x\}) - \Phi_0(x)$ и будет равномерно малой при всех $x \in \mathcal{R}^1$, но неверно, что отношение

$P(\{\eta_n < x\})/\Phi_0(x) \rightarrow 1$ равномерно по $x \in \mathcal{R}^1$, т.е. «хвосты» распределения $P(\{\eta_n < x\})$, $x \in \mathcal{R}^1$, требуют очень аккуратной оценки.

В частности, если случайные величины $\xi_1, \xi_2, \dots$ взаимно независимы и одинаково распределены как случайная величина ξ , у которой существуют моменты $M\xi = m$, $D\xi = \sigma^2$ и $M(|\xi - m|^3) = \alpha^3$, то скорость сходимости в теореме 1.11.5 оценивается следующим неравенством (Берри–Эссеен)

$$\sup_{-\infty < x < \infty} |P(\{\sum_{k=1}^n (\xi_k - m)/(\sigma\sqrt{n}) < x\}) - \Phi(x)| \leq c\alpha^3/(\sigma^3\sqrt{n}),$$

где $1/\sqrt{2\pi} \leq c \leq 0.8$.

Пример 1.11.8. Пользуясь теоремой 1.11.3, оценим вероятность уклонения частоты успехов от вероятности успеха p в серии n испытаний Бернулли. Оценка основана на соотношении

$$\begin{aligned} P\left(\left\{\left|\frac{\eta_n}{n} - p\right| > \varepsilon\right\}\right) &= P\left(\left\{\left|\frac{\eta_n - np}{\sqrt{npq}}\right| > \varepsilon\sqrt{\frac{n}{pq}}\right\}\right) \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\varepsilon\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_{\varepsilon\sqrt{\frac{n}{pq}}}^{\infty} e^{-\frac{x^2}{2}} dx = 2\Phi_0\left(-\varepsilon\sqrt{\frac{n}{pq}}\right), \quad \eta_n - \end{aligned}$$

число успехов при n испытаниях. Поскольку при $p + q = 1$, очевидно, $pq \leq \frac{1}{4}$, то эта вероятность не превосходит $2\Phi_0(-2\varepsilon\sqrt{n})$. Поэтому

$$P\left(\left\{\frac{\eta_n}{n} - \varepsilon < p < \frac{\eta_n}{n} + \varepsilon\right\}\right) \approx 1 - 2\Phi_0\left(-\varepsilon\sqrt{\frac{n}{pq}}\right) \geq 1 - 2\Phi_0(-2\varepsilon\sqrt{n}).$$

Таким образом, зная число успехов η_n в n испытаниях Бернулли, мы можем построить случайный интервал $\left(\frac{\eta_n}{n} - \varepsilon, \frac{\eta_n}{n} + \varepsilon\right)$, который накрывает истинное (неизвестное) значение вероятности p с любой заданной вероятностью $1 - \alpha$. Для этого следует лишь выбрать $\varepsilon = \varepsilon(\alpha)$ из соотношения $2\Phi_0(-2\varepsilon\sqrt{n}) = \alpha$, пользуясь таблицами нормального распределения. Тогда $P\left\{\frac{\eta_n}{n} - \varepsilon(\alpha) < p < \frac{\eta_n}{n} + \varepsilon(\alpha)\right\} = 1 - \alpha$. Интервал $\left(\frac{\eta_n}{n} - \varepsilon(\alpha), \frac{\eta_n}{n} + \varepsilon(\alpha)\right)$ называется *доверительным интервалом* для p с *уровнем доверия* $1 - \alpha$ (см. подробнее об этом §2.2).

Пример 1.11.9. Пусть $\xi_1, \xi_2, \dots$ — попарно независимые случайные величины, $M\xi_k = \mu$, $D\xi_k = \sigma^2$ для всех $k = 1, 2, \dots$. Закон больших чисел (см. §1.10) утверждает, что для любого $\varepsilon > 0$ при $n \rightarrow \infty$

$$P\left(\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n} - \mu\right| > \varepsilon\right\}\right) \rightarrow 0.$$

Если $\xi_1, \xi_2, \dots$ не только попарно независимы (что достаточно для применимости закона больших чисел), но и независимы в совокупности, то можно применить теорему 1.11.3. Для любого $\varepsilon > 0$ это дает

$$P\left(\left|\frac{\xi_1 + \dots + \xi_n}{n} - \mu\right| > \varepsilon\right) = P\left(\left|\frac{\xi_1 + \dots + \xi_n - n\mu}{\sigma\sqrt{n}}\right| > \frac{\varepsilon\sqrt{n}}{\sigma}\right) = \\ = P(\{\eta_n < -\frac{\varepsilon\sqrt{n}}{\sigma}\}) + P(\{\eta_n > \frac{\varepsilon\sqrt{n}}{\sigma}\}) = 2\Phi_0\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right), \quad \text{где} \\ \eta_n = (\xi_1 + \dots + \xi_n - n\mu)/(\sigma\sqrt{n}).$$

Из таблиц нормального распределения следует, что при $\frac{\varepsilon\sqrt{n}}{\sigma} = 3$, т. е. при $\varepsilon = \frac{3\sigma}{\sqrt{n}}$ вероятность $P\left(\left|\eta_n\right| < \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - 2\Phi_0\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)$ равна 0,997. Это так называемое *правило 3σ*.

Рассмотрим еще несколько примеров применения центральных предельных теорем.

Пример 1.11.10. Пусть ξ_k , $k = 1, 2, \dots$, — взаимно независимые дискретные случайные величины, $P(\{\xi_k = \pm k\}) = 1/2$, $k = 1, 2, \dots$, $M\xi_k = k \cdot \frac{1}{2} - k \cdot \frac{1}{2} = 0$, $D\xi_k = M\xi_k^2 = k^2 \cdot \frac{1}{2} + k^2 \cdot \frac{1}{2} = k^2$, $M|\xi_k - M\xi_k|^3 = M|\xi_k|^3 = k^3$, $k = 1, 2, \dots$. Легко по индукции доказываются формулы $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$, $\sum_{k=1}^n k^3 = \left[\frac{n(n+1)}{2}\right]^2$.

Отсюда $B_n = \sqrt{\sum_{k=1}^n D\xi_k} = \sqrt{\frac{n(n+1)(2n+1)}{6}}$, и условие (63) Ляпунова выполнено

$$\frac{1}{B_n^3} \sum_{k=1}^n M|\xi_k - M\xi_k|^3 = \left(\frac{n(n+1)(2n+1)(2n+1)}{6}\right)^{-3/2} \left(\frac{n(n+1)}{2}\right)^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Таким образом согласно теореме 1.11.5 случайные величины $\eta_n = (\xi_1 + \dots + \xi_n)/B_n$ асимптотически при $n \rightarrow \infty$ нормальны $N(0, 1)$.

Пример 1.11.11. Докажем, что

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}. \quad (11.79)$$

Доказательство. Рассмотрим последовательность независимых случайных величин ξ_k , $k = 1, 2, \dots$, распределенных по закону Пуассона с параметром $\lambda = 1$. Так как $M\xi_k = 1$, $D\xi_k = 1$, $k = 1, 2, \dots$, то согласно теореме 1.11.3 для $\eta_n = \frac{\xi_1 + \dots + \xi_n - n}{\sqrt{n}}$, $n = 1, 2, \dots$, при $n \rightarrow \infty$ $P(\{\eta_k < x\}) = P\left(\left\{\frac{\xi_1 + \dots + \xi_n - n}{\sqrt{n}} < x\right\}\right) \rightarrow \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$ равномерно по $x \in \mathcal{R}^1$. Положив здесь $x = 0$, получим

$$P(\{\xi_1 + \dots + \xi_n < n\}) = P\left(\left\{\frac{\xi_1 + \dots + \xi_n - n}{\sqrt{n}} < 0\right\}\right) \rightarrow \frac{1}{2}. \quad (11.80)$$

Согласно следствию 1.11.2 сумма $\xi_1 + \dots + \xi_n$ распределена по закону Пуассона с параметром $\lambda = n$, поэтому

$$P(\{\xi_1 + \dots + \xi_n < n\}) = \sum_{k=0}^n \frac{n^k}{k!} e^{-n} - \frac{n^n e^{-n}}{n!}. \quad (11.81)$$

Сопоставляя (11.80) и (11.81), видим, что справедливо (11.79).

Пример 1.11.12. Театр, вмещающий 1000 человек, имеет два разных входа. Около каждого входа имеется гардероб. Сколько мест должно быть в каждом гардеробе для того, чтобы в 99% случаев все зрители могли раздеться в гардеробе того входа, через который они вошли? Предполагается, что зрители приходят парами и каждая пара независимо от других выбирает с вероятностью $1/2$ любой из входов.

Очевидно, это схема Бернулли с числом испытаний $n = 500$ (число пар), p — вероятность выбора определенного входа, $p = 1/2$, $q = 1 - p = 1/2$.

Обозначим x число мест в каждом гардеробе, m — число успехов, т. е. случаев, когда пришедшей паре есть место в гардеробе соответствующего входа. Тогда x определяется из условия $P\left(\left\{m \leq \frac{x}{2}, n - m \leq \frac{x}{2}\right\}\right) = 0,99$. Для нахождения x воспользуемся интегральной теоремой Муавра–Лапласа:

$$0,99 = P\left(\left\{m \leq \frac{x}{2}, n - m \leq \frac{x}{2}\right\}\right) \approx -2\Phi_0\left(\frac{\frac{x}{2} - \frac{500}{2}}{5\sqrt{5}}\right) + 1.$$

Отсюда по таблицам определяем: $x = 537$.

1.12. Конечные однородные цепи Маркова

1.12.1. Определение цепи Маркова. Понятие конечной цепи Маркова является простейшим и вместе с тем важным обобщением схемы независимых испытаний на случай, когда испытания взаимно зависимы. Напомним, что последовательностью n взаимно независимых испытаний мы называли вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n)$; элементами $\Omega_n = \Omega \times \dots \times \Omega$ являются всевозможные последовательности $(\omega_{i_1}, \dots, \omega_{i_n})$, где $\omega_k \in \Omega$, $k = 1, 2, \dots$, $\Omega = \{\omega_1, \omega_2, \dots\}$ — множество исходов в каждом испытании, $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ — σ -алгебра всех подмножеств Ω_n , при этом независимость испытаний фиксировалась определением вероятности P_n , $P_n(\{(\omega_{i_1}, \dots, \omega_{i_n})\}) = p_{i_1} \dots p_{i_n}$, где $p_k = P(\{\omega_k\})$ — вероятность исхода ω_k в каждом испытании¹⁾, $k = 1, 2, \dots$

¹⁾ ω_i , $(\omega_{i_1}, \omega_{i_2})$, ..., $(\omega_{i_1}, \dots, \omega_{i_n})$ — исходы первого, первого и второго, ..., первого, второго, ..., n -го испытаний, $\{\omega_1\} \subset \Omega$, $\{(\omega_{i_1}, \omega_{i_2})\} \subset \Omega \times \Omega = \Omega_2$, ..., $\{(\omega_{i_1}, \dots, \omega_{i_n})\} \subset \Omega \times \dots \times \Omega = \Omega_n$ — соответствующие элементарные события.

Предположим теперь, что в последовательности n испытаний вероятность исхода в любом s -м испытании зависит от исхода $(s-1)$ -го испытания и не зависит от исходов испытаний с номерами $s-2, s-3, \dots, 1$. Обозначим эту «условную вероятность»

$$p_{ik} = P(\{\omega_k\}|\{\omega_i\}), \quad 0 \leq p_{ik} \leq 1, \quad i, k = 1, 2, \dots \quad (12.1)$$

В этом обозначении отражен тот факт, что вероятность события $\{\omega_k\}$ в s -м испытании при условии, что в $(s-1)$ -м испытании произошло событие $\{\omega_i\}$, не зависит от номера s . Это свойство называется *однородностью последовательности испытаний*.

Пусть задано распределение вероятностей для первого испытания:

$$P(\{\omega_j\}) = a_j, \quad j = 1, 2, \dots, \quad a_j \geq 0, \quad \sum_{j=1}^{\infty} a_j = 1. \quad (12.2)$$

Тогда вероятность исхода (ω_i, ω_j) первого и второго испытаний равна $P(\{\omega_i, \omega_j\}) = a_i p_{ij}$ и аналогично для исхода k испытаний, $k = 2, 3, \dots$,

$$P(\{\omega_{i_1}, \dots, \omega_{i_k}\}) = a_{i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}, \quad i_1, i_2, \dots, i_k = 1, 2, \dots \quad (12.3)$$

Убедимся, что равенства (12.3) при $k = n$ задают вероятность на $\mathcal{F}_n$, т. е. на всех подмножествах Ω_n . Для этого достаточно проверить, что

$$\sum_{(\omega_{i_1}, \dots, \omega_{i_n}) \in \Omega_n} P(\{\omega_{i_1}, \dots, \omega_{i_n}\}) = \sum_{i_1=1, \dots, i_n=1}^{\infty} a_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} = 1, \quad (12.4)$$

так как в остальном проверка дословно повторяет случай независимых испытаний (см. §5). Поскольку для всякого исхода ω_i в $(s-1)$ -м испытании в следующем (s) -м испытании непременно наблюдается один из исходов $\omega_1, \omega_2, \dots$, то для каждого $i = 1, 2, \dots$ $\sum_{j=1}^{\infty} p_{ij} = 1$. Это влечет следующую цепочку равенств:

$$\begin{aligned} & \sum_{i_1=1, \dots, i_n=1}^{\infty} a_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} = \\ &= \sum_{i_1=1, \dots, i_{n-1}=1}^{\infty} a_{i_1} p_{i_1 i_2} \dots p_{i_{n-2} i_{n-1}} \sum_{i_n=1}^{\infty} p_{i_{n-1} i_n} = \\ &= \sum_{i_1=1, \dots, i_{n-1}=1}^{\infty} a_{i_1} p_{i_1 i_2} \dots p_{i_{n-2} i_{n-1}} = \dots = 1, \end{aligned}$$

тем самым равенство (12.4) доказано.

Замечание 1.12.1. На самом деле приведенная схема построения марковской цепи не вполне корректна. Дело в том, что некорректно задавать условные вероятности в (12.1), не определив вначале вероятности $P_2(\{\omega_k, \omega_j\})$, $k, j = 1, 2, \dots$. Чтобы исправить схему построения марковской цепи введем понятие *переходной вероятности*

для пары (измеримых) пространств $(\Omega^{(1)}, \mathcal{P}(\Omega^{(1)}))$, $(\Omega^{(2)}, \mathcal{P}(\Omega^{(2)}))$, в которых $\Omega^{(i)} = \{\omega_1^{(i)}, \omega_2^{(i)}, \dots\}$, $\mathcal{P}(\Omega^{(i)})$ — σ -алгебра всех подмножеств $\Omega^{(i)}$, $i = 1, 2$.

Определение 1.12.1. Переходной вероятностью для $(\Omega^{(1)}, \mathcal{P}(\Omega^{(1)}))$, $(\Omega^{(2)}, \mathcal{P}(\Omega^{(2)}))$ называется отображение $P_{(2)}^{(1)}(\cdot | \times) : \Omega^{(1)} \times \mathcal{P}(\Omega^{(2)}) \rightarrow [0, 1]$, удовлетворяющее следующим условиям:

1) для любого $\omega_i^{(1)} \in \Omega^{(1)}$ отображение $P_{(2)}^{(1)}(\omega_i^{(1)} | \cdot) : \mathcal{P}(\Omega^{(2)}) \rightarrow [0, 1]$ является вероятностью на $(\Omega^{(2)}, \mathcal{P}(\Omega^{(2)}))$ и

2) для любого $A^{(2)} \in \mathcal{P}(\Omega^{(2)})$ отображение $P_{(2)}^{(1)}(\cdot | A^{(2)}) : \Omega^{(1)} \rightarrow [0, 1]$ является случайной величиной на $(\Omega^{(1)}, \mathcal{P}(\Omega^{(1)}), P_{(1)})$, где $P_{(1)}(\times) : \mathcal{P}(\Omega^{(1)}) \rightarrow [0, 1]$ — произвольная вероятность.

В нашем случае дискретного $\Omega^{(1)}$ второе условие выполняется автоматически.

Следующее *свойство переходной вероятности* определяет ее роль в конструкции цепи Маркова. Пусть $P_{(1)}(\cdot) : \mathcal{P}(\Omega^{(1)}) \rightarrow [0, 1]$ — произвольная вероятность на $(\Omega^{(1)}, \mathcal{P}(\Omega^{(1)}))$. Тогда существует единственная вероятность P на $(\Omega^{(1)} \times \Omega^{(2)}, \mathcal{P}(\Omega^{(1)} \times \Omega^{(2)}))$ такая, что

$$P(\{(\omega_i^{(1)}, \omega_j^{(2)})\}) = P_{(1)}(\{\omega_i^{(1)}\})P_{(2)}^{(1)}(\omega_i^{(1)} | \{\omega_j^{(2)}\}), \quad i, j = 1, 2, \dots$$

В самом деле, в силу теоремы о суммировании по блокам

$$\begin{aligned} & \sum_{i, j: (\omega_i^{(1)}, \omega_j^{(2)}) \in \Omega^{(1)} \times \Omega^{(2)}} P_{(1)}(\{\omega_i^{(1)}\})P_{(2)}^{(1)}(\omega_i^{(1)} | \{\omega_j^{(2)}\}) = \\ &= \sum_{i: \omega_i^{(1)} \in \Omega^{(1)}} P_{(1)}(\{\omega_i^{(1)}\}) \sum_{j: \omega_j^{(2)} \in \Omega^{(2)}} P_{(2)}^{(1)}(\omega_i^{(1)} | \{\omega_j^{(2)}\}) = \\ &= \sum_{i: \omega_i^{(1)} \in \Omega^{(1)}} P_{(1)}(\{\omega_i^{(1)}\}) = 1, \end{aligned}$$

ибо по свойству 1) переходной вероятности $\sum_{j: \omega_j^{(2)} \in \Omega^{(2)}} P_{(2)}^{(1)}(\omega_i^{(1)} | \{\omega_j^{(2)}\}) =$

$= 1$ при любом $\omega_i^{(1)} \in \Omega^{(1)}$. ■

Понятно, что схема построения марковской цепи станет вполне корректной, если в (12.1) определить p_{ik} как *переходную вероятность* равенством

$$p_{ik} = P(\omega_i | \{\omega_k\}), \quad i, k = 1, 2, \dots \quad (12.5)$$

Далее для простоты предполагается, что пространство Ω конечно, $\Omega = \{\omega_1, \dots, \omega_N\}$.

Определение 1.12.2. Конечной однородной цепью Маркова, состоящей из n испытаний, называется вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n)$, в котором Ω_n — множество всех последовательностей $(\omega_{i_1}, \dots, \omega_{i_n})$, $1 \leq i_1, i_2, \dots, i_n \leq N$, $\mathcal{F}_n$ — алгебра всех подмножеств Ω_n , и вероятность P_n определена для каждого элементарного события

Ω_n равенством (12.3) с $k = n$, где $a_i \geq 0$, $\sum_{i=1}^N a_i = 1$ — начальное распределение, $p_{ij} \geq 0$, $\sum_{j=1}^N p_{ij} = 1$ — переходные вероятности, $i, j = 1, \dots, N$.

Название «*переходные вероятности*» естественно связано со следующей интерпретацией цепей Маркова. Будем считать, что ω_j , $j = 1, 2, \dots, N$, отмечают состояния некоторой системы, которая может находиться в одном из N состояний. Пусть в дискретные моменты времени $1, 2, \dots$ система может менять свое состояние, переходя из состояния $\omega_{i_{s-1}}$, в котором она находилась в $(s-1)$ -й момент времени, в состояние ω_{i_s} в s -й момент времени с *вероятностью перехода* $p_{i_{s-1}i_s}$. Если распределение вероятностей состояний в начальный момент времени задается равенствами $P(\{\omega_i\}) = a_i$, $i = 1, 2, \dots, N$, то получаем интерпретацию цепи Маркова в терминах состояний некоторой системы. Далее мы пользуемся как той, так и другой точкой зрения на цепи Маркова.

Совокупность вероятностей перехода p_{ij} , $i, j = 1, \dots, N$ образует матрицу

$$\pi_1 = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix}, \quad (12.6)$$

называемую матрицей перехода. По определению все элементы матрицы π_1 неотрицательны и сумма элементов каждой строки равна единице, $\sum_{j=1}^N p_{ij} = 1$, $i = 1, \dots, N$. Матрицы, обладающие этими свойствами, называются *стохастическими*.

Пример 1.12.1. Заметим, что в цепи Маркова «*прошлое*» — $\omega_{i_{s-1}}$ и «*будущее*» — $\omega_{i_{s+1}}$ независимы, если известно «*настоящее*» — ω_{i_s} . Действительно, условная вероятность $P(\{\omega_{i_{s-1}}, \omega_{i_{s+1}}\} | \{\omega_{i_s}\}) = \frac{P(\{\omega_{i_{s-1}}, \omega_{i_s}, \omega_{i_{s+1}}\})}{P(\{\omega_{i_{s-1}}, \omega_{i_s}\})} = \frac{P(\{\omega_{i_{s+1}}\} | \{\omega_{i_{s-1}}, \omega_{i_s}\}) P(\{\omega_{i_{s-1}}, \omega_{i_s}\})}{P(\{\omega_{i_{s-1}}, \omega_{i_s}\})} = \frac{P(\{\omega_{i_s}\})}{P(\{\omega_{i_{s-1}}\})} \cdot \frac{P(\{\omega_{i_{s+1}}\} | \{\omega_{i_s}\})}{P(\{\omega_{i_{s-1}}\} | \{\omega_{i_s}\})}$, ибо для цепи Маркова $P(\{\omega_{i_{s+1}}\} | \{\omega_{i_{s-1}}, \omega_{i_s}\}) = P(\{\omega_{i_{s+1}}\} | \{\omega_{i_s}\})$.

Пример 1.12.2. (*Случайное блуждание с поглощением*). Пусть частица может передвигаться по прямой под действием случайных толчков. В точках $x = 1$ и $x = N$ стоят поглощающие экраны, и при каждом случайном толчке частица передвигается на единицу длины вправо с вероятностью p , влево с вероятностью q , $p + q = 1$, однако, попав в точки $x = 1$ и $x = N$, частица остается в этих точках. Таким образом, состояния частицы суть: $\omega_i = \{\text{частица находится в точке } x = i\}$, $i = 1, \dots, N$. Переходные вероятности в этом случае равны: $p_{11} = 1$, $p_{NN} = 1$, $p_{ii+1} = p$, $p_{ii-1} = q$, $p_{ij} = 0$ для остальных комбинаций i и j .

Матрица перехода имеет вид

$$\pi_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & q & 0 & p \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \quad (*)$$

Полезным представлением цепи Маркова является ее граф (рис. 1.19).

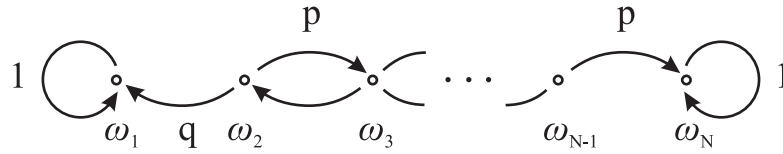


Рис. 1.19. Граф марковской цепи, соответствующей матрице (*).

Пример 1.12.3. (*Случайное блуждание с отражением*). Рассмотрим ту же схему, что и в предыдущем примере, с той лишь разницей, что теперь в точках $x = 1$ и $x = N$ стоят отражающие экраны. В этом случае матрица перехода имеет вид

$$\pi_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ q & 0 & p & \dots & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \quad (**)$$

а цепь Маркова представляется графом, изображенным на рис. 1.20.

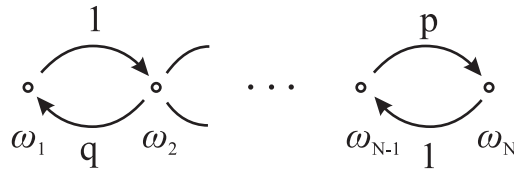


Рис. 1.20. Граф марковской цепи, соответствующей матрице (**).

Пример 1.12.4. (*Схема независимых испытаний*). В этом случае $p_{ij} = P(w_i | \{w_j\}) = a_j$, т. е. вероятность состояния ω_j не зависит от исхода предыдущего испытания и совпадает с вероятностью события

$\{\omega_j\}$ в начальном испытании $i, j = 1, \dots, N$. Матрица π_1 имеет вид

$$\pi_1 = \begin{pmatrix} a_1 & a_2 & \dots & a_N \\ a_1 & a_2 & \dots & a_N \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_N \end{pmatrix} \quad (***)$$

Соответствующий граф представлен на рис. 1.21.

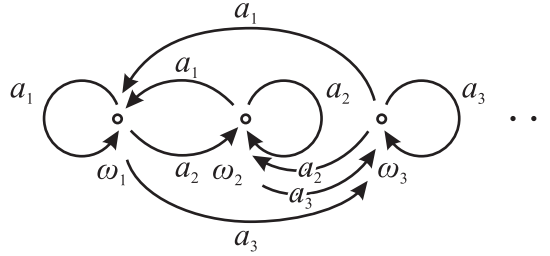


Рис. 1.21. Граф марковской цепи, соответствующей матрице $(***)$.

1.12.2. Вероятность перехода за n шагов. Рассмотрим вероятность перехода из состояния ω_i , которое реализовано, например, в s -м испытании, в состояние ω_j за n шагов, т. е. в состояние ω_j в $s + n$ -м испытании. Ясно, что в силу однородности цепи Маркова эта вероятность зависит только от n (и не зависит от s). Обозначим ее p_{ij}^n . Тогда p_{ik}^m — вероятность перехода за $m < n$ шагов из состояния ω_i в состояние ω_k и p_{kj}^{n-m} — вероятность перехода за $n - m$ шагов из состояния ω_k в ω_j .

Пользуясь формулой полной вероятности и учитывая, что промежуточные состояния $\omega_1, \omega_2, \dots, \omega_N$ в $(s + m)$ -м испытании образуют полную группу попарно несовместных событий, найдем

$$p_{ij}^n = \sum_{k=1}^N p_{ik}^m p_{kj}^{n-m}, \quad i, j = 1, \dots, N, \quad (12.7)$$

здесь m — любое целое от 1 до $n - 1$.

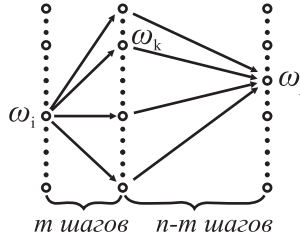


Рис. 1.22. Графическая иллюстрация суммирования в (12.8).

Обозначим π_n матрицу, составленную из вероятностей p_{ij}^n , $i, j = 1, \dots, N$; таким образом, π_n — матрица перехода через n испытаний. Пользуясь формулой для перемножения квадратных матриц, равенство (12.7) можно записать в матричном виде

$$\pi_n = \pi_m \pi_{n-m}, \quad m = 1, \dots, n-1. \quad (12.8)$$

Применяя последовательно формулу (12.8), получим

$$\pi_n = \pi_1 \pi_{n-1} = \pi_1 \pi_1 \pi_{n-2} = \dots = \pi_1^n. \quad (12.9)$$

1.12.3. Эргодичность. Матрица π_1 , соответствующая схеме независимых испытаний (см. пример 1.12.4), обладает свойством $\pi_1 = \pi_n = \pi_1^n$ для любого натурального n . Интуитивно можно ожидать, что и в случае любой цепи Маркова при n переходах влияние начального распределения вероятностей с ростом n должно ослабевать в том смысле, что $p_{ij}^n \rightarrow p_j$ при $n \rightarrow \infty$ независимо от $i = 1, \dots, N$, $j = 1, \dots, N$. Предельная матрица переходов имеет вид матрицы $\pi_1(* **)$ в схеме независимых испытаний и, следовательно, удовлетворяет условию $\pi_1^k = \pi_1$, $k = 1, 2, \dots$. Это свойство цепей Маркова, называемое *эргодичностью*, действительно имеет место, хотя и не во всех случаях. Далее мы укажем достаточное условие эргодичности (теорему Маркова), а сейчас рассмотрим некоторые свойства эргодических цепей Маркова. Обозначим p_j^k вероятность того, что в k -м испытании осуществится событие $\{\omega_j\}$, p_j^k называется *абсолютной вероятностью*, $j = 1, \dots, N$, $k = 1, 2, \dots$.

Пусть числа $a_1, a_2, \dots, a_N$ задают начальное распределение вероятностей состояний $\omega_1, \omega_2, \dots, \omega_N$, $a_i \geq 0$, $i = 1, \dots, N$, $\sum_{i=1}^N a_i = 1$. Согласно формуле полной вероятности распределение вероятностей состояний $\omega_1, \dots, \omega_N$ во втором испытании определяется равенством $p_j^2 = \sum_{k=1}^N a_k p_{kj}$, в третьем испытании $p_j^3 = \sum_{l=1}^N p_l^2 p_{lj} = \sum_{k=1}^N a_k p_{kj}^2$, $j = 1, \dots, N$, и т. д., в n -м испытании задается равенством

$$p_j^n = \sum_{l=1}^N p_l^{n-1} p_{lj} = \sum_{k=1}^N a_k p_{kj}^{n-1}, \quad n = 2, 3, \dots, \quad j = 1, \dots, N. \quad (12.10)$$

Если цепь Маркова эргодическая, то есть если при $n \rightarrow \infty$ существуют пределы

$$\lim_{n \rightarrow \infty} p_{kj}^n = p_j, \quad k, j = 1, \dots, N, \quad (12.11)$$

то существуют в силу (12.10) и пределы для абсолютных вероятностей:

$$\lim_{n \rightarrow \infty} p_j^n = \sum_{k=1}^N a_k \lim_{n \rightarrow \infty} p_{kj}^{n-1} = \sum_{k=1}^N a_k p_j = p_j, \quad j = 1, \dots, N. \quad (12.12)$$

Таким образом, согласно (12.12), если имеет место эргодичность, то существует предельное, называемое *финальным*, распределение вероятностей $p_1, \dots, p_N$, состояний $\omega_1, \dots, \omega_N$, не зависящее от начального распределения $a_1, \dots, a_N$.

Переходя к пределу при $n \rightarrow \infty$ в первом равенстве формулы (12.10), найдем

$$p_j = \lim_{n \rightarrow \infty} p_j^n = \sum_{l=1}^N \lim_{n \rightarrow \infty} p_l^{n-1} p_{lj} = \sum_{l=1}^N p_l p_{lj}, \quad j = 1, \dots, N. \quad (12.13)$$

Таким образом, финальное распределение вероятностей $p_1, \dots, p_N$ удовлетворяет системе линейных однородных уравнений

$$p_j = \sum_{l=1}^N p_l p_{lj}, \quad j = 1, \dots, N. \quad (12.14)$$

и очевидным дополнительным условиям

$$p_j \geq 0, \quad j = 1, \dots, N, \quad \sum_{j=1}^N p_j = 1, \quad (12.15)$$

которые также следуют из (12.10). Система (12.14) в силу стохастичности матрицы перехода π_1 имеет определитель, равный нулю.

Сопоставляя (12.14) и первое равенство в формуле (12.10), приходим к выводу, что если задача (12.14), (12.15) разрешима, а в качестве начального распределения $a_1, \dots, a_N$ выбрать финальное распределение $p_1, \dots, p_N$, то распределение состояний $\omega_1, \dots, \omega_N$ не будет изменяться от испытания к испытанию. Тем самым показано, что *финальное распределение стационарно*.

Отметим также следующий глубокий результат, характеризующий свойство эргодичности. Пусть в каждом состоянии до момента перехода в следующее состояние система находится τ секунд. Пусть, далее, A — некоторое множество состояний, T_A — время, которое система находится в состояниях из множества A , а $T = n\tau$ — общее время функционирования системы. Тогда в случае эргодичности

$$\lim_{n\tau=T \rightarrow \infty} \left(\frac{T_A}{T} \right) = \sum_{i: \omega_i \in A} p_i.$$

Однако доказательство этого результата потребовало бы более глубокого математического аппарата исследования цепей Маркова.

В следующей теореме сформулировано достаточное условие эргодичности цепи Маркова.

Теорема 1.12.1. (Марков А.А.). Пусть существует целое число $\nu \geq 1$, такое, что в матрице $\pi_\nu = \{p_{ij}^\nu\}$ найдется столбец,

все элементы которого больше или равны $\delta > 0$, т.е. найдется $j \in \{1, \dots, N\}$, при котором

$$\min_{1 \leq i \leq N} p_{ij}^\nu = \delta > 0. \quad (12.16)$$

Тогда существуют

$$\lim_{n \rightarrow \infty} p_{ij}^n = p_j, \quad i, j = 1, \dots, N. \quad (12.17)$$

Доказательство. Заметим прежде всего, что матрица $\pi_n = \{p_{ij}^n\}$ — стохастическая при любом $n = 1, 2, \dots$. Действительно, в силу (12.7):

$$\begin{aligned} \sum_{j=1}^N p_{ij}^n &= \sum_{j=1}^N \left(\sum_{k=1}^N p_{ik}^{n-1} p_{kj} \right) = \sum_{k=1}^N p_{ik}^{n-1} \sum_{j=1}^N p_{kj} = \\ &= \sum_{k=1}^N p_{ik}^{n-1} = \dots = \sum_{k=1}^N p_{ik} = 1. \end{aligned} \quad (12.18)$$

Обозначим $m_j^r = \min_{1 \leq i \leq N} p_{ij}^r$, $M_j^r = \max_{1 \leq i \leq N} p_{ij}^r$, $m_j^r \leq M_j^r$, $j = 1, \dots, N$, $r = 1, 2, \dots$. Для всех $i, j = 1, \dots, N$

$$m_j^1 \leq m_j^2 \leq \dots \leq m_j^r \leq p_{ij}^r \leq M_j^r \leq M_j^{r-1} \leq \dots \leq M_j^2 \leq M_j^1,$$

так как

$$m_j^{r+1} = \min_i \sum_{k=1}^N p_{ik} p_{kj}^r \geq \min_i \sum_{k=1}^N p_{ik} m_j^r = m_j^r \min_i \sum_{k=1}^N p_{ik} = m_j^r,$$

и

$$M_j^{r+1} = \max_i \sum_{k=1}^N p_{ik} p_{kj}^r \leq \max_i \sum_{k=1}^N p_{ik} M_j^r = M_j^r \max_i \sum_{k=1}^N p_{ik} = M_j^r.$$

Поскольку последовательности $\{m_j^n\}$ и $\{M_j^n\}$ монотонные и ограниченные, а значит, сходятся при $n \rightarrow \infty$, то $\lim_{n \rightarrow \infty} m_j^n = m_j \leq \varliminf_{n \rightarrow \infty} p_{ij}^n \leq \varlimsup_{n \rightarrow \infty} p_{ij}^n \leq M_j = \lim_{n \rightarrow \infty} M_j^n$.

Покажем, что при условии теоремы $m_j = M_j$, $j = 1, \dots, n$, и как следствие, $m_j = M_j = \varliminf_{n \rightarrow \infty} p_{ij}^n = \varlimsup_{n \rightarrow \infty} p_{ij}^n = p_j$, $i, j = 1, \dots, N$, что и утверждается в теореме.

Фиксируем любые две строки α и β матрицы π_ν т. е. рассмотрим строки $p_{\alpha j}^\nu$ и $p_{\beta j}^\nu$, $j = 1, \dots, N$. Обозначим $\sum_k^+$ суммирование по тем k , для которых $p_{\alpha k}^\nu \geq p_{\beta k}^\nu$, а $\sum_k^-$ суммирование по тем k , для которых

$p_{\alpha k}^\nu < p_{\beta k}^\nu$, тогда

$$\sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) + \sum_k^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) = \sum_{k=1}^N p_{\alpha k}^\nu - \sum_{k=1}^N p_{\beta k}^\nu = 1 - 1 = 0. \quad (12.19)$$

Покажем, что

$$\sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) \leq 1 - \delta. \quad (12.20)$$

Действительно, поскольку в одну из сумм $\sum_k^+$ или $\sum_k^-$ входит матричный элемент j -го столбца, который по условию теоремы больше либо равен $\delta > 0$, см. рис. 1.23, то для любых $\alpha, \beta \in \{1, \dots, N\}$

$$\begin{aligned} \sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) &= \sum_{k=1}^N p_{\alpha k}^\nu - \sum_k^- p_{\alpha k}^\nu - \sum_k^+ p_{\beta k}^\nu = \\ &= 1 - \sum_k^- p_{\alpha k}^\nu - \sum_k^+ p_{\beta k}^\nu \leq 1 - \delta. \end{aligned} \quad (12.21)$$

А так как неравенство в (12.21) верно для любых $\alpha, \beta \in \{1, \dots, N\}$, то

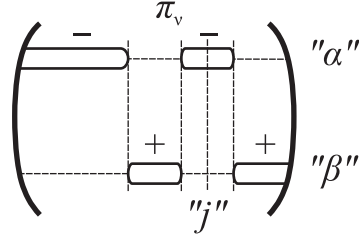


Рис. 1.23. Матрица π_ν с выбранными строками « α » и « β » и столбцом « j », все элементы которого $\geq \delta > 0$. Выделены матричные элементы строк « α » и « β », входящие в суммы $\sum^-$ и $\sum^+$.

и

$$\max_{1 \leq \alpha, \beta \leq N} \sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) \leq 1 - \delta. \quad (12.22)$$

Далее, при любом $n > 0$ с учетом (12.19) и (12.22):

$$\begin{aligned} M_j^{\nu+n} - m_j^{\nu+n} &= \max_i \sum_{k=1}^N p_{ik}^\nu p_{kj}^n - \min_i \sum_{k=1}^N p_{ik}^\nu p_{kj}^n = \\ &= \max_{\alpha, \beta} \sum_{k=1}^N (p_{\alpha k}^\nu - p_{\beta k}^\nu) p_{kj}^n \leq \max_{\alpha, \beta} \left\{ \sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) M_j^n + \right. \\ &\quad \left. + \sum_k^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) m_j^n \right\} = \max_{\alpha, \beta} \sum_k^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) (M_j^n - m_j^n) \leq \\ &\leq (1 - \delta) (M_j^n - m_j^n). \end{aligned} \quad (12.23)$$

Итак, $0 \leq M_j^{\nu+n} - m_j^{\nu+n} \leq (1-\delta)(M_j^n - m_j^n)$, n — любое. Переходя к пределу в этом неравенстве при $n \rightarrow \infty$, найдем

$$0 \leq M_j - m_j \leq (1-\delta)(M_j - m_j). \quad (12.24)$$

Так как $\delta > 0$, то $1-\delta < 1$ и из (12.24) следует равенство $M_j - m_j = 0$.

Пример 1.12.5. Пусть ξ_k , $k = 1, 2, \dots$, — взаимно независимые целочисленные одинаково распределенные случайные величины, $P\{\xi_k = i\} = p_i \geq 0$, $i = 0, 1, 2, \dots$, $\sum_{i=0}^{\infty} p_i = 1$. Пусть $\eta_n = \xi_1 + \dots + \xi_n$. Покажем, что случайные величины η_n , $n = 1, 2, \dots$, удовлетворяют условиям (1.12.1) и (1.12.3) однородной цепи Маркова. В самом деле, $\eta_{n+1} = \eta_n + \xi_{n+1}$, так что при заданном η_n распределение η_{n+1} известно. Подсчитаем условные вероятности:

$$\begin{aligned} p_{ij} &= P(\{\eta_{n+1} = j | \eta_n = i\}) = P(\{\eta_n + \xi_{n+1} = j | \eta_n = i\}) = \\ &= P(\{\xi_{n+1} = j - i\}) = \begin{cases} p_{j-i}, & j \geq i, \\ 0, & j < i, \end{cases} \quad i, j = 1, 2, \dots; \end{aligned}$$

таким образом цепь Маркова однородна.

Матрица перехода имеет вид

$$\pi_1 = \begin{pmatrix} p_0 & p_1 & p_2 & \dots \\ 0 & p_0 & p_1 & \dots \\ . & . & . & . \end{pmatrix}.$$

Пример 1.12.6. Рассмотрим цепь Маркова с двумя состояниями ω_1 и ω_2 , с вероятностями перехода $p_{11} = p_{22} = p$, $p_{12} = p_{21} = q$ ($0 \leq p \leq 1$, $p + q = 1$) и начальными вероятностями: $P(\{\omega_1\}) = a_1$, $P(\{\omega_2\}) = a_2$, $a_1 + a_2 = 1$. Найдем матрицу π_n перехода через n испытаний, абсолютные вероятности p_i^n и предельные вероятности p_i , $i = 1, 2$.

Поскольку в силу (12.9) $\pi_n = \pi_1^n$, то нам надо возвести матрицу перехода π_1 в n -ю степень. Для этого рассмотрим $\pi_1 = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$ как матрицу линейного оператора A в евклидовом пространстве $\mathcal{R}^2$ с базисом $f_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $f_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Найдем собственные векторы оператора A . Если $e = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$, то $\pi_1 e = \lambda e$, и мы получаем систему уравнений

$$\begin{cases} p\xi + q\eta = \lambda\xi, \\ q\xi + p\eta = \lambda\eta. \end{cases}$$

Характеристический определитель

$$\begin{vmatrix} p - \lambda & q \\ q & p - \lambda \end{vmatrix} = 0,$$

отсюда $\lambda_1 = 1$, $\lambda_2 = p - q$ (обозначим для удобства $p - q = c$) и собственные векторы: $e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. В базисе e_1, e_2 матрица $\hat{\pi}_1$ линейного оператора A диагональна $\hat{\pi}_1 = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$, и поэтому $\hat{\pi}_n = \hat{\pi}_1^n = \begin{pmatrix} 1 & 0 \\ 0 & c^n \end{pmatrix}$. Остается вернуться к старому базису. Если B — матрица перехода от $\{e_1, e_2\}$ к $\{f_1, f_2\}$, то, как известно, $\pi_n = B^{-1}\hat{\pi}_n B$. Найдем B ; если $B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, то $Be = f$, т.е. $\begin{cases} \alpha + \beta = 1 \\ \gamma + \delta = 0 \end{cases}$ и $\begin{cases} \alpha - \beta = 0 \\ \gamma - \delta = 1 \end{cases}$, отсюда $B = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ и $\pi_n = B^{-1}\hat{\pi}_n B = \frac{1}{2} \begin{pmatrix} 1 + c^n & 1 - c^n \\ 1 - c^n & 1 + c^n \end{pmatrix}$. Таким образом, $p_{11}^n = p_{22}^n = \frac{1}{2}(1 + (p - q)^n)$, $p_{12}^n = p_{21}^n = \frac{1}{2}(1 - (p - q)^n)$, $n = 1, 2, \dots$

Далее,

$$\begin{aligned} p_1^{n+1} &= a_1 p_{11}^n + a_2 p_{21}^n = \frac{1}{2} + \frac{1}{2}(a_1 - a_2)(p - q)^n, \\ p_2^{n+1} &= a_1 p_{12}^n + a_2 p_{22}^n = \frac{1}{2} - \frac{1}{2}(a_1 - a_2)(p - q)^n. \end{aligned}$$

Наконец,

$$\lim_{n \rightarrow \infty} p_{i1}^n = \lim_{n \rightarrow \infty} p_{i2}^n = \frac{1}{2}, \quad \lim_{n \rightarrow \infty} p_i^n = \frac{1}{2}, \quad i = 1, 2,$$

т.е. цепь Маркова — эргодическая.

1.13. Случайные процессы

Случайные процессы (случайные функции) — это случайные величины, зависящие от параметров, принимающих дискретное или непрерывное множество значений. Как правило, в автоматических системах, в системах, связанных с управлением, и т. п. приходится учитывать воздействие случайных помех, являющихся функциями времени, температуры, давления и т. д. Теория случайных процессов находит широкое применение во многих разделах физики и техники.

Определение 1.13.1. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, T — некоторое числовое множество. Действительная функция $\xi(t) = f(t, \omega)$, определенная при $t \in T$, $\omega \in \Omega$, называется случайным процессом (случайной функцией), если при каждом $t \in T$ $f(t, \omega)$ как функция $\omega \in \Omega$ является случайной величиной.

Если $\omega = \omega_0$ — фиксировано, то $f(t, \omega_0)$, рассматриваемая как функция t , $t \in T$, называется *реализацией* случайного процесса $\xi(t)$, или *выборочной функцией*. Если же фиксировано $t = t_0$, то случайная величина $\xi(t_0)$ называется *сечением* случайного процесса в точке $t = t_0$.

В каждом сечении распределение вероятностей случайного процесса задается одномерной функцией распределения $F(t, x) = P(\{\xi(t) < x\})$, $x \in \mathcal{R}^1$, $t \in T$. Однако эта функция не дает исчерпывающую

вероятностную характеристику случайного процесса $\xi(\cdot)$, поскольку не учитывает зависимости случайных величин $\xi(t)$ при разных $t \in T$ (т. е. зависимости различных сечений случайного процесса). Более полно вероятностные свойства $\xi(\cdot)$ описывает k -мерная функция распределения — функция распределения случайного вектора $(\xi(t_1), \dots, \xi(t_k))$, $(t_1, \dots, t_k) \in T^k$, для сечений случайного процесса:

$$F(t_1, x_1; \dots; t_k, x_k) = P(\{\xi(t_1) < x_1, \dots, \xi(t_k) < x_k\}), \\ (t_1, \dots, t_k) \in T^k, (x_1, \dots, x_k) \in \mathcal{R}^k$$

Однако практическое применение находят лишь функции распределения первого и второго порядков.

Рассмотренные в предыдущих параграфах последовательности независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$, или случайных величин, связанных в марковскую цепь, представляют собой примеры случайных процессов. Вообще любая последовательность случайных величин может быть интерпретирована как случайный процесс, при этом $T = \{1, 2, \dots\}$. Процессы, в которых T является подмножеством множества целых чисел, называются *случайными процессами с дискретным временем*, или *случайными последовательностями*¹⁾.

Переходим к изучению основных типов случайных процессов.

1.13.1. Процессы Пуассона и Винера. Эти два процесса являются наиболее важными представителями так называемых *процессов с независимыми приращениями*.

Начнем с процесса Пуассона, или *пуассоновского потока событий*. Рассмотрим некоторое событие A , которое может происходить в случайные моменты времени, и пусть $\xi(t)$ — число наступлений события A в промежутке времени длины t . Пусть выполнены следующие условия:

- 1) случайные величины $\xi(t_j)$, $j = 1, 2, \dots$, для непересекающихся промежутков времени независимы в совокупности;
- 2) для любого промежутка времени вероятность наступления события A в этом промежутке зависит лишь от его длины и не зависит от того, где на оси времени он расположен — свойство однородности процесса по времени;
- 3) при $t \rightarrow 0$

$$P(\{\xi(t) = 1\}) = \lambda t + o(t), \quad \lambda > 0; \quad P(\{\xi(t) > 1\}) = o(t).$$

Сформулированным условиям подчиняются многие реальные явления: распад радиоактивного вещества, отказы радиоэлектронной аппаратуры, вызовы на телефонной станции, запросы на обслуживание станков и т. д.

¹⁾ Случайная последовательность $\xi_1, \xi_2, \dots$, — последовательность случайных величин, рассматриваемая, как «единный объект», как функция $\xi: \{1, 2, \dots\} \rightarrow (\Omega, \mathcal{F}, P)$

Теорема 1.13.1. Если выполнены условия 1)–3), то распределение случайного вектора $(\xi(t_1), \dots, \xi(t_j))$ для непересекающихся промежутков времени $t_1, \dots, t_j$ является пуассоновским:

$$P(\{\xi(t_1) = k_1, \dots, \xi(t_j) = k_j\}) = \frac{(\lambda t_1)^{k_1}}{k_1!} \dots \frac{(\lambda t_j)^{k_j}}{k_j!} e^{-\lambda(t_1 + \dots + t_j)},$$

$$k_1, \dots, k_j = 0, 1, \dots \quad (13.1)$$

Доказательство. Ввиду условия 1) достаточно доказать, что

$$P(\{\xi(t) = k\}) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots \quad (13.2)$$

Чтобы использовать условие 3), разобьем промежутки времени t на n непересекающихся интервалов $\Delta_1, \dots, \Delta_n$ длины $\Delta = t/n$, а затем устремим $n \rightarrow \infty$. Событие $\{\xi(t) = k\}$ представим в виде $\{\xi(t) = k\} = A_1 + A_2$, где

$$A_1 = \bigcup_{\{i_1, \dots, i_k\}} \{\xi(\Delta_{i_1}) = 1, \xi(\Delta_{i_2}) = 1, \dots, \xi(\Delta_{i_k}) = 1, \xi(\Delta_{i_{k+1}}) = 0, \dots, \xi(\Delta_{i_n}) = 0\}, \quad (13.3)$$

т. е. A_1 представляет собой сумму всех тех событий, входящих в событие $\{\xi(t) = k\}$, при наступлении которых в каждом частичном интервале Δ_i , $i = 1, \dots, n$, событие A наступает не более одного раза, тогда как A_2 представляет собой сумму всех остальных событий, входящих в $\{\xi(t) = k\}$, т. е. таких, что по крайней мере в одном частичном интервале Δ_i событие A наступает не менее двух раз. Поскольку $A_2 \subset \bigcup_{i=1}^n \{\xi(\Delta_i) > 1\}$, то в силу свойства 3) при $n \rightarrow \infty$

$$P(A_2) \leq P\left(\bigcup_{i=1}^n \{\xi(\Delta_i) > 1\}\right) \leq \sum_{i=1}^n P(\{\xi(\Delta_i) > 1\}) = n o\left(\frac{t}{n}\right) = o(1). \quad (13.4)$$

Обозначим

$$p_0 = P(\{\xi(\Delta) = 0\}), \quad p_1 = P(\{\xi(\Delta) = 1\}). \quad (13.5)$$

Согласно (13.3) события, входящие в A_1 , попарно несовместны. Отсюда и из свойства 1) следует, что

$$P(A_1) = C_n^k p_1^k p_0^{n-k}. \quad (13.6)$$

Согласно (13.4) и (13.6) при $n \rightarrow \infty$

$$P(\{\xi(t) = k\}) = C_n^k p_1^k p_0^{n-k} + o(1). \quad (13.7)$$

Далее, ввиду свойства 3)

$$\begin{aligned} p_0 &= 1 - P(\{\xi(\Delta) \geq 1\}) = 1 - P(\{\xi(\Delta) = 1\}) - P(\{\xi(\Delta) > 1\}) = \\ &= 1 - \lambda\Delta + o(\Delta), \\ p_1 &= P(\{\xi(\Delta) = 1\}) = \lambda\Delta + o(\Delta), \quad \Delta = t/n. \end{aligned} \quad (13.8)$$

Подставляя (13.8) в (13.7) и переходя к пределу при $n \rightarrow \infty$, как при выводе теоремы Пуассона (см. §1.6), получим (13.2)

$$P(\{\xi(t) = k\}) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots \quad \blacksquare$$

Аргументом t случайного процесса $\xi(t)$, охарактеризованного в теореме 1.13.1, является длина промежутка времени. В силу однородности процесса в качестве промежутка времени можно взять интервал $(0, t)$. Обозначим $\eta(t) = \xi(t)$, где аргументом процесса $\eta(t)$ является текущий момент времени t , так что, например, $\eta(t)$ выражает число α -частиц, испущенных источником к моменту t , или число вызовов на телефонной станции, поступивших к моменту времени t .

Случайный процесс $\eta(t)$, $t \in [0, \infty)$, называется процессом Пуассона. Точное определение таково.

Определение 1.13.2. Случайный процесс $\eta(t)$, $0 \leq t < \infty$, называется процессом Пуассона, или пуассоновским потоком событий, если:

- 1) для любых $0 \leq t_1 < t_2 < \dots < t_n$ случайные величины $\eta(t_2) - \eta(t_1), \dots, \eta(t_n) - \eta(t_{n-1})$ независимы в совокупности (такие процессы называются процессами с независимыми приращениями);
- 2) случайная величина $\eta(t) - \eta(s)$, $t - s \geq 0$, имеет распределение Пуассона с параметром $\lambda(t - s)$:

$$P(\{\eta(t) - \eta(s) = k\}) = \frac{(\lambda(t - s))^k}{k!} e^{-\lambda(t - s)}, \quad k = 0, 1, \dots, \quad t > s.$$

Если дополнительно потребовать, чтобы $\eta(0) = 0$, то говорят, что процесс начинается в нуле. Если условиться считать каждую выборочную функцию процесса непрерывной справа, то последние являются функциями, возрастающими только целочисленными скачками, см. рис. 1.24.

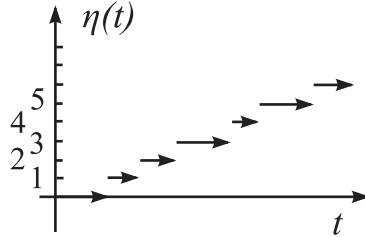


Рис. 1.24. Выборочная функция пуассоновского процесса.

Обозначим τ время ожидания первого события в пуассоновском потоке событий. Найдем *функцию распределения случайной величины τ* . Ясно, что $P(\{\tau \geq t\}) = P(\{\eta(t) - \eta(0) = 0\}) = e^{-\lambda t}$, и, таким образом, функция распределения времени ожидания:

$$F_\tau(t) = P(\{\tau < t\}) = 1 - e^{-\lambda t}, \quad 0 \leq t \leq \infty. \quad (13.9)$$

Следовательно, *время ожидания τ распределено с плотностью вероятности*

$$p_\tau(t) = \lambda e^{-\lambda t}, \quad 0 \leq t < \infty. \quad (13.10)$$

Точно так же, если время ожидания отсчитывается от произвольного момента t_1 , то

$$P(\{\tau \geq t\}) = P(\{\eta(t + t_1) - \eta(t_1) = 0\}) = e^{-\lambda t}. \quad (13.11)$$

Поэтому время ожидания очередного события не уменьшается от того, что это событие уже ожидалось, скажем, и до момента t_1 . В самом деле, так как $\{\tau - t_1 \geq t\} \cap \{\tau \geq t_1\} = \{\tau - t_1 \geq t\}$, то

$$\begin{aligned} P(\{\tau - t_1 \geq t | \tau \geq t_1\}) &= \frac{P(\{\tau - t_1 \geq t\})}{P(\{\tau \geq t_1\})} = \frac{e^{-\lambda(t+t_1)}}{e^{-\lambda t_1}} = \\ &= e^{-\lambda t} = P(\{\tau \geq t\}). \end{aligned} \quad (13.12)$$

Пусть τ_k — случайный момент времени k -го наступления события в пуассоновском потоке. Так как $P(\{\xi(t) \geq k\}) = P(\{\tau_k \leq t\}) = \sum_{j=k}^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t}$, $t \geq 0$, то τ_k — абсолютно непрерывная случайная величина с плотностью

$$\begin{aligned} \frac{dP(\{\tau_k \leq t\})}{dt} &= \sum_{j=k}^{\infty} \left(\lambda j \frac{(\lambda t)^{j-1}}{j!} - \lambda \frac{(\lambda t)^j}{j!} \right) e^{-\lambda t} = \\ &= \frac{\lambda (\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}, \quad t \geq 0, \quad k = 1, 2, \dots \end{aligned}$$

Вероятностный смысл параметра λ определяется равенством

$$M\tau = \int_0^{\infty} t p_\tau(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}, \quad (13.13)$$

т. е. λ — *среднее число событий, происходящих в единицу времени*.

Пример 1.13.1. Вероятность радиоактивного распада для каждого атома к моменту t равна $P(\{\tau < t\}) = 1 - e^{-\lambda t}$. Следовательно, к моменту t в среднем распадается $1 - e^{-\lambda t}$ часть всего вещества, и так называемая постоянная t_0 полураспада определяется из равенства $1 - e^{-\lambda t} = 1/2$, т. е. $t_0 = (\ln 2)/\lambda$; для радия экспериментально найдено, что $t_0 = 1590$ лет.

Пример 1.13.2. Покажем, что для пуассоновского процесса, для которого

$$P(\{\xi(t) = k\}) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0, \quad t > 0,$$

имеет место сходимость по вероятности:

$$\frac{\xi(t)}{t} \xrightarrow{P} \lambda \quad \text{при } t \rightarrow \infty.$$

В самом деле, как показано в §1.9, $M\xi(t) = \lambda t$, $D\xi(t) = \lambda t$. Поэтому для случайной величины $\xi(t)/t$ с помощью неравенства Чебышева (10.4) получим при любом $\varepsilon > 0$

$$P\left(\left|\frac{\xi(t)}{t} - \lambda\right| > \varepsilon\right) \leq \frac{D\left(\frac{\xi(t)}{t}\right)}{\varepsilon^2} = \frac{\lambda}{t\varepsilon^2} \rightarrow 0 \quad \text{при } t \rightarrow \infty,$$

что и доказывает высказанное утверждение.

Определение 1.13.3. Винеровским процессом $\xi(t)$, $0 \leq t < \infty$, начинающимся в нуле, называется случайный процесс, обладающий свойствами:

- 1) для любых $0 \leq t_0 < t_1 < \dots < t_n$ случайные величины $\eta_1 = \xi(t_1) - \xi(t_0)$, $\dots$, $\eta_n = \xi(t_n) - \xi(t_{n-1})$ взаимно независимы;
- 2) случайная величина $\xi(t) - \xi(s)$ ($0 \leq s < t$ — любые) имеет нормальное распределение $N(0, t - s)$;
- 3) $\xi(0) = 0$.

Согласно определению случайный вектор $\eta = (\eta_1, \dots, \eta_n)$ распределен нормально с плотностью

$$p_\eta(x_1, \dots, x_n) = \prod_{j=1}^n (2\pi(t_j - t_{j-1}))^{-1/2} \exp\left(-\frac{x_j^2}{2(t_j - t_{j-1})}\right), \quad (13.14)$$

$$(x_1, \dots, x_n) \in \mathcal{R}^n.$$

Полагая $t_0 = 0$, в силу условия $\xi(0) = 0$ найдем $\xi(t_1) = \eta_1$, $\xi(t_2) = \eta_1 + \eta_2$, $\dots$, $\xi(t_n) = \eta_1 + \dots + \eta_n$, т. е. случайные векторы $\eta = (\eta_1, \dots, \eta_n)$ и $\xi = (\xi(t_1), \dots, \xi(t_n))$ связаны невырожденным линейным преобразованием с определителем, равным единице. Поэтому случайный вектор ξ также нормальный и имеет плотность вероятности, задаваемую формулой

$$p_\xi(x_1, \dots, x_n) = \prod_{j=1}^n (2\pi(t_j - t_{j-1}))^{-1/2} \exp\left(-\frac{(x_j - x_{j-1})^2}{2(t_j - t_{j-1})}\right),$$

$$x_0 = 0, \quad (x_1, \dots, x_n) \in \mathcal{R}^n.$$

Пример 1.13.3. Модель броуновского движения. Представим себе частицу, взвешенную в однородной жидкости. Она испытывает хаотические столкновения с молекулами жидкости и в результате беспорядочно движется. Это так называемое броуновское движение частицы. Рассмотрим одномерную дискретную модель броуновского движения,

в которой частица изменяет свое положение лишь скачками в фиксированные моменты времени, кратные Δt . Обозначим величину скачка Δx . Тогда, находясь в момент t в точке x , частица в следующий момент $t + \Delta t$ с равными вероятностями окажется в одной из соседних точек $x - \Delta x$ или $x + \Delta x$. Модель броуновского движения можно получить в результате предельного перехода при $\Delta t \rightarrow 0$, $\Delta x \rightarrow 0$.

Смещение частицы к моменту $t + s$ равно

$$\xi(t + s) = [\xi(t + s) - \xi(s)] + [\xi(s) - \xi(0)], \quad \xi(0) = 0, \quad (13.15)$$

причем в рассматриваемой модели случайные величины в скобках, очевидно, независимы для любых $s, t \geq 0$ и, более того, распределение вероятностей $\xi(t + s) - \xi(s)$ точно такое же, как распределение $\xi(t) - \xi(0) = \xi(t)$. Поэтому

$$D\xi(t + s) = D[\xi(t + s) - \xi(s)] + D[\xi(s) - \xi(0)] = D\xi(t) + D\xi(s).$$

Отсюда следует, что дисперсия является линейной функцией времени

$$D\xi(t) = \sigma^2 t, \quad 0 \leq t < \infty, \quad (13.16)$$

постоянная σ^2 называется *коэффициентом диффузии*.

Пусть в начальный момент броуновская частица находилась в нуле, $\xi(0) = 0$. Тогда к моменту t она совершит $n = t/\Delta t$ скачков, и если обозначить $\xi_1, \dots, \xi_n$ случайные величины, принимающие значения $\pm \Delta x$ с вероятностью $1/2$, то $\xi(t) = \sum_{k=1}^{t/\Delta t} \xi_k$, $M\xi(t) = 0$, $D\xi(t) = \sum_{k=1}^{t/\Delta t} D\xi_k$, а поскольку $D\xi_k = \frac{1}{2}(\Delta x)^2 + \frac{1}{2}(\Delta x)^2 = (\Delta x)^2$, то

$$D\xi(t) = (\Delta x)^2 \cdot \frac{t}{\Delta t}. \quad (13.17)$$

Сравнивая (13.16) и (13.17), найдем, что $\sigma^2 = (\Delta x)^2/\Delta t$.

Рассмотрим случайную величину

$$\xi_n(t) = \frac{\xi(t)}{\sqrt{D\xi(t)}} = \frac{\xi(t)}{\sigma\sqrt{t}} = \frac{1}{\sigma\sqrt{t}} \sum_{k=1}^{\frac{t}{\Delta t}} \xi_k, \quad n = \frac{t}{\Delta t}. \quad (13.18)$$

Поскольку $M\xi_n(t) = 0$ и $D\xi_n(t) = 1$, то в силу центральной предельной теоремы при $\Delta t \rightarrow 0$ и $\Delta x \rightarrow 0$, так что $(\Delta x)^2/\Delta t = \sigma^2$ и $n \rightarrow \infty$, $\xi_n(t)$ *сходится по распределению к нормальному закону*. Поэтому для предельной случайной величины $\tilde{\xi}(t)$, которую естественно рассматривать как «непрерывную» *модель броуновского движения*, получим

$$P\left(\left\{x_1 \leq \frac{\tilde{\xi}(t)}{\sigma\sqrt{t}} \leq x_2\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-x^2/2} dx. \quad (13.19)$$

Мы нашли распределение вероятностей $\tilde{\xi}(t)$ — положений броуновской частицы в момент $t \in [0, \infty)$. Это же распределение справедливо и для любого смещения частицы за время t , ибо в силу однородности процесса броуновского движения приращение $\tilde{\xi}(t+s) - \tilde{\xi}(s)$ имеет такое же распределение вероятностей, как и приращение $\xi(t) - \xi(0) = \xi(t)$, и, стало быть,

$$P\left(\left\{x_1 \leq \frac{\tilde{\xi}(t+s) - \tilde{\xi}(s)}{\sigma\sqrt{t}} \leq x_2\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-x^2/2} dx, \quad x_1, x_2 \in \mathcal{R}^1. \quad (13.20)$$

Таким образом, броуновское движение частицы — винеровский процесс.

В общем случае винеровский процесс, в отличие от рассмотренного в примере, определяется так, что случайная величина $\xi(t) - \xi(s)$, $0 \leq s < t$, распределена нормально $N(\vartheta(t-s), \sigma^2(t-s))$, где ϑ — коэффициент сноса, σ^2 — коэффициент диффузии. Ранее мы ввели так называемый *стандартный винеровский процесс*: $\xi(t) - \xi(s) \sim N(0, t-s)$. Общий случай получается в результате преобразования $\xi(t) \rightarrow \sigma\xi(t) + \vartheta(t)$.

Оба рассмотренных процесса, пуассоновский и винеровский, являются примерами однородных процессов с независимыми приращениями $\xi(t_1) - \xi(t_0)$, $\xi(t_2) - \xi(t_1)$, ... *Однородным*, как уже сказано выше, называется случайный процесс, у которого распределение $\xi(t) - \xi(s)$ зависит только от разности $t-s$. Отличное от условий 1)–3) определение винеровского процесса мы дадим при рассмотрении непрерывных марковских процессов.

Для винеровского процесса 1)–3) можно показать, что его траектории (реализации) непрерывны с вероятностью единица. Это позволяет ввести для винеровского процесса (в частности, непрерывного броуновского движения) важную случайную величину τ_x — *момент времени первого достижения траекторией точки x , иначе говоря τ_x — время ожидания события: первый раз $\{\xi(t) = x\}$, $\xi(0) = 0$* . Пусть $x > 0$. Частица может оказаться правее точки x , $\xi(t) \geq x$, лишь при условии, что до этого она побывала в точке x (так как траектория непрерывна), т. е. $\tau_x \leq t$. Таким образом, событие $\{\xi(t) \geq x\}$ влечет событие $\{\tau_x \leq t\}$. Поэтому

$$P(\{\xi(t) \geq x\} | \{\tau_x \leq t\}) = \frac{P(\{\xi(t) \geq x\})}{P(\{\tau_x \leq t\})}, \quad -\infty < x < \infty, \quad 0 < t < \infty. \quad (13.21)$$

Очевидно, что если в момент τ_x , $\tau_x \leq t$, частица находится в точке x , то после выхода из x к моменту t частица с вероятностью $1/2$ окажется правее x , и с той же вероятностью она окажется левее x . Поэтому

$P(\{\xi(t) \geq x | \tau_x \leq t\}) = 1/2$ и из (13.21) находим

$$\begin{aligned} P(\{\tau_x \leq t\}) &= 2P(\{\xi(t) \geq x\}) = 2P\left(\left\{\frac{\xi(t)}{\sigma\sqrt{t}} \geq \frac{x}{\sigma\sqrt{t}}\right\}\right) = \\ &= \sqrt{\frac{2}{\pi}} \int_{x/(\sigma\sqrt{t})}^{\infty} \exp(-\frac{z^2}{2}) dz, \quad t > 0. \end{aligned} \quad (13.22)$$

Поэтому плотность распределения случайной величины τ_x

$$p_{\tau_x}(t) = \frac{x}{\sqrt{2\pi}\sigma t^{3/2}} \exp(-\frac{x^2}{2\sigma^2 t}), \quad 0 < t < \infty. \quad (13.23)$$

Рассмотрим теперь другую важную случайную величину: $\hat{\xi}(t)$ — *максимальное смещение* броуновской частицы за время t , $\hat{\xi}(t) = \max_{0 \leq s \leq t} \xi(s)$, $\xi(0) = 0$. Согласно (13.22)

$$\begin{aligned} P(\{\hat{\xi}(t) \geq x\}) &\equiv P(\{\max_{0 \leq s \leq t} \xi(s) \geq x\}) = P(\{\tau_x \leq t\}) = \\ &= \sqrt{\frac{2}{\pi}} \int_{x/(\sigma\sqrt{t})}^{\infty} \exp(-\frac{z^2}{2}) dz. \end{aligned} \quad (13.24)$$

Отсюда, поскольку $\sqrt{\frac{2}{\pi}} \int_0^{\infty} \exp(-\frac{z^2}{2}) dz = 1$, следует, что

$$F_{\hat{\xi}(t)}(x) = 1 - P(\{\hat{\xi}(t) \geq x\}) = \sqrt{\frac{2}{\pi}} \int_0^{x/(\sigma\sqrt{t})} \exp(-\frac{z^2}{2}) dz, \quad x \in \mathcal{R}^1, t > 0, \quad (13.25)$$

и, значит, $\hat{\xi}(t)$ имеет плотность распределения

$$p_{\hat{\xi}(t)}(x) = \frac{1}{\sigma} \sqrt{\frac{2}{\pi t}} e^{-\frac{x^2}{2\sigma^2 t}}, \quad 0 \leq x < \infty, \quad t > 0, \quad (13.26)$$

$(p_{\hat{\xi}(t)}(x) = 0$ при $x < 0$, так как $\hat{\xi}(t) = \max_{0 \leq s \leq t} \xi(s) \geq \xi(0) = 0$). Распределение (13.25) называется удвоенным нормальным законом, поскольку ввиду (13.22) и (13.24) $P(\{\hat{\xi}(t) \geq x\}) = 2P(\{\xi(t) \geq x\})$.

1.13.2. Процессы второго порядка (Корреляционная теория).

Далее будут встречаться комплексные случайные процессы $\xi(t) = \eta(t) + i\zeta(t)$, $t \in [0, T]$, где $\eta(\cdot)$ и $\zeta(\cdot)$ — действительные случайные процессы.

Ковариационной функцией случайного процесса $\xi(t)$, $t \in [0, T]$, называется

$$K(t, s) = M(\xi(t) - M\xi(t))(\overline{\xi(s)} - \overline{M\xi(s)}), \quad s, t \in [0, T].$$

Функция $\tilde{K}(t, s) = M\xi(t)\bar{\xi}(s)$, $s, t \in [0, T]$ называется *корреляционной*, $K(t, s) = \tilde{K}(t, s) - M\xi(t)M\bar{\xi}(s) = \overline{K(s, t)}$, $M|K(t, s)|^2 \leq M|\xi(t) - M\xi(t)|^2 \cdot M|\xi(s) - M\xi(s)|^2$, $s, t \in [0, T]$.

Пример 1.13.4. Ковариационная функция винеровского случайного процесса.

$$\begin{aligned} M\xi(t)\xi(s) &= M[(\xi(t) - \xi(s)) + (\xi(s) - \xi(0))](\xi(s) - \xi(0)) \\ &= M(\xi(s) - \xi(0))^2 = s, \quad 0 < s < t < \infty. \end{aligned}$$

Следовательно, $M\xi(t)\xi(s) = \min(t, s)$, $s, t \in [0, T]$.

Пример 1.13.5. Ковариационная функция пуассоновского процесса. Так как $P(\{\xi(t) = k\}) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$, $k = 0, 1, \dots$, $M\xi(t) = \lambda t$, $D\xi(t) = \lambda t$, то для $0 < s < t < \infty$ $K(t, s) = M(\xi(t) - \lambda t)(\xi(s) - \lambda s) = M[(\xi(t) - \xi(s) - \lambda(t - s)) + (\xi(s) - \lambda s)](\xi(s) - \lambda s) = M(\xi(s) - \lambda s)^2 = \lambda s$. Поэтому и в этом случае $K(t, s) = \lambda \min(t, s)$, $t, s \in (0, \infty)$.

В обоих примерах использована независимость приращений.

Охарактеризуем вначале свойства в среднем квадратичном дискретного случайного процесса как характеристик в среднем квадратичном последовательности его сечений (как последовательности случайных величин). Напомним, что последовательность случайных величин $\{\xi_k\}$ сходится в среднем квадратичном (с.к.) к случайной величине ξ , $\xi = \text{l.i.m.}_{k \rightarrow \infty} \xi_k$, если при $k \rightarrow \infty$ $M|\xi_k - \xi|^2 \rightarrow 0$.

Критерий Коши с.к. сходимости. Для сходимости в с.к. последовательности случайных величин $\{\xi_k\}$ (к некоторой случайной величине¹⁾) необходимо и достаточно, чтобы $M|\xi_k - \xi_n|^2 \rightarrow 0$ при $k, n \rightarrow \infty$.

Свойства с.к. сходимости. Если $\xi = \text{l.i.m.}_{k \rightarrow \infty} \xi_k$, то

1. $M\xi = \lim_{k \rightarrow \infty} M\xi_k$. Действительно, $|M\xi - M\xi_k| \leq M|\xi - \xi_k| \leq (M|\xi - \xi_k|^2)^{1/2} \rightarrow 0$ при $k \rightarrow \infty$.

2. $M|\xi|^2 = \lim_{k \rightarrow \infty} M|\xi_k|^2$. Действительно, $M|\xi - \xi_k|^2 = M|\xi|^2 - M\xi\bar{\xi}_k - M\xi_k\bar{\xi} + M|\xi_k|^2 \geq M|\xi|^2 - 2(M|\xi|^2 M|\xi_k|^2)^{1/2} + M|\xi_k|^2 = (\sqrt{M|\xi|^2} - \sqrt{M|\xi_k|^2})^2$. Более того, если и $\eta = \text{l.i.m.}_{k \rightarrow \infty} \eta_k$, то $M\xi\bar{\eta} = \lim_{k \rightarrow \infty} M\xi_k\bar{\eta}_k$.

Это утверждение и равенство $M\xi\bar{\eta} = \lim_{k \rightarrow \infty} M\xi_k\bar{\eta}_k$ следуют из критерия Лозэва.

Критерий Лозэва с.к. сходимости. Последовательность $\{\xi_k\}$ сходится в среднем квадратичном, если и только если сходится числовая последовательность $M\xi_k\bar{\xi}_n \rightarrow M\xi\bar{\xi}$ при $k, n \rightarrow \infty$.

¹⁾ Это свойство полноты гильбертова пространства $L^2(\Omega)$ случайных величин $\xi(\omega)$, $\omega \in \Omega$, с конечным моментом $M|\xi(\cdot)|^2 = \int |\xi(\omega)|^2 dP(\omega)$, (где $\int$ — символ интеграла Лебега).

Доказательство. Достаточность. При $k, n \rightarrow \infty$ $M(\xi_k - \xi_n)(\bar{\xi}_k - \bar{\xi}_n) = M\xi_k\bar{\xi}_k + M\xi_n\bar{\xi}_n - M\xi_k\bar{\xi}_n - M\xi_n\bar{\xi}_k \rightarrow 0$, т.е. выполнен критерий Коши, если последовательность $M\xi_k\bar{\xi}_n$ сходится при $k, n \rightarrow \infty$.

Необходимость. При $k, n \rightarrow \infty$ $M\xi_k\bar{\xi}_n - M\xi\bar{\xi} = M(\xi_k - \xi)(\bar{\xi}_n - \bar{\xi}) + M\xi(\bar{\xi}_n - \bar{\xi}) + M(\xi_k - \xi)\bar{\xi} \rightarrow 0$, если $\xi_k \xrightarrow{c.k.} \xi$, ибо $|M(\xi_k - \xi)(\bar{\xi}_n - \bar{\xi})| \leq (M|\xi_k - \xi|^2 M|\bar{\xi}_n - \bar{\xi}|^2)^{1/2}$, $|M\xi(\bar{\xi}_n - \bar{\xi})| \leq (M|\xi|^2 M|\bar{\xi}_n - \bar{\xi}|^2)^{1/2}$, $|M(\xi_k - \xi)\bar{\xi}| \leq (M|\xi_k - \xi|^2 M|\bar{\xi}|^2)^{1/2}$. ■

Замечание 1.13.1. Если $\xi_k \xrightarrow[k \rightarrow \infty]{c.k.} \xi$ «достаточно быстро», т.е., если

$$\sum_{k=1}^{\infty} M|\xi - \xi_k|^2 < \infty, \text{ то } \xi_k \xrightarrow[k \rightarrow \infty]{п.н.} \xi.$$

Действительно, пусть $A_k = \{|\xi_k - \xi| > e\}$, $k = 1, 2, \dots$. Тогда $P(A_k) < \frac{M|\xi - \xi_k|^2}{e^2}$, откуда следует, что $\sum_{k=1}^{\infty} P(A_k) < \infty$. Следовательно, согласно лемме Бореля—Кантелли с вероятностью единица происходит не более конечного числа событий среди $A_1, A_2, \dots$ для любого $e > 0$. Это означает, что $\xi_k \xrightarrow[k \rightarrow \infty]{п.н.} \xi$. ■

Перечисленные свойства дискретного случайного процесса нетрудно переформулировать как *свойства в среднем квадратичном случайного процесса* $\xi(t)$, $-\infty < t < \infty$. Например,

Определение 1.13.4. Комплексная случайная величина ξ — *предел в среднем квадратичном случайной функции* $\xi(t)$, $-\infty < t < \infty$ при $t \rightarrow t_0$, если при $t \rightarrow t_0$ $M|\xi(t) - \xi|^2 \rightarrow 0$. Соответственно случайный процесс $\xi(t)$, $t \in [0, T]$ называется *непрерывным в среднем квадратичном* (с.к.) *в точке* $t \in [0, T]$, если $M|\xi(t+h) - \xi(t)|^2 \rightarrow 0$ при $h \rightarrow 0$, или, иначе говоря, если $\xi(t) = \lim_{h \rightarrow 0} \xi(t+h)$, а если последнее равенство выполнено для всех $t \in [0, T]$, то случайный процесс $\xi(\cdot)$ с.к. непрерывен на $[0, T]$. Речь идет о случайных процессах с конечными моментами второго порядка, $M|\xi(t)|^2 < \infty$, $-\infty < t < \infty$, при этом $|M\xi(t)| \leq M|\xi(t)| \leq [M|\xi(t)|^2]^{1/2} < \infty$, $-\infty < t < \infty$.

Всюду далее будем считать, что $M\xi(t) = 0$, $t \in [0, T]$.

Теорема 1.13.2. 1. Случайный процесс $\xi(t)$, $t \in [0, T]$ с.к. непрерывен в точке t_0 тогда и только тогда, когда его ковариационная функция $K(t, s)$, $t, s \in [0, T]$, непрерывна в точке (t_0, t_0) .

2. Если $K(t, s)$, $t, s \in [0, T]$, непрерывна на диагонали $t = s$, то $K(\cdot, \cdot)$ непрерывна всюду.

Доказательство. 1. Достаточность. Пусть $K(\cdot, \cdot)$ непрерывна в точке (t_0, t_0) . Тогда $M|\xi(t_0+h) - \xi(t_0)|^2 = K(t_0+h, t_0+h) - K(t_0+h, t_0) - K(t_0, t_0+h) + K(t_0, t_0) \rightarrow 0$ при $h \rightarrow 0$.

Необходимость. Пусть $\xi(\cdot)$ с.к. непрерывна при $t = t_0$. Так как

$$\begin{aligned} K(t_0+h, t_0+k) - K(t_0, t_0) &= M[(\xi(t_0+h) - \xi(t_0))(\bar{\xi}(t_0+k) - \bar{\xi}(t_0)) + \\ &+ (\xi(t_0+h) - \xi(t_0))\bar{\xi}(t_0) + \xi(t_0)(\bar{\xi}(t_0+k) - \bar{\xi}(t_0))], \end{aligned}$$

то для доказательства непрерывности $K(\cdot, \cdot)$ в (t_0, t_0) достаточно показать, что все (три) слагаемые в правой части стремятся к нулю при $h \rightarrow 0$ и $k \rightarrow 0$. Это следует из с.к. непрерывности $\xi(\cdot)$ в t_0 и оценок

$$\begin{aligned} & |M(\xi(t_0 + h) - \xi(t_0))(\bar{\xi}(t_0 + k) - \bar{\xi}(t_0))| \leq \\ & \leq \sqrt{M|\xi(t_0 + h) - \xi(t_0)|^2 M|\bar{\xi}(t_0 + k) - \bar{\xi}(t_0)|^2}, \\ & |M(\xi(t_0 + h) - \xi(t_0))\bar{\xi}(t_0)| \leq \sqrt{M|\xi(t_0 + h) - \xi(t_0)|^2 M|\bar{\xi}(t_0)|^2} \end{aligned}$$

и

$$|M(\xi(t_0)(\bar{\xi}(t_0 + k) - \bar{\xi}(t_0)))| \leq \sqrt{M|\xi(t_0)|^2 M|\bar{\xi}(t_0 + k) - \bar{\xi}(t_0)|^2}$$

(по существу речь идет о критерии Лозева).

2. Если $K(t, s)$ $t, s \in [0, T]$, непрерывна на диагонали, например, в точках $(t, t), (s, s)$, то при $h \rightarrow 0$ $\xi(t + h) \xrightarrow{\text{с.к.}} \xi(t)$ и $\xi(s + h) \xrightarrow{\text{с.к.}} \xi(s)$. Следовательно, при $h \rightarrow 0, k \rightarrow 0$, $M\xi(t + h)\bar{\xi}(s + k) \rightarrow M\xi(t)\bar{\xi}(s)$, т.е. $K(t + h, s + k) \rightarrow K(t, s)$, $t, s \in [0, T]$. ■

Следствие 1.13.1. Пуассоновский и винеровский случайные процессы непрерывны в среднем квадратичном, поскольку их ковариационные функции непрерывны. Однако реализации пуассоновского процесса — разрывны с вероятностью единица (ступенчатые) функции. Следовательно с.к. непрерывность случайного процесса $\xi(t)$, $t \in [0, T]$ не влечет непрерывность его выборочных функций.

Определение 1.13.5. Случайная функция $\xi(\cdot)$ называется с.к. дифференцируемой в точке t , если существует с.к. предел $\frac{\xi(t + h) - \xi(t)}{h}$ при $h \rightarrow 0$. Этот предел называется с.к. производной $\xi(\cdot)$,

$$\xi'(t) = \text{l.i.m.}_{h \rightarrow 0} \frac{\xi(t + h) - \xi(t)}{h}, \text{ или } M\left|\frac{\xi(t + h) - \xi(t)}{h} - \xi'(t)\right|^2 \rightarrow 0, h \rightarrow 0.$$

Согласно критерию Лозева, $\xi(\cdot)$ с.к. дифференцируема в точке t , если и только если дробь

$$Q = \frac{1}{hk} M(\xi(t + h) - \xi(t))(\xi(t + k) - \xi(t)) \text{ имеет предел при } h, k \rightarrow 0.$$

$$Q = \frac{K(t + h, t + k) - K(t, t + k) - K(t + h, t) + K(t, t)}{hk} \rightarrow \left. \frac{\partial^2 K(t, s)}{\partial t \partial s} \right|_{t=s}.$$

Здесь слева разностное отношение для второй производной. Оно сходится при $h, k \rightarrow 0$ к $\left. \frac{\partial^2 K(t, s)}{\partial t \partial s} \right|_{t=s}$ если, например, смешанная производная $\frac{\partial^2 K(t, s)}{\partial t \partial s}$, $t, s \in [0, T]$, непрерывна.

В этом случае $\frac{\partial^2 K(t, s)}{\partial t \partial s} = M\xi'(t)\xi'(s)$ — ковариационная функция с.к. производной $\xi(\cdot)$. Это утверждение — следствие свойства 2 ковариационной функции, приведенного в теореме 1.13.2, а именно если при $h, k \rightarrow 0$ $\frac{\xi(t+h) - \xi(t)}{h} \rightarrow \xi'(t)$, $\frac{\xi(s+k) - \xi(s)}{k} \rightarrow \xi'(s)$, то $M \frac{\xi(t+h) - \xi(t)}{h} \frac{\xi(s+k) - \xi(s)}{k} \rightarrow M\xi'(t)\xi'(s) = \frac{\partial^2 K(t, s)}{\partial t \partial s} \Big|_{t=s}$, если эта производная с.к. непрерывна. При этом согласно свойствам 1), 2) с.к. сходимости $M\xi'(t) = \lim_{h \rightarrow 0} \frac{1}{h} M(\xi(t+h) - \xi(t))$, $M|\xi'(t)|^2 = \lim_{h \rightarrow 0} \frac{1}{h^2} |\xi(t+h) - \xi(t)|^2$.

Определение 1.13.6. Случайная функция $\xi(t)$, $t \in [0, T]$, с.к. интегрируема по Риману на $[a, b]$, если последовательность римановских интегральных сумм $\sum_{[a, b]} \xi(t_k) \Delta t_k$ с.к. сходится при $\max_k \Delta t_k \rightarrow 0$.

Теорема 1.13.3. Случайная функция $\xi(\cdot)$ с.к. интегрируема на $[a, b]$ по Риману, если $K(\cdot, \cdot)$ интегрируема по Риману на $[a, b] \times [a, b]$. Доказательство. Действительно, пусть $S = \sum_{[a, b]} \xi(t_k) \Delta t_k$, $S' = \sum_{[a, b]} \xi(t'_k) \Delta t'_k$. Тогда $MSS' = \sum_{[a, b] \times [a, b]} K(t_i, t'_k) \Delta t_i \Delta t'_k$ — интегральная сумма интеграла $\int_{[a, b] \times [a, b]} K(t, s) dt ds$, и факт существования предела интегральной суммы при $\max_{i, k} (\Delta t_i, \Delta t'_k) \rightarrow 0$ эквивалентен с.к. интегрируемости $\xi(\cdot)$ на $[a, b]$ (с.к. сходимости S). ■

1.13.3. Марковские процессы. Мы уже рассматривали частный случай марковских процессов — конечные однородные цепи Маркова. В этом пункте мы изучим основные свойства марковских процессов с непрерывным временем.

Определение 1.13.7. Случайный процесс $\xi(\cdot)$ называется *марковским*, если для любого момента времени t_1 при известном значении $\xi(t_1)$ случайные величины $\xi(t)$ с $t > t_1$ не зависят от случайных величин $\xi(s)$ с $s < t_1$. Таким образом, марковские процессы характеризуются тем, что вероятностные свойства процесса в момент $t \geq t_1$ определяются состоянием в момент t_1 , и не зависят от состояний процесса до момента t_1 , сравните с примером 1.12.1.

1.13.3.1 Процессы с дискретным множеством состояний. Рассмотрим вначале марковские процессы с непрерывным временем и с конечным или счетным множеством состояний $\Omega = (\omega_1, \omega_2, \dots)$. Остановимся подробно на случае конечного числа состояний, он отличается от цепей Маркова тем, что здесь время изменяется непрерывно, и переход системы из одного состояния ω_i в другое

ω_j происходит в любой момент времени. Введем *переходные вероятности*, см определение 1.12.1,

$$p_{ij}(s, t) = P(\xi(t) = \omega_j | \{\xi(s) = \omega_i\}), \quad p_{ij}(s, t) \geq 0, \\ s \leq t, \quad s, t \in (-\infty, \infty), \quad i, j = 1, \dots, N. \quad (13.27)$$

Очевидно,

$$\sum_{j=1}^N p_{ij}(s, t) = 1, \quad s \leq t, \quad p_{ii}(t, t) = 1, \quad p_{ij}(t, t) = 0, \quad i \neq j, \\ -\infty < s < t < \infty, \quad i, j = 1, \dots, N. \quad (13.28)$$

Обозначим $a_i, i = 1, \dots, N$, начальное распределение вероятностей при $t = t_0$,

$$a_i \geq 0, \quad \sum_{i=1}^N a_i = 1, \quad (13.29)$$

и $p_j(t)$ — *абсолютную вероятность*, т. е. вероятность того, что система будет в состоянии ω_j в момент $t, t \geq t_0, j = 1, \dots, N$. В силу формулы полной вероятности очевидны следующие равенства:

$$p_j(t) = \sum_{k=1}^N a_k p_{kj}(t_0, t), \quad j = 1, \dots, N, \quad t \geq t_0, \quad (13.30)$$

$$p_{ij}(s, t) = \sum_{k=1}^N p_{ik}(s, t_1) p_{kj}(t_1, t), \quad i, j = 1, \dots, N, \quad s < t_1 < t. \quad (13.31)$$

Теорема 1.13.4. Пусть переходные вероятности $p_{ij}(s, t)$ имеют частные производные по t и $s, s, t \in (t_0, \infty)$. Тогда при $s \leq t$

$$\frac{\partial p_{ij}(s, t)}{\partial t} = \sum_{k=1}^N p_{ik}(s, t) A_{kj}(t), \quad i, j = 1, \dots, N, \quad (13.32)$$

$$\frac{\partial p_{ij}(s, t)}{\partial s} = - \sum_{k=1}^N A_{ik}(s) p_{kj}(s, t), \quad i, j = 1, \dots, N, \quad (13.33)$$

где

$$A_{ik}(t) = \left[\frac{\partial p_{ik}(t, u)}{\partial u} \right] \Big|_{u=t}, \quad A_{ik}(t) \geq 0, \quad i \neq k, \quad A_{kk}(t) \leq 0, \\ \sum_{k=1}^N A_{ik}(t) = 0, \quad i, k = 1, \dots, N. \quad (13.34)$$

Доказательство. Для $t > s$ в силу (13.31)

$$\begin{aligned} \frac{\partial p_{ij}(s, t)}{\partial t} &= \lim_{\Delta \rightarrow 0} \frac{p_{ij}(s, t + \Delta) - p_{ij}(s, t)}{\Delta} = \\ &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \left[\sum_{k=1}^N p_{ik}(s, t) p_{kj}(t, t + \Delta) - p_{ij}(s, t) \right] = \\ &= \lim_{\Delta \rightarrow 0} \left[\sum_{k=1, k \neq j}^N p_{ik}(s, t) \frac{p_{kj}(t, t + \Delta)}{\Delta} + p_{ij}(s, t) \frac{p_{jj}(t, t + \Delta) - 1}{\Delta} \right]. \end{aligned} \quad (13.35)$$

Рассмотрим выражение в последней квадратной скобке. В силу (13.28) и дифференцируемости $p_{ij}(s, t)$ по t

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \frac{p_{jj}(t, t + \Delta) - 1}{\Delta} &= \left. \frac{\partial p_{jj}(t, u)}{\partial u} \right|_{u=t} = A_{jj}(t) \\ \lim_{\Delta \rightarrow 0} \frac{p_{kj}(t, t + \Delta)}{\Delta} &= \left. \frac{\partial p_{kj}(t, u)}{\partial u} \right|_{u=t} = A_{kj}(t), \quad k \neq j. \end{aligned} \quad (13.36)$$

Подставляя эти выражения в (13.35), получим систему уравнений (13.32). Сопоставляя (13.36) и (13.28), заключаем ввиду неравенств $0 \leq p_{ij}(t, u) \leq 1$, что $A_{ik} \geq 0$, $i \neq k$, $A_{kk} \leq 0$, $\sum_{k=1}^N A_{ik}(t) = 0$, и (13.34) также доказано. Далее, при $t > s$ в силу (13.31)

$$\begin{aligned} \frac{\partial p_{ij}(s, t)}{\partial s} &= \lim_{\Delta \rightarrow 0} \frac{p_{ij}(s + \Delta, t) - p_{ij}(s, t)}{\Delta} = \\ &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \left[p_{ij}(s + \Delta, t) - \sum_{k=1}^N p_{ik}(s, s + \Delta) p_{kj}(s + \Delta, t) \right] = \\ &= - \lim_{\Delta \rightarrow 0} \left[\frac{p_{ii}(s, s + \Delta) - 1}{\Delta} p_{ij}(s + \Delta, t) + \right. \\ &\quad \left. + \sum_{k=1, k \neq i}^N \frac{p_{ik}(s, s + \Delta)}{\Delta} p_{kj}(s + \Delta, t) \right]. \end{aligned} \quad (13.37)$$

Отсюда, рассуждая как и выше, получим систему дифференциальных уравнений (13.33). ■

Уравнения (13.32) называются *прямой*, а уравнения (13.33) *обратной системой уравнений А.Н. Колмогорова*.

Физический смысл функций $A_{ij}(\cdot)$ состоит в том, что $A_{ij}(t)dt$ есть вероятность перехода из состояния ω_i в состояние ω_j за время от t до $t + dt$. Если функции $A_{ij}(\cdot)$, $i, j = 1, \dots, N$, непрерывны, то функции $p_{ij}(s, t)$, $s, t \in [t_0, \infty)$, $i, j = 1, \dots, N$, представляют единственное решение системы уравнений (13.32), удовлетворяющее условиям $p_{ii}(s, s) = 1$, $p_{ij}(s, s) = 0$, $i \neq j$, $i, j = 1, \dots, N$. Таким образом, рассматриваемый марковский процесс полностью определяется заданием функций $A_{ij}(\cdot)$, $i, j = 1, \dots, N$. Нетрудно показать также, что если заданы любые непрерывные функции $A_{ij}(t)$, $t \in [t_0, \infty)$, $i, j = 1, \dots, N$,

удовлетворяющие условиям $A_{ij}(t) \geq 0$, $i \neq j$, $A_{jj}(t) \leq 0$, $\sum_{j=1}^N A_{ij}(t) = 0$, то решение $p_{ij}(s, t)$, $s \leq t$, $i, j = 1, \dots, N$, системы (13.32) при начальных условиях $p_{ii}(s, s) = 1$, $p_{ij}(s, s) = 0$, $i \neq j$, $i, j = 1, \dots, N$, будет неотрицательным ($p_{ij}(s, t) \geq 0$), справедливо равенство (13.31), так что переходные вероятности $p_{ij}(s, t)$, $s \leq t$, $i, j = 1, \dots, N$, будут определять некоторый марковский процесс.

Отправляясь от прямой системы уравнений Колмогорова, установим дифференциальные уравнения для абсолютных вероятностей $p_i(t)$. Если a_i — начальное распределение вероятностей (13.29), то

$$p_i(t) = \sum_{j=1}^N a_j p_{ji}(t_0, t), \quad t \neq t_0, \quad i = 1, \dots, N. \quad (13.38)$$

Дифференцируя (13.38) по t и используя сначала (13.32), а затем (13.38), найдем

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^N a_j \sum_{k=1}^N p_{jk}(t_0, t) A_{kj}(t) = \sum_{k=1}^N A_{kj}(t) p_k(t), \quad t > t_0, \quad i = 1, \dots, N. \quad (13.39)$$

Таким образом, получаем систему уравнений

$$\frac{dp_i(t)}{dt} = \sum_{k=1}^N A_{kj}(t) p_k(t), \quad t > t_0, \quad i = 1, \dots, N. \quad (13.40)$$

Начальные условия для этой системы таковы: $p_i(t_0) = a_i$, $i = 1, \dots, N$.

Для однородного марковского процесса переходные вероятности $p_{ij}(s, t)$, $i, j = 1, \dots, N$, зависят лишь от разности $t - s$. В этом случае согласно определению (13.34) A_{ij} , $i, j = 1, \dots, N$, — константы, и система уравнений (13.32) принимает вид

$$\frac{dp_{ij}(t)}{dt} = \sum_{k=1}^N p_{ik}(t) A_{kj}, \quad t > t_0, \quad i, j = 1, \dots, N. \quad (13.41)$$

Замечание 1.13.2. В случае счетного числа состояний $\Omega = (\omega_1, \omega_2, \dots)$ прямая и обратная системы уравнений Колмогорова (13.32) и (13.34) остаются справедливыми, но для их обоснования надо дополнительно требовать равномерную сходимость соответствующих рядов

Пример 1.13.6. *Двусторонняя реакция.* Система может находиться в двух состояниях ω_1 и ω_2 (ω_1 — нераспавшаяся частица, ω_2 — распавшаяся). Пусть возможен как процесс распада — переход из ω_1 в ω_2 с вероятностью $\alpha \Delta t + o(\Delta t)$ за время Δt , так и процесс восстановления — переход из ω_2 в ω_1 с вероятностью $\beta \Delta t + o(\Delta t)$ за

время Δt . В этом случае $A_{12} = \alpha$, $A_{21} = \beta$, а, стало быть, $A_{11} = -\alpha$, $A_{22} = -\beta$. Уравнения (13.40) дают

$$\frac{dp_1(t)}{dt} = -\alpha p_1(t) + \beta p_2(t), \quad \frac{dp_2(t)}{dt} = \alpha p_1(t) - \beta p_2(t), \quad t \in (0, \infty). \quad (13.42)$$

Пусть задано начальное распределение вероятностей при $t = 0$

$$a_1 = 1, \quad a_2 = 0. \quad (13.43)$$

Подставляя в первое уравнение (13.42) $p_2(t) = 1 - p_1(t)$, найдем $\frac{dp_1(t)}{dt} = -(\alpha + \beta)p_1(t) + \beta$, откуда получим решение

$$p_1(t) = e^{-(\alpha+\beta)t} + \frac{\beta}{\alpha+\beta}(1 - e^{-(\alpha+\beta)t}), \quad p_2(t) = \frac{\alpha}{\alpha+\beta}(1 - e^{-(\alpha+\beta)t}). \quad (13.44)$$

При $t \rightarrow \infty$ $p_1(t) \rightarrow \beta/(\alpha + \beta)$, $p_2(t) \rightarrow \alpha/(\alpha + \beta)$, т. е. процесс эргодичен. Если восстановление невозможно (например, радиоактивный распад), то $\beta = 0$ и

$$p_1(t) = e^{-\alpha t}, \quad p_2(t) = 1 - e^{-\alpha t}.$$

Пример 1.13.7. *Двухпозиционное реле.* Пусть двухпозиционное реле находится под воздействием случайной последовательности управляющих импульсов, имеющих с одинаковой вероятностью знаки плюс или минус, причем положительный импульс создает или сохраняет состояние ω_1 а отрицательный — состояние ω_2 . Тогда $A_{12} = A_{21} = \alpha$, и, следовательно, согласно (13.34) $A_{11} = A_{22} = -\alpha$. Это частный случай предыдущего примера: $\beta = \alpha$. Найдем переходные вероятности. Решая систему уравнений (13.41) с начальными условиями $p_{ik}(t_0) = \delta_{ik}$, $i, k = 1, 2$, получим

$$\begin{aligned} p_{12}(t) &= p_{21} = \frac{1}{2}(1 - e^{-2\alpha(t-t_0)}), \\ p_{11}(t) &= p_{22} = \frac{1}{2}(1 + e^{-2\alpha(t-t_0)}), \quad t \in [t_0, \infty). \end{aligned} \quad (13.45)$$

При $t \rightarrow \infty$ все вероятности перехода стремятся к значению $1/2$.

Пример 1.13.8. *Пуассоновский поток требований.* Пусть на некоторую систему обслуживания поступают требования так, что $\xi(t)$ — число требований за время t — образует однородный марковский процесс со счетным числом состояний $\omega_i = 0, 1, 2, \dots$, $t \in [0, \infty)$. Из состояния i система непосредственно может перейти только в состояние $i + 1$, $i = 0, 1, 2, \dots$. Таким образом, $A_{i, i+1} = \alpha$, остальные $A_{ij} = 0$ при $i \neq j$ и согласно (34) $A_{ii} = -\alpha$, $i, j = 0, 1, 2, \dots$. Абсолютные вероятности $p_j(t)$, т. е. вероятности того, что за время t поступит j требований, $j = 1, 2, \dots$, суть решения бесконечной системы уравнений (13.40)

$$\frac{dp_i(t)}{dt} = \sum_{k=0}^{\infty} A_{ki} p_k(t), \quad t > 0, \quad i = 1, 2, \dots, \quad (13.46)$$

с начальными условиями

$$p_0(0) = 1, \quad p_i(0) = 0, \quad i = 1, 2, \dots \quad (13.47)$$

Перепишем систему (13.46) подробнее

$$\frac{dp_0(t)}{dt} = -\alpha p_0(t), \quad \frac{dp_i(t)}{dt} = \alpha p_{i-1}(t) - \alpha p_i(t), \quad t > 0, \quad i = 1, 2, \dots \quad (13.48)$$

Система (13.48) легко решается с помощью введения новых функций $q_k(\cdot)$ по формуле $p_k(t) = e^{-\alpha t} q_k(t)$, $t \geq 0$, $k = 0, 1, 2, \dots$, для которых получаем систему уравнений

$$\begin{aligned} \frac{dq_0(t)}{dt} &= 0, \\ \frac{dq_i(t)}{dt} &= \alpha q_{i-1}(t), \quad t > 0, \quad i = 1, 2, \dots, \end{aligned} \quad (13.49)$$

с начальными условиями

$$q_0(0) = 1, \quad q_i(0) = 0, \quad i = 1, 2, \dots$$

Интегрируя последовательно уравнения (13.49) при $i = 0, 1, 2, \dots$, получим

$$\begin{aligned} q_0(t) &= 1, \quad q_1(t) = \alpha t, \dots, \quad q_k(t) = \frac{(\alpha t)^k}{k!}, \dots, \\ p_k(t) &= \frac{(\alpha t)^k}{k!} e^{-\alpha t}, \quad k = 0, 1, 2, \dots, \end{aligned} \quad (13.50)$$

т. е. рассмотренный поток требований является пуассоновским.

Пример 1.13.9. Пусть однородный марковский процесс обладает свойством эргодичности: $p_{ij}(t) \rightarrow p_j$, $i, j = 1, 2, \dots$, при $t \rightarrow \infty$. Тогда, как видно из системы (13.41), стационарные вероятности p_j , $j = 1, 2, \dots$, удовлетворяют системе уравнений

$$\sum_{i=1}^{\infty} p_i A_{ik} = 0, \quad k = 1, 2, \dots,$$

с условием нормировки $\sum_{i=1}^{\infty} p_i = 1$. Пусть, например, $A_{i,i+1} = A$, $A_{i+1,i} = B$, $B > A$, $A_{11} = -A$, $A_{ii} = -(A+B)$, $i > 1$, $A_{ij} = 0$ для других i и $j = 1, 2, \dots$. Тогда указанная выше система уравнений имеет вид

$$\begin{aligned} Bp_2 - Ap_1 &= 1, \quad k = 1, \\ Bp_{k+1} - Ap_k &= Bp_k - Ap_{k-1}, \quad k = 2, 3, \dots \end{aligned}$$

Отсюда $Bp_k = Ap_{k-1}$, $k = 2, 3, \dots$, т. е. $p_k = (A/B)^{k-1} p_1$ и условие нормировки $\sum_{i=1}^{\infty} p_i = 1$ дает $p_1 = 1 - A/B$. Таким образом, стационарные вероятности в этом случае равны

$$p_k = \left(1 - \frac{A}{B}\right) \left(\frac{A}{B}\right)^{k-1}, \quad k = 1, 2, \dots$$

1.13.3.2 Процессы с непрерывным множеством состояний. Наиболее интересными и важными для физики являются процессы, n -мерная функция распределения которых при любом $n = 1, 2, \dots$ имеет плотность распределения вероятностей. Точнее: пусть $\xi(t)$, $t \in T$, — случайный процесс и пусть при каждом множестве моментов времени $t_1, t_2, \dots, t_n \in T$ n -мерная случайная величина $(\xi(t_1), \dots, \xi(t_n))$ имеет n -мерную плотность вероятности $p_n(t_1, x_1; \dots; t_n, x_n)$, $x_1, \dots, x_n \in \mathcal{R}^1$, $t_1, \dots, t_n \in T$. Эта плотность обладает двумя очевидными свойствами:

1) $p_n(t_1, x_1; \dots; t_n, x_n)$ симметрична относительно любых перестановок пар аргументов (t_i, x_i) , ибо $p_n(t_1, x_1; \dots; t_n, x_n) dx_1 \dots dx_n$ выражает вероятность совместного осуществления событий $x_i \leq \xi(t_i) < x_i + dx_i$, $i = 1, \dots, n$, и, стало быть, не зависит от порядка их перечисления;

2) все конечномерные плотности p_n для различных n должны быть согласованы в том смысле, что плотность любого k -мерного распределения при $k < n$ определяется с помощью n -мерного распределения:

$$p_k(t_1, x_1; \dots; t_k, x_k) = \int \dots \int_{-\infty}^{\infty} p_n(t_1, x_1; \dots; t_k, x_k; t_{k+1}, x_{k+1}; \dots; t_n, x_n) dx_{k+1} \dots dx_n, \quad x_1, \dots, x_k \in \mathcal{R}^1, \quad t_1, \dots, t_k \in T. \quad (13.51)$$

Согласно определению плотности условной вероятности (см. §1.8) для всех $x_1, \dots, x_k \in \mathcal{R}^1$, $t_1, \dots, t_k \in T$ выполнены равенства

$$p_n(t_1, x_1; \dots; t_n, x_n) = p_{n-1}(t_1, x_1; \dots; t_{n-1}, x_{n-1}) q_n(t_n, x_n | t_1, x_1; \dots; t_{n-1}, x_{n-1}). \quad (13.52)$$

Свойство марковости процесса означает, как уже известно, что его вероятностные свойства в момент t_n определяются состоянием в момент t_{n-1} и не зависят от протекания процесса в предшествующие моменты времени, в силу чего в (13.53)

$$q_n(t_n, x_n | t_1, x_1; \dots; t_{n-1}, x_{n-1}) = q(t_n, x_n | t_{n-1}, x_{n-1}). \quad (13.53)$$

Функцию $q(t, x | \tau, y)$, $t > \tau$, $x, y \in \mathcal{R}^1$, называют ¹⁾ *плотностью переходной вероятности*. Сопоставив (13.52) и (13.53), найдем

$$p_n(t_1, x_1; \dots; t_n, x_n) = p_{n-1}(t_1, x_1; \dots; t_{n-1}, x_{n-1}) q(t_n, x_n | t_{n-1}, x_{n-1}), \quad (13.54)$$

¹⁾ В §1.13.3.2 плотность переходной вероятности обозначена как плотность условной вероятности, ср. с обозначением (12.6).

а применив эту формулу последовательно для $n, n-1, \dots, 2$, получим

$$\begin{aligned} p_n(t_1, x_1; \dots; t_n, x_n) = \\ = p_1(t_1, x_1)q(t_2, x_2|t_1, x_1) \dots q(t_n, x_n|t_{n-1}, x_{n-1}), \end{aligned} \quad (13.55)$$

$t_1, \dots, t_n \in T, x_1, \dots, x_n \in \mathcal{R}^1.$

Равенство (13.55) означает, что для задания n -мерной плотности вероятности марковского процесса достаточно знать лишь две функции: одномерную плотность $p_1(t, x)$ и плотность переходной вероятности $q(t, x|\tau, y)$, $\tau \leq t, x, y \in \mathcal{R}^1$.

Уравнение Смолуховского (Колмогорова–Чепмена). Основным в теории непрерывных марковских процессов является *уравнение Э. Смолуховского* (оно также называется *уравнением Колмогорова–Чепмена*):

$$q(t, x|t_0, x_0) = \int_{-\infty}^{\infty} q(t, x|\tau, y)q(\tau, y|t_0, x_0)dy \quad (13.56)$$

для любых трех моментов времени $t_0 < \tau < t$, $t_0, \tau, t \in T$, и $x_0, x \in \mathcal{R}^1$. В самом деле, согласно (13.55)

$$p_3(t_0, x_0; \tau, y; t, x) = p_1(t_0, x_0)q(\tau, y|t_0, x_0)q(t, x|\tau, y), \quad t_0 < \tau < t, \quad (13.57)$$

и

$$p_2(t_0, x_0; t, x) = p_1(t_0, x_0)q(t, x|t_0, x_0), \quad t_0 < t. \quad (13.58)$$

В силу условия согласования (13.51)

$$p_2(t_0, x_0; t, x) = \int_{-\infty}^{\infty} p_3(t_0, x_0; \tau, y; t, x)dy, \quad \tau \in (t_0, t). \quad (13.59)$$

Подставив (13.57) и (13.58) в (13.59) и сократив обе части равенства на $p_1(t_0, x_0)$, получим (13.56).

Для *однородного* марковского процесса плотность вероятности перехода зависит только от разности моментов времени $q(t, x|\tau, y) = q(x|t - \tau, y)$. В этом случае уравнение Смолуховского (13.56) имеет вид

$$q(x|t - t_0, x_0) = \int_{-\infty}^{\infty} q(x|t - \tau, y)q(y|\tau - t_0, x_0)dy, \quad t_0 < \tau < t, \quad x, x_0 \in \mathcal{R}^1. \quad (13.60)$$

Уравнение Эйнштейна–Фоккера–Планка. В дальнейшем при формулировке условий и интерпретации результатов нам будет удобно использовать терминологию, связанную с блужданием частицы под действием случайных толчков. Эта интерпретация тем более уместна, что мы ограничиваемся рассмотрением так называемых *диффузионных*

процессов, которые характеризуются выполнением следующих трех условий:

$$1) \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y)q(t + \Delta, x|t, y)dx = A(y, t), \quad y \in \mathcal{R}^1, \quad t \in T; \quad (13.61)$$

интеграл в этом выражении означает условное среднее значение перемещения частицы за время Δ из фиксированной точки y , так что $A(y, t)$ — средняя скорость изменения состояния в точке y в момент времени t , $A(y, t)$ называется *коэффициентом сноса*.

$$2) \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y)^2 q(t + \Delta, x|t, y)dx = B(y, t) > 0, \quad y \in \mathcal{R}^1, \quad t \in T; \quad (13.62)$$

здесь интеграл определяет меру разброса конечных точек x относительно исходной точки $y \in \mathcal{R}^1$. Условие 2) означает, что этот разброс растет при малых Δ пропорционально Δ , т. е. по диффузионному закону. Коэффициент $B(y, t)/2$ называется *коэффициентом диффузии*.

$$3) \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} |x - y|^3 q(t + \Delta, x|t, y)dx = 0, \quad y \in \mathcal{R}^1, \quad t \in T. \quad (13.63)$$

Это условие означает, что в малых промежутках времени вероятность больших значений $|x - y|$ мала (в самом деле, при малых Δ интеграл в (13.62) больше интеграла в (13.63), т. е. основную роль в интегралах играют малые значения $|x - y|$). С физической точки зрения именно это условие позволяет рассматривать $\xi(t)$, $t \in T$, как непрерывно (во времени) меняющуюся величину, т. е. как среднее за время, гораздо большее промежутка между последовательными случайными толчками.

Теорема 1.13.5. Пусть выполнены условия (13.61)–(13.63) и $q(t, x|t_0, x_0) \in C^{(2,1)}(\mathcal{R}^1 \times T)$ как функция $x \in \mathcal{R}^1$ и $t \in T$, $A(x, t) \in C^{(1,0)}(\mathcal{R}^1 \times T)$, $B(x, t) \in C^{(2,0)}(\mathcal{R}^1 \times T)$. Тогда плотность переходной вероятности

$q(t, x|t_0, x_0)$, как функция $x \in \mathcal{R}^1$ и $t \in T$, удовлетворяет уравнению

$$\begin{aligned} \frac{\partial q(t, x|t_0, x_0)}{\partial t} = & -\frac{\partial}{\partial x}(A(x, t)q(t, x|t_0, x_0)) + \\ & + \frac{1}{2} \frac{\partial^2}{\partial x^2}(B(x, t)q(t, x|t_0, x_0)). \end{aligned} \quad (13.64)$$

Доказательство. Умножив уравнение Смолуховского (13.56) на произвольную финитную функцию $\varphi(\cdot) \in C_0^\infty(\mathcal{R}^1)$, проинтегрируем его по x от $-\infty$ до ∞ и, предварительно изменив порядок интегрирования

(что возможно в силу финитности $\varphi(\cdot)$), представим $\varphi(x)$ в виде разложения по формуле Тейлора:

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \varphi(x) q(t + \Delta, x | t_0, x_0) dx = \\
 &= \int_{-\infty}^{\infty} \varphi(x) dx \int_{-\infty}^{\infty} q(t + \Delta, x | t, y) q(t, y | t_0, x_0) dy = \\
 &= \int_{-\infty}^{\infty} q(t, y | t_0, x_0) dy \int_{-\infty}^{\infty} q(t + \Delta, x | t, y) \times \\
 & \times \left\{ \varphi(y) + \varphi'(y)(x - y) + \frac{\varphi''(y)}{2}(x - y)^2 + \varphi'''(z) \frac{(x - y)^3}{6} \right\} dx, \\
 & \quad z \in (y, x), \quad \Delta > 0.
 \end{aligned} \tag{13.65}$$

Поскольку $\int_{-\infty}^{\infty} q(t + \Delta, x | t, y) dx = 1$, то после деления на Δ преобразуем (13.65) к виду

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \varphi(x) \frac{q(t + \Delta, x | t_0, x_0) - q(t, y | t_0, x_0)}{\Delta} dx = \\
 &= \int_{-\infty}^{\infty} q(t, y | t_0, x_0) dy \left\{ \varphi'(y) \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y) q(t + \Delta, x | t, y) dx + \right. \\
 & \quad + \frac{1}{2} \varphi''(y) \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y)^2 q(t + \Delta, x | t, y) dx + \\
 & \quad \left. + \frac{1}{6} \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y)^3 \varphi'''(z) q(t + \Delta, x | t, y) dx \right\}.
 \end{aligned} \tag{13.66}$$

В силу финитности $\varphi(\cdot)$ и непрерывности подынтегральных функций можно в равенстве (13.66) перейти к пределу при $\Delta \rightarrow 0$ под знаком интеграла. Получим с учетом условий (13.61)–(13.63)

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \varphi(x) \frac{\partial q(t, y | t_0, x_0)}{\partial t} dx = \\
 &= \int_{-\infty}^{\infty} q(t, y | t_0, x_0) \left\{ \varphi'(y) A(y, t) + \frac{1}{2} \varphi''(y) B(y, t) \right\} dy.
 \end{aligned} \tag{13.67}$$

Последний член в (13.66) пропадает ввиду (13.63):

$$\begin{aligned} & \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \left| \int_{-\infty}^{\infty} (x-y)^3 q(t+\Delta, x|t, y) \varphi'''(z) dx \right| \leq \\ & \leq c \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} |x-y|^3 q(t+\Delta, x|t, y) dx = 0, \quad c = \sup_{z \in \mathcal{R}^1} |\varphi'''(z)| < \infty. \end{aligned}$$

Правую часть в (13.67) проинтегрируем по частям, получим с учетом финитности $\varphi(x)$:

$$\begin{aligned} & \int_{-\infty}^{\infty} q(t, y|t_0, x_0) \left\{ \varphi'(y) A(y, t) + \frac{1}{2} \varphi''(y) B(y, t) \right\} dy = \\ & = - \int_{-\infty}^{\infty} \varphi(y) \left[\frac{\partial}{\partial y} (qA) - \frac{1}{2} \frac{\partial^2}{\partial y^2} (qB) \right] dy. \end{aligned} \quad (13.68)$$

Подставляя (68) в (67), найдем

$$\begin{aligned} & \int_{-\infty}^{\infty} \varphi(x) \left\{ \frac{\partial q(t, y|t_0, x_0)}{\partial t} + \frac{\partial}{\partial x} (q(t, x|t_0, x_0) A(x, t)) - \right. \\ & \left. - \frac{1}{2} \frac{\partial^2}{\partial x^2} (q(t, y|t_0, x_0) B(x, t)) \right\} dx = 0. \end{aligned} \quad (13.69)$$

Поскольку здесь $\varphi(x)$, $x \in \mathcal{R}^1$, —любая финитная функция, а выражение в фигурных скобках в (13.69) является непрерывной функцией $x \in \mathcal{R}^1$, то из (13.69) следует справедливость равенства (13.64). ■

Уравнение (13.64) называется *уравнением Эйнштейна–Фоккера–Планка* или *прямым (первым) уравнением Колмогорова*. Это — уравнение параболического типа. По самому смыслу плотности переходной вероятности следует искать решения уравнения (13.64), удовлетворяющие условиям

$$q(t, x|t_0, x_0) \geq 0, \quad \int_{-\infty}^{\infty} q(t, x|t_0, x_0) dx = 1, \quad (13.70)$$

и начальному условию (см. ниже формулу (13.77))

$$q(t, x|t_0, x_0)|_{t=t_0} = \delta(x - x_0). \quad (13.71)$$

Получим теперь *обратное (второе) уравнение Колмогорова*.

Теорема 1.13.6. Пусть выполнены условия (13.61)–(13.63) и $q(t, x|t_0, x_0) \in C^{(3,1)}(\mathcal{R}^1 \times T)$ как функция $x_0 \in \mathcal{R}^1$ и $t_0 \in T$, причем

$\sup_{x_0 \in \mathcal{R}^1} \left| \frac{\partial^3 q}{\partial x_0^3} \right| < \infty$, $t \neq t_0$. Тогда переходная плотность вероятности как функция x_0 и t_0 удовлетворяет уравнению

$$\frac{\partial q(t, x|t_0, x_0)}{\partial t_0} = -A(x_0, t_0) \frac{\partial q(t, x|t_0, x_0)}{\partial x_0} - \frac{B(x_0, t_0)}{2} \frac{\partial^2 q(t, x|t_0, x_0)}{\partial x_0^2}. \quad (13.72)$$

Доказательство. Взяв $\Delta > 0$ так, что $t_0 + \Delta < t$, разложим по формуле Тейлора переходную плотность вероятности под знаком интеграла в уравнении Смолуховского (13.56):

$$\begin{aligned} q(t, x|t_0, x_0) &= \int_{-\infty}^{\infty} q(t, x|t_0 + \Delta, y) q(t_0 + \Delta, y|t_0, x_0) dy = \\ &= \int_{-\infty}^{\infty} q(t_0 + \Delta, y|t_0, x_0) \{ q(t, x|t_0 + \Delta, x_0) + \\ &+ \frac{\partial q(t, x|t_0 + \Delta, x_0)}{\partial x_0} (y - x_0) + \frac{1}{2} \frac{\partial^2 q(t, x|t_0 + \Delta, x_0)}{\partial x_0^2} (y - x_0)^2 + \\ &+ \frac{1}{6} \frac{\partial^3 q(t, x|t_0 + \Delta, z)}{\partial x_0^3} (y - x_0)^3 \} dy, \quad z \in (y, x_0). \end{aligned} \quad (13.73)$$

Поскольку

$$\int_{-\infty}^{\infty} q(t_0 + \Delta, y|t_0, x_0) dy = 1,$$

то на основании соотношения (13.73) получим

$$\begin{aligned} &\frac{q(t, x|t_0, x_0) - q(t, x|t_0 + \Delta, x_0)}{\Delta} = \\ &= \frac{\partial q(t, x|t_0 + \Delta, x_0)}{\partial x_0} \frac{1}{\Delta} \int_{-\infty}^{\infty} q(t_0 + \Delta, y|t_0, x_0) (y - x_0) dy + \\ &+ \frac{1}{2} \frac{\partial^2 q(t, x|t_0 + \Delta, x_0)}{\partial x_0^2} \frac{1}{\Delta} \int_{-\infty}^{\infty} q(t_0 + \Delta, y|t_0, x_0) (y - x_0)^2 dy + \\ &+ \frac{1}{6} \frac{1}{\Delta} \int_{-\infty}^{\infty} \frac{\partial^3 q(t, x|t_0 + \Delta, z)}{\partial x_0^3} q(t_0 + \Delta, y|t_0, x_0) (y - x_0)^3 dy. \end{aligned} \quad (13.74)$$

Используя условие $\sup_{z \in \mathcal{R}^1} \left| \frac{\partial^3 q(t, x|t_0 + \Delta, z)}{\partial x_0^3} \right| \leq c$ для всех $\Delta > 0$, таких, что $t_0 + \Delta < t$, и условия (13.61)–(13.63), перейдем в (13.74) к пределу при $\Delta \rightarrow 0$, получим уравнение (13.72)

$$-\frac{\partial q(t, x|t_0, x_0)}{\partial t_0} = A(x_0, t_0) \frac{\partial q(t, x|t_0, x_0)}{\partial x_0} + \frac{B(x_0, t_0)}{2} \frac{\partial^2 q(t, x|t_0, x_0)}{\partial x_0^2}, \quad x \in \mathcal{R}^1, \quad t \in T. \quad \blacksquare \quad (13.75)$$

Уравнение (13.72) следует решать в обратную сторону по времени, для $t_0 \leq t$, при этом решение $q(t, x|t_0, x_0)$ снова должно быть подчинено условиям (13.70) и (13.71).

Пусть в начальный момент t_0 задана плотность распределения вероятности $p(t_0, x)$, $x \in \mathcal{R}^1$, случайной величины $\xi(t_0)$. Тогда двумерная плотность вероятности для произвольного момента времени $t \geq t_0$ и начального момента t_0 равна

$$p_2(t, x; t_0, x_0) = p(t_0, x_0)q(t, x|t_0, x_0), \quad x \in \mathcal{R}^1, \quad (13.76)$$

а одномерная плотность случайного процесса $\xi(t)$ для момента времени t равна согласно (13.51)

$$p_1(t, x) = \int_{-\infty}^{\infty} p(t_0, x_0)q(t, x|t_0, x_0)dx_0, \quad t \geq t_0, \quad x \in \mathcal{R}^1. \quad (13.77)$$

Если мы теперь уравнение (13.64) Эйнштейна–Фоккера–Планка умножим на $p(t_0, x_0)$ и проинтегрируем по x_0 , то в силу (13.77), очевидно, найдем, что плотность $p_1(t, x)$ удовлетворяет тому же уравнению

$$\frac{\partial p_1(t, x)}{\partial t} = -\frac{\partial}{\partial x}(A(x, t)p_1(t, x) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(B(x, t)p_1(t, x))), \quad x \in \mathcal{R}^1, \quad t \in T. \quad (13.78)$$

Решение уравнения (13.78) должно удовлетворять условиям

$$p_1(t, x) \geq 0, \quad \int_{-\infty}^{\infty} p_1(t, x)dx = 1, \quad p_1(t, x)|_{t=t_0} = p(t_0, x), \quad x \in \mathcal{R}^1, \quad t \in T. \quad (13.79)$$

Пример 1.13.10. Пусть марковский процесс *однороден по времени*, т. е. $q(t, x|t_0, x_0) = q(x|t - t_0, x_0)$. В этом случае в условиях (13.61), (13.62) функции A и B не зависят от t . Пусть, кроме того, одномерная плотность p_1 также не зависит от времени (стационарный марковский процесс). Тогда уравнение (13.78) записывается в виде

$$\frac{d}{dx} \left[A(x)p_1(x) - \frac{1}{2} \frac{d}{dx}(B(x)p_1(x)) \right] = 0. \quad (13.80)$$

Если на границах изменения x (т. е. области значений процесса $\xi(\cdot)$) поток $A p_1 - \frac{1}{2} \frac{d}{dx}(B p_1)$ равен нулю, то в силу (13.80) он равен нулю всюду:

$$A(x)p_1(x) - \frac{1}{2} \frac{d}{dx}(B(x)p_1(x)) = 0. \quad (13.81)$$

Интегрируя уравнение (13.81) (для этого полагаем $v = Bp_1$), получим

$$p_1(x) = \frac{c}{B(x)} \exp \left(\int_0^x (2A(y)/B(y)) dy \right), \quad (13.82)$$

где c — постоянная, определяемая из условия нормировки (13.79). Физическим примером, в котором существует стационарное распределение $p_1(\cdot)$, является броуновское движение частицы над отражающей границей при наличии силы тяжести. Здесь $A = -mg$. Ясно, что на отражающей границе выполнено условие обращения потока в нуль, так что уравнение (13.81) имеет место. Выражение (13.82) с постоянными A и B дает

$$p_1(x) = ce^{-2mgx/B}, \quad x \geq 0, \quad (13.83)$$

т. е. мы получили барометрическую формулу. Известно, что $B = 2kT$, k — постоянная Больцмана, T — абсолютная температура. Итак,

$$p_1(x) = ce^{-mgx/(kT)}, \quad x \geq 0. \quad (13.84)$$

Пример 1.13.11. Рассмотрим марковский процесс, *однородный по координате*: $q(t, x|t_0, x_0) = q(t, x - x_0|t_0)$. В этом случае в условиях (13.61) и (13.62) функции A и B зависят лишь от t и уравнение (13.64) имеет вид

$$\frac{\partial q(t, x - x_0|t_0)}{\partial t} = -A(t) \frac{\partial q}{\partial x} + \frac{B(t)}{2} \frac{\partial^2 q}{\partial x^2}. \quad (13.85)$$

С помощью замены независимых переменных $(x, t) \rightarrow (y, \tau)$ по формулам

$$y = x - x_0 - \int_{t_0}^t A(s) ds, \quad \tau = \int_{t_0}^t B(s) ds, \quad (13.86)$$

приходим к уравнению теплопроводности

$$\frac{\partial q}{\partial \tau} = \frac{1}{2} \frac{\partial^2 q}{\partial y^2}. \quad (13.87)$$

Его решение, удовлетворяющее условиям (70) и (71), имеет вид $q(\tau, y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{y^2}{2\tau}}$, или в старых переменных

$$\begin{aligned} q(t, x - x_0|t_0) &= \\ &= \left(2\pi \int_{t_0}^t B(s) ds \right)^{-1/2} \exp \left\{ -\frac{1}{2} \left(x - x_0 - \int_{t_0}^t A(s) ds \right)^2 / \int_{t_0}^t B(s) ds \right\}. \end{aligned} \quad (13.88)$$

Пусть, в частности, коэффициент сноса $A(t) = 0$, а $B(t) = B = \text{const}$. Тогда

$$q(t, x - x_0|t_0) = \frac{1}{(2\pi B(t - t_0))^{1/2}} e^{-\frac{(x - x_0)^2}{2B(t - t_0)}}, \quad t > t_0, \quad x \in \mathcal{R}^1.$$

Отсюда среднее смещение частицы равно нулю, а средний квадрат смещения частицы равен $B(t - t_0)$, т. е. растет пропорционально времени. Этот результат впервые был получен А. Эйнштейном.

1.13.4. Стационарные процессы. В математических моделях, описывающих ряд явлений в радиофизике, в теории квантовых генераторов, в астрономии и геофизике встречаются случайные функции, при изучении которых нельзя пренебрегать эффектами последствия. Наиболее важными среди них являются так называемые стационарные процессы. Для их описания наиболее естественно рассматривать комплексные случайные величины. Напомним соответствующее определение: комплексная случайная величина $\xi = \eta + i\zeta$ есть пара (η, ζ) действительных случайных величин, относительно их совместного распределения не делается никаких предположений, $\bar{\xi} = \eta - i\zeta$, произведение $\xi\bar{\xi} = |\xi|^2$ играет ту же роль, что и ξ^2 в действительном случае, так что, например, $D\xi = M[(\xi - M\xi)(\bar{\xi} - M\bar{\xi})]$.

Определение 1.13.8. Случайный процесс $\xi(t)$, $-\infty < t < \infty$, называется *стационарным*, если математическое ожидание $M\xi(t)$ и дисперсия $D\xi(t)$ не зависят от t , а корреляционная функция случайных величин $\xi(t+u)$ и $\xi(t)$ зависит лишь от u , т. е.

$$M\xi(t) = M\xi(0) = a, \quad (13.89)$$

$$D\xi(t) = D\xi(0) = \sigma^2, \quad (13.90)$$

$$R[\xi(t+u), \xi(t)] \equiv M[(\xi(t+u) - M\xi(t+u))(\bar{\xi}(t) - M\bar{\xi}(t))] = R(u), \quad t, u \in (-\infty, \infty). \quad (13.91)$$

Не ограничивая общности, будем считать, что $a = 0$, $\sigma^2 = 1$ (достаточно от $\xi(\cdot)$ перейти к центрированной и нормированной случайной функции $[\xi(\cdot) - a]/\sigma$). Тогда

$$R(u) = M\{\xi(t+u)\bar{\xi}(t)\} = M\{\xi(u)\bar{\xi}(0)\}, \quad (13.92)$$

$$R(0) = M\{\xi(0)\bar{\xi}(0)\} = D\xi(0) = 1. \quad (13.93)$$

Из определения (13.91) следует, что если $\xi(t)$, $-\infty < t < \infty$, — вещественный процесс, то $R(u)$, $-\infty < u < \infty$, — четная функция.

Излагаемая ниже теория стационарных процессов называется корреляционной теорией ввиду того, что она формулируется в терминах свойств корреляционной функции $R(u)$, $u \in (-\infty, \infty)$, §1.13.2.

Напомним, что, случайный процесс называется *непрерывным в среднем квадратичном* (с. к.), если для каждого t , $-\infty < t < \infty$, см. определение 1.13.4,

$$M|\xi(t+h) - \xi(t)|^2 \rightarrow 0 \text{ при } h \rightarrow 0. \quad (13.94)$$

В дальнейшем будем рассматривать только с. к. непрерывные стационарные процессы. Для с. к. непрерывного стационарного процесса кор-

реляционная функция $R(\cdot)$ является непрерывной функцией. В самом деле, используя неравенство Коши-Буняковского, получим

$$\begin{aligned} |R(u + \Delta u) - R(u)| &= |M\{\xi(u + \Delta u)\bar{\xi}(0)\} - M\{\xi(u)\bar{\xi}(0)\}| = \\ &= |M\{[\xi(u + \Delta u) - \xi(u)]\bar{\xi}(0)\}| \leq \{M|\xi(0)|^2 \times \\ &\quad \times M|\xi(u + \Delta u) - \xi(u)|^2\}^{1/2} \rightarrow 0 \end{aligned} \quad (13.95)$$

при $\Delta u \rightarrow 0$ в силу (13.94), см. теорему 1.13.2.

Пример 1.13.12. *Стационарный процесс с дискретным спектром.* Рассмотрим случайный процесс $\xi(\cdot)$, представляющий собой сумму гармонических колебаний со случайными амплитудами и фазами:

$$\xi(t) = \sum_{k=-n}^n \eta_k e^{i\lambda_k t}, \quad -\infty < t < \infty,$$

где λ_k , $k = -n, \dots, n$, — действительные числа, η_k , $k = -n, \dots, n$, — некоррелированные случайные величины (комплексные) с нулевым математическим ожиданием: $M\eta_k = 0$, $M(\eta_k \bar{\eta}_j) = 0$, $k \neq j$, $k, j = -n, \dots, n$. Тогда для всех $t, u \in (-\infty, \infty)$

$$\begin{aligned} M\xi(t) &= \sum_{k=-n}^n e^{i\lambda_k t} M\eta_k = 0, \\ M\{\xi(t+u)\bar{\xi}(t)\} &= \sum_{k=-n}^n \sum_{j=-n}^n e^{i\lambda_k(t+u) - i\lambda_j t} M(\eta_k \bar{\eta}_j) = \\ &= \sum_{k=-n}^n e^{i\lambda_k u} F_k^2 = R(u), \end{aligned} \quad (13.96)$$

где $F_k^2 = M|\eta_k|^2$, $k = -n, \dots, n$. Таким образом, $\xi(t)$, $t \in (-\infty, \infty)$, — стационарный случайный процесс. Он служит примером случайных колебаний (например, в электрической цепи или нагретой плазме). Его составляющие — гармонические колебания. Величину $M|\xi(\cdot)|^2$ называют *энергией стационарного процесса*. Из (13.96) видим, что (при $u = 0$)

$$M|\xi(t)|^2 = \sum_{k=-n}^n F_k^2 = \sum_{k=-n}^n M|\eta_k|^2, \quad (13.97)$$

т. е. энергия стационарного процесса $\xi(\cdot)$ складывается из энергий его гармонических составляющих. Значения λ_k , $k = -n, \dots, n$, называют частотами, а их совокупность $\lambda_{-n}, \dots, \lambda_n$ образуют *частотный спектр* стационарного процесса $\xi(\cdot)$.

Для дальнейшего нам понадобятся некоторые сведения о положительно определенных функциях

Определение 1.13.9. Непрерывная функция $f(t)$, $-\infty < t < \infty$, называется *положительно определенной*, если для любых натурально-

го n , вещественных чисел $t_1, t_2, \dots, t_n$ и комплексных чисел $z_1, z_2, \dots, z_n$ справедливо неравенство

$$\sum_{k,j=1}^n f(t_k - t_j) z_k \bar{z}_j \geq 0. \quad (13.98)$$

Примером положительно определенной функции является характеристическая функция любой случайной величины. Действительно, в силу определения $f_\xi(t) = M e^{it\xi}$, $-\infty < t < \infty$, непрерывна и

$$\sum_{k,j=1}^n f(t_k - t_j) z_k \bar{z}_j = \sum_{k,j=1}^n M e^{i(t_k - t_j)\xi} z_k \bar{z}_j = M \left| \sum_{k=1}^n e^{it_k \xi} z_k \right|^2 \geq 0. \quad (13.99)$$

Мы будем опираться на следующую известную теорему анализа.

Лемма 1.13.1. (Бохнер–Хинчин). Для того чтобы непрерывная функция $f(t)$, $-\infty < t < \infty$, такая, что $f(0) = 1$, являлась характеристической функцией некоторой случайной величины, необходимо и достаточно, чтобы она была положительно определенной. (Необходимость сформулированных условий очевидна в силу примера (13.99).)

Теперь сформулируем и докажем основную теорему теории стационарных процессов.

Теорема 1.13.7. (Хинчин). Для того чтобы абсолютно интегрируемая на $\mathcal{R}^1$ функция $R(\cdot)$ являлась корреляционной функцией некоторого с. к. непрерывного стационарного процесса, необходимо и достаточно, чтобы она могла быть представлена в виде

$$R(u) = \int_{-\infty}^{\infty} e^{iu\lambda} p(\lambda) d\lambda, \quad -\infty < u < \infty, \quad (13.100)$$

где $p(\lambda)$, $-\infty < \lambda < \infty$, — некоторая плотность вероятности.

Доказательство. Н е о б х о д и м о с т ь. Пусть $R(\cdot)$ — корреляционная функция с. к. непрерывного стационарного процесса $\xi(t)$. Тогда она непрерывна (см. (13.95)), $R(0) = 1$. Покажем, что $R(\cdot)$ положительно определена. В самом деле, каковы бы ни были натуральное n , действительные числа $u_1, \dots, u_n$ и комплексные $z_1, \dots, z_n$

$$\begin{aligned} \sum_{k,j=1}^n R(u_k - u_j) z_k \bar{z}_j &= \sum_{k,j=1}^n M \{ \xi(u_k) \bar{\xi}(u_j) \} z_k \bar{z}_j = \\ &= M \left\{ \sum_{k,j=1}^n \xi(u_k) \bar{\xi}(u_j) z_k \bar{z}_j \right\} = M \left| \sum_{k=1}^n \xi(u_k) z_k \right|^2 \geq 0. \end{aligned} \quad (13.101)$$

Таким образом, согласно лемме Бохнера–Хинчина, $R(u)$, $-\infty < u < \infty$, — характеристическая функция некоторой случайной величины η , а так как по условию теоремы $R(\cdot)$ абсолютно интегрируема,

$\int_{-\infty}^{\infty} |R(u)| du < \infty$, то в силу теоремы 1.11.4 случайная величина η непрерывна, ее плотность распределения вероятности — $p(\cdot)$, и, стало быть,

$$R(u) = \int_{-\infty}^{\infty} e^{iu\lambda} p(\lambda) d\lambda, \quad (13.102)$$

причем

$$p(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu\lambda} R(u) du, \quad -\infty < \lambda < \infty. \quad (13.103)$$

Необходимость доказана.

Д о с т а т о ч н о с т ь. Пусть справедливо представление (13.100). Покажем, что существует с. к. непрерывный стационарный процесс, для которого $R(\cdot)$ является корреляционной функцией. Рассмотрим непрерывную случайную величину η с плотностью вероятности $p(\cdot)$ из (13.100). Обозначим $\xi_1(\cdot)$ случайную функцию, определяемую равенством $\xi_1(t) = e^{it\eta}$, $-\infty < t < \infty$. Пусть ξ_2 — случайная величина, не зависящая от $\xi(t)$ при всех $t \in (-\infty, \infty)$ и принимающая два значения ± 1 с вероятностями $1/2$. Положим $\xi(t) = \xi_1(t)\xi_2$, $-\infty < t < \infty$. Случайный процесс $\xi(\cdot)$ обладает следующими свойствами:

1) $M\xi(t) = M\xi_1(t)M\xi_2 = 0$, $-\infty < t < \infty$, так как $M\xi_2 = 1 \cdot 1/2 - 1 \cdot 1/2 = 0$;

2) $\xi(t)$, $-\infty < t < \infty$, — с. к. непрерывен, ибо $M|\xi(t+h) - \xi(t)|^2 = M\xi_2^2 \cdot M|e^{i(t+h)\eta} \cdot e^{it\eta}|^2 = M|e^{ih\eta} - 1|^2 = \int_{-\infty}^{\infty} |e^{ihx} - 1|^2 p(x) dx \rightarrow 0$, $h \rightarrow 0$;

3) $R(u) = M\{\xi(t+u)\bar{\xi}(t)\} = M\xi_2^2 \cdot M\{e^{i(t+u)\eta} - e^{-it\eta}\} = Me^{iu\eta} = \int_{-\infty}^{\infty} e^{iux} p(x) dx$, $-\infty < u < \infty$, что совпадает с (13.100). ■

Заметим, что для случая вещественного стационарного процесса представление (13.100) имеет вид

$$R(u) = \int_{-\infty}^{\infty} \cos(u\lambda) p(\lambda) d\lambda, \quad -\infty < u < \infty. \quad (13.104)$$

Стационарный процесс, корреляционная функция которого представима в виде интеграла (13.100) или (13.104), называется *стационарным процессом с непрерывным спектром*, а функция $p(\lambda)$, $-\infty < \lambda < \infty$, называется *спектральной плотностью* стационарного процесса.

Средняя энергия процесса $\xi(\cdot)$ выражается через его спектральную плотность следующим образом: для любого $t \in (-\infty, \infty)$

$$M|\xi(t)|^2 = R(0) = \int_{-\infty}^{\infty} p(\lambda) d\lambda \quad (13.105)$$

(при нашем соглашении о нормировке $R(0) = 1$). Эта формула вполне аналогична выражению для энергии случайного процесса с дискретным спектром, пример 1.13.12 которого рассмотрен выше, и объясняет термин «спектральная плотность»: $p(\lambda)$ определяет вклад в энергию, отвечающий спектральному интервалу $(\lambda, \lambda + d\lambda)$. Величина

$$\tau = \int_0^{\infty} |R(t)| dt$$

называется *временем корреляции*. Время корреляции τ дает приближенное представление о том, на каких интервалах времени имеется корреляция между значениями (сечениями) случайного процесса. При существенно больших интервалах парными корреляциями можно пренебрегать. Ниже будут даны примеры подсчета времени корреляции.

Одним из важнейших свойств стационарных процессов является их эргодичность (см. ниже теорему 1.13.9).

Теорема 1.13.8. *Для корреляционной функции $R(u)$, $0 < u < \infty$, с.к. непрерывного стационарного вещественного процесса существует предел*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T R(u) du = 0. \quad (13.106)$$

Доказательство. Ограничимся случаем вещественного стационарного процесса с *непрерывным спектром*. На основании представления (13.104) имеем

$$\begin{aligned} \frac{1}{T} \int_0^T R(u) du &= \frac{1}{T} \int_0^T du \int_{-\infty}^{\infty} \cos(u\lambda) p(\lambda) d\lambda = \int_{-\infty}^{\infty} \frac{\sin(T\lambda)}{T\lambda} p(\lambda) d\lambda = \\ &= \int_{-\delta}^{\delta} \frac{\sin(T\lambda)}{T\lambda} p(\lambda) d\lambda + \int_{|\lambda| \geq \delta} \frac{\sin(T\lambda)}{T\lambda} p(\lambda) d\lambda. \end{aligned} \quad (13.107)$$

Пусть $\varepsilon > 0$ — любое. Первый интеграл в правой части (13.107) оценим следующим образом, учитывая, что $|\sin(T\lambda)| \leq T|\lambda|$ для всех

λ и T :

$$\left| \int_{-\delta}^{\delta} \frac{\sin(T\lambda)}{T\lambda} p(\lambda) d\lambda \right| \leq \int_{-\delta}^{\delta} p(\lambda) d\lambda < \frac{\varepsilon}{2}, \quad (13.108)$$

если $\delta = \delta(\varepsilon)$ достаточно мало, поскольку $p(\cdot)$ — непрерывная функция. Зафиксировав такое δ , оценим второй интеграл в правой части (13.107)

$$\left| \int_{|\lambda| \geq \delta} \frac{\sin(T\lambda)}{T\lambda} p(\lambda) d\lambda \right| \leq \frac{1}{T\delta} \int_{|\lambda| \geq \delta} p(\lambda) d\lambda \leq \frac{1}{T\delta} < \frac{\varepsilon}{2}, \quad (13.109)$$

если $T > 0$ — достаточно велико ($T \geq 2/(\delta(\varepsilon)\varepsilon)$). Из (13.108) и (13.109) заключаем, что

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T R(u) du = 0. \blacksquare \quad (13.110)$$

Рассмотрим ряд примеров, иллюстрирующих теорию стационарных процессов.

Пример 1.13.13. Пусть $\xi(t)$, $-\infty < t < \infty$, — экспоненциально-коррелированный стационарный процесс, т. е. процесс, имеющий корреляционную функцию

$$R(u) = e^{-\beta|u|}, \quad -\infty < u < \infty, \quad \beta > 0.$$

Его спектральная плотность согласно (13.103) равна $p(\lambda) = \frac{\pi^{-1}\beta}{\lambda^2 + \beta^2}$, $-\infty < \lambda < \infty$, а время корреляции $\tau = 1/\beta$.

Пример 1.13.14. Если корреляционная функция $R(\cdot)$ стационарного процесса имеет вид $R(u) = e^{-\beta|u|} \cos(\omega u)$, $-\infty < u < \infty$, $\beta, \omega > 0$, то его спектральная плотность равна $p(\lambda) = \frac{\beta}{\pi} \frac{\lambda^2 + \omega^2 + \beta^2}{(\lambda^2 - \omega^2 - \beta^2)^2 + 4\beta^2\lambda^2}$, $-\infty < \lambda < \infty$. При выполнении условия $\beta \ll \omega$ спектральная плотность сосредоточена в относительно узкой полосе частот $\lambda - \omega \sim \beta$, и справедливо приближенное соотношение $p(\lambda) \approx \frac{\beta}{\pi} \frac{\lambda^2 + \omega^2}{(\lambda^2 - \omega^2)^2 + 4\beta^2\lambda^2} \approx \frac{\beta}{2\pi[(\lambda - \omega)^2 + \beta^2]}$, а время корреляции $\tau \approx \frac{2}{\pi} \frac{1}{\beta}$.

Пример 1.13.15. Если корреляционная функция $R(\cdot)$ имеет вид $R(u) = e^{-\frac{1}{4}\beta^2 u^2}$, $-\infty < u < \infty$, то $p(\lambda) = \frac{1}{\sqrt{\pi}\beta} e^{-\lambda^2/\beta^2}$, а $\tau = \frac{1}{\beta} \sqrt{\frac{\pi}{2}}$, $-\infty < \lambda < \infty$.

Пример 1.13.16. Фильтрация. Пусть $\xi(t) = \sum_{k=1}^n c_k \eta(t - \tau_k)$, $-\infty < t < \infty$, где c_k, τ_k — постоянные, $k = 1, \dots, n$, причем τ_k — ве-

шественные числа, а $\eta(t)$, $-\infty < t < \infty$, — стационарный случайный процесс с непрерывным спектром, $M\eta(t) = 0$. Тогда случайный процесс $\xi(\cdot)$, называемый линейным преобразованием процесса $\eta(\cdot)$, также стационарен. Действительно, для $u, t \in (-\infty, \infty)$

$$\begin{aligned} M\xi(t) &= 0, R_\xi(u) = M\{\xi(u)\bar{\xi}(0)\} = M\left\{\sum_{k=1}^n c_k \eta(u - \tau_k) \sum_{j=1}^n \overline{c_j \eta(-\tau_j)}\right\} = \\ &= \sum_{k,j=1}^n c_k \bar{c}_j M\{\eta(u - \tau_k) \overline{\eta(-\tau_j)}\} = \sum_{k,j=1}^n c_k \bar{c}_j R_\eta(u - \tau_k + \tau_j). \end{aligned}$$

Отсюда получаем выражение для спектральной плотности процесса $\xi(t)$.

$$\begin{aligned} p_\xi(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda u} R_\xi(u) du = \frac{1}{2\pi} \sum_{k,j=1}^n c_k \bar{c}_j \int_{-\infty}^{\infty} R_\eta(u - \tau_k + \tau_j) e^{-i\lambda u} du = \\ &= \sum_{k,j=1}^n c_k \bar{c}_j e^{-i\lambda(\tau_k - \tau_j)} p_\eta(\lambda) = \left| \sum_{k=1}^n c_k e^{-i\lambda \tau_k} \right|^2 p_\eta(\lambda), \quad -\infty < \lambda < \infty. \end{aligned}$$

Таким образом, спектральная плотность процесса $\xi(\cdot)$ получается из спектральной плотности процесса $\eta(\cdot)$ умножением на квадрат модуля частотной характеристики преобразования (фильтра).

Пример 1.13.17. *Эргодичность стационарного процесса.* Оказывается, что аналог равенства (13.106) имеет место и для самого стационарного процесса $\xi(\cdot)$. Однако прежде чем сформулировать и доказать соответствующую теорему, мы введем новое понятие: *интеграла от случайной функции, так называемого стохастического интеграла*. Случайная функция (случайный процесс) $\xi(t)$, $t \in T$, — это функция двух переменных: $\xi(t, \omega)$, $t \in T$, $\omega \in \Omega$; пусть множество T , ради определенности, интервал $[a, b]$. Даже в тех (довольно редких) случаях, когда $\xi(t, \omega)$ как функция t , $t \in [a, b]$, оказывается интегрируемой (по Риману) функцией при всех $\omega \in \Omega$, так что интеграл $\int_a^b \xi(t) dt \equiv \int_a^b \xi(t, \omega) dt$ имеет смысл при каждом фиксированном $\omega \in \Omega$, вопрос о том, является ли этот интеграл случайной величиной, определенной на Ω , нетривиален. Мы дадим определение стохастического интеграла, основанное на сходимости в среднем квадратичном, см. §1.13.2.

Выше (см. (13.94)) было дано определение с. к. непрерывного случайного процесса. Усилим это определение: случайная функция $\xi(t)$, $t \in [a, b]$, называется *равномерно непрерывной в среднеквадратичном*, если

$$M|\xi(t+h) - \xi(t)|^2 \rightarrow 0, \quad h \rightarrow 0, \quad (13.111)$$

равномерно по $t \in [a, b]$.

Для определения *стохастического интеграла равномерно с. к. непрерывной случайной функции* $\xi(t)$, $a \leq t \leq b$, возьмем произвольное

разбиение сегмента $[a, b]$ точками t_k , $k = 0, \dots, n$:

$$a = t_0 < t_1 < \dots < t_n = b, \\ \Delta t_k = t_{k+1} - t_k, \quad \Delta = \max_k \Delta t_k, \quad k = 0, 1, \dots, n.$$

Составим интегральную сумму

$$\sum_{k=0}^{n-1} \xi(t_k) \Delta t_k = \eta(\Delta), \quad (13.112)$$

$\eta(\Delta)$ — случайная величина, $M\eta(\Delta) = \sum_{k=0}^{n-1} M\xi(t_k) \Delta t_k$. Определим для случайной величины ξ норму по формуле (если $M|\xi|^2 < \infty$)

$$\|\xi\| = \sqrt{M|\xi|^2}. \quad (13.113)$$

Так определенная величина обладает всеми свойствами нормы, в частности, для нее справедливо неравенство треугольника (следствие неравенства Коши-Буняковского, см. §1.9):

$$\|\xi_1 - \xi_2\| \leq \|\xi_1 - \xi_3\| + \|\xi_2 - \xi_3\|.$$

Покажем, что при измельчении разбиения сегмента $[a, b]$ так, что $\Delta \rightarrow 0$, последовательность интегральных сумм (13.112) фундаментальна по норме (13.113). В самом деле, пусть $\{t'_k\}$ и $\{t''_k\}$ — два произвольных разбиения сегмента $[a, b]$, $\Delta_1 = \max_k \Delta t'_k$, $\Delta_2 = \max_k \Delta t''_k$. Обозначим $\{t_k\}$ суммарное разбиение, т. е. разбиение сегмента $[a, b]$, в которое входят все различные точки обоих разбиений $\{t'_k\}$ и $\{t''_k\}$, тогда если $\Delta = \max_k \Delta t_k$, то $\Delta \leq \min\{\Delta_1, \Delta_2\}$. При этом

$$\begin{aligned} \left\| \sum_k \xi(t'_k) \Delta t'_k - \sum_k \xi(t''_k) \Delta t''_k \right\| &= \left\| \sum_k [\xi(t'_k) - \xi(t''_k)] \Delta t_k \right\| \leq \\ &\leq \sum_k \|\xi(t'_k) - \xi(t''_k)\| \Delta t_k < \varepsilon \sum_k \Delta t_k = \varepsilon(b-a) \end{aligned} \quad (13.114)$$

ввиду (13.111), если Δ_1 и $\Delta_2 < \delta(\varepsilon)$, предпоследнее неравенство справедливо в силу неравенства Минковского. Можно показать что фундаментальная в норме (13.113) последовательность $\eta(\Delta)$ сходится при $\Delta \rightarrow 0$ к некоторой случайной величине, см. Критерий Коши в §1.13.2. Эта случайная величина называется (по определению) *стохастическим интегралом* от $\xi(t)$, $a \leq t \leq b$, и обозначается $\int_a^b \xi(t) dt$. Очевидно, что этот интеграл обладает свойством линейности

$$\int_a^b (c_1 \xi_1(t) + c_2 \xi_2(t)) dt = c_1 \int_a^b \xi_1(t) dt + c_2 \int_a^b \xi_2(t) dt, \quad (13.115)$$

и, кроме того,

$$M \int_a^b \xi(t) dt = \int_a^b M \xi(t) dt. \quad (13.116)$$

Теорема 1.13.9. Для равномерно с. к. непрерывного стационарного процесса $\xi(t)$, $t \in [a, b]$ случайная величина

$$\bar{\xi}(T) = \frac{1}{T} \int_0^T \xi(t) dt \xrightarrow{P} 0, \quad T \rightarrow \infty \quad (13.117)$$

(сходится по вероятности к нулю при $T \rightarrow \infty$).

Доказательство. Ограничимся случаем вещественного стационарного процесса $\xi(\cdot)$ с непрерывным спектром. Покажем, что для любого $\varepsilon > 0$

$$\lim_{T \rightarrow \infty} P\{|\bar{\xi}(T)| > \varepsilon\} = 0. \quad (13.118)$$

Поскольку $M\bar{\xi}(T) = \frac{1}{T} \int_0^T M \xi(t) dt = 0$, то в силу неравенства Чебышева

$$P\{|\bar{\xi}(T)| > \varepsilon\} \leq \frac{D\bar{\xi}(T)}{\varepsilon^2}, \quad (13.119)$$

и достаточно доказать, что $\lim_{T \rightarrow \infty} D\bar{\xi}(T) = 0$. Так как $M\bar{\xi}(T) = 0$, то

$$\begin{aligned} D\bar{\xi}(T) &= M(\bar{\xi}(T))^2 = \frac{1}{T^2} M \left(\int_0^T \xi(t) dt \int_0^T \xi(u) du \right) = \\ &= \frac{1}{T^2} \int_0^T \int_0^T M\{\xi(t)\xi(u)\} dt du = \frac{1}{T^2} \int_0^T \int_0^T R(t-u) dt du. \end{aligned} \quad (13.120)$$

Отсюда, пользуясь представлением (13.104), получим

$$\begin{aligned} D\bar{\xi}(T) &= \frac{1}{T^2} \int_0^T \int_0^T dt du \int_{-\infty}^{\infty} \cos((t-u)\lambda) p(\lambda) d\lambda = \\ &= \frac{1}{T^2} \int_{-\infty}^{\infty} p(\lambda) d\lambda \int_0^T \int_0^T \cos(\lambda(t-u)) dt du = 2 \int_{-\infty}^{\infty} \frac{1 - \cos(\lambda T)}{\lambda^2 T^2} p(\lambda) d\lambda. \end{aligned} \quad (13.121)$$

Теперь интеграл в (121) оценим, как в теореме 1.13.6,

$$\begin{aligned} D\bar{\xi}(T) &= \int_{-\infty}^{\infty} \frac{1 - \cos(\lambda T)}{\lambda^2 T^2} p(\lambda) d\lambda = \\ &= 2 \int_{-\delta}^{\delta} \frac{1 - \cos(\lambda T)}{\lambda^2 T^2} p(\lambda) d\lambda + 2 \int_{|\lambda| \geq \delta} \frac{1 - \cos(\lambda T)}{\lambda^2 T^2} p(\lambda) d\lambda. \end{aligned} \quad (13.122)$$

Поскольку $0 \leq 1 - \cos(\lambda T) \leq \lambda^2 T^2 / 2$ для всех λ и T , то первый интеграл в (13.122) меньше любого $\varepsilon_1 > 0$, если $\delta = \delta(\varepsilon_1)$ достаточно мало. Зафиксировав это δ , второй интеграл в (13.122) оценим так

$$\left| 2 \int_{|\lambda| \geq \delta} \frac{1 - \cos(\lambda T)}{\lambda^2 T^2} p(\lambda) d\lambda \right| \leq \frac{4}{\delta^2 T^2} \int_{|\lambda| \geq \delta} p(\lambda) d\lambda \leq \frac{4}{\delta^2 T^2} < \varepsilon_1,$$

если T достаточно велико. Тем самым равенство (13.120) доказано. Отсюда, как указано выше, следует (13.118). ■

Пример 1.13.18. *Спектр колебания с флуктуирующей частотой.* Пусть задано колебание

$$\xi(t) = A_0 \cos[\omega_0 t + \int_{t_0}^t \eta(t) dt + \varphi_0], \quad t \geq t_0, \quad (13.123)$$

в котором девиация частоты $\eta(\cdot)$ — вещественный стационарный равномерно с. к. непрерывный случайный процесс, а φ_0 — случайная величина, равномерно распределенная в интервале $[0, 2\pi]$. Рассмотрим вопрос о том, как связан спектр колебания $\xi(\cdot)$ со спектром девиации частоты $\eta(\cdot)$. Естественно считать, что $M\eta(t) = 0$, $t_0 \leq t$. Согласно представлению (13.104) корреляционная функция $R_\eta(\cdot)$ имеет вид

$$R_\eta(u) = M\{\eta(t+u)\eta(t)\} = \int_0^\infty p_\eta(\lambda) \cos(\lambda u) d\lambda, \quad (13.124)$$

где $p_\eta(\cdot)$ — спектральная плотность стационарного процесса $\eta(\cdot)$. Для случайного набегу фазы $\psi(\tau)$ за время от t до $t + \tau$

$$\psi(\tau) = \int_t^{t+\tau} \eta(z) dz, \quad (13.125)$$

а так как $M\eta(t) = 0$, то $M\psi(\tau) = 0$. Обозначим $\sigma^2(\tau)$ дисперсию $\psi(\tau)$

$$\sigma^2(\tau) = M\psi^2(\tau) = \int_t^{t+\tau} \int_t^{t+\tau} M\{\eta(z)\eta(u)\}dzdu = \int_t^{t+\tau} \int_t^{t+\tau} R_\eta(z-u)dzdu. \quad (13.126)$$

Подставляя в (13.126) выражение (13.124), получим (ср. теорему 1.13.7)

$$\sigma^2(\tau) = 2 \int_0^\infty p_\eta(\lambda) \frac{1 - \cos(\tau\lambda)}{\lambda^2} d\lambda. \quad (13.127)$$

Таким образом, средний квадрат набега фазы определяется спектром девиации частоты.

Пусть теперь девиация частоты $\eta(\cdot)$ распределена по нормальному закону. В силу равенства (13.125) это же верно и в отношении набега фазы $\psi(\tau)$ (13.125), а так как $M\psi(\tau) = 0$, и дисперсия $\psi(\tau)$ равна $\sigma^2(\tau)$, то плотность распределения $\psi(\tau)$ имеет вид

$$p(x, \tau) = \frac{1}{\sigma(\tau)\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2(\tau)}}, \quad -\infty < x < \infty. \quad (13.128)$$

Найдем корреляционную функцию $R_\xi(\cdot)$ стационарного процесса (13.123), $R_\xi(u) = M\{\xi(t+u)\xi(t)\}$. Так как

$$\begin{aligned} \xi(t+u)\xi(t) = & \frac{A_0^2}{2} \{ \cos[\omega_0 u - \psi(u)] + \cos[\omega_0(2t+u) - \\ & - \int_{t_0}^{t+u} \eta(\tau)d\tau - \int_{t_0}^t \eta(\tau)d\tau + 2\varphi_0] \}, \end{aligned} \quad (13.129)$$

то для вычисления $R_\xi(\cdot)$ следует усреднить (13.129) по распределению (13.128) для $\psi(\tau)$ и равномерному распределению для φ_0 . Интегрирование по φ_0 от 0 до 2π обращает в нуль интеграл от второго косинуса

в (13.129). Таким образом, получаем

$$\begin{aligned}
 R_\xi(u) &= \frac{A_0^2}{2} \frac{1}{\sigma(u)\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(\omega_0 u - x) e^{-\frac{x^2}{2\sigma^2(u)}} dx = \\
 &= \frac{A_0^2}{2} \frac{\cos(\omega_0 u)}{\sigma(u)\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(x) e^{-\frac{x^2}{2\sigma^2(u)}} dx = \\
 &= \frac{A_0^2}{2} \frac{\cos(\omega_0 u)}{\sigma(u)\sqrt{2\pi}} \frac{1}{2} \int_{-\infty}^{\infty} (e^{ix} + e^{-ix}) e^{-\frac{x^2}{2\sigma^2(u)}} dx = \\
 &= \frac{A_0^2}{2} \cos(\omega_0 u) e^{-\frac{\sigma^2(u)}{2}}, \quad -\infty < u < \infty.
 \end{aligned} \tag{13.130}$$

(для вычисления интеграла надо выделить полный квадрат в показателе экспоненты). Поскольку по теореме Хинчина для $R_\xi(\cdot)$ справедливо спектральное представление

$$R_\xi(u) = \int_0^\infty p_\xi(\lambda) \cos(\lambda u) d\lambda, \tag{13.131}$$

то согласно формуле обращения для интеграла Фурье из (13.130) и (13.131) найдем

$$p_\xi(\lambda) = \frac{A_0^2}{2\pi} \int_0^\infty e^{-\frac{\sigma^2(u)}{2}} \cos(\omega_0 u) \cos(\lambda u) du. \tag{13.132}$$

Для определения спектра девиации частоты $\eta(\cdot)$ по спектру наблюдаемых колебаний $\xi(\cdot)$ мы приходим к следующему нелинейному интегральному уравнению для $p_\eta(\cdot)$ (подставляем (13.127) в (13.132)):

$$\begin{aligned}
 p_\xi(\lambda) &= \frac{A_0^2}{2\pi} \int_0^\infty \exp\left\{-\int_0^\infty p_\eta(\lambda') \frac{1 - \cos(u\lambda')}{\lambda'} d\lambda'\right\} \cos(\omega_0 u) \cos(\lambda u) du, \\
 &\quad 0 \leq \lambda < \infty.
 \end{aligned} \tag{13.133}$$

Это сложное интегральное уравнение может быть решено при различных упрощающих предположениях.

Пример 1.13.19. *Корреляционная функция стационарного марковского процесса.* Целью этого примера является доказательство **теоремы Дуба:** корреляционная функция $R_\xi(\cdot)$ нормального $N(0, 1)$ стационарного марковского процесса равна

$$R_\xi(u) = e^{-\alpha|u|}, \quad -\infty < u < \infty, \quad \alpha > 0. \tag{13.134}$$

В самом деле, двумерная плотность нормального $N(0, 1)$ процесса $\xi(\cdot)$ равна (см. §1.9)

$$p_2(t_1, x_1; t_2, x_2) = \frac{1}{2\pi\sqrt{1-\sigma^2}} e^{-\frac{1}{2(1-\sigma^2)}(x_1^2 + x_2^2 - 2\sigma x_1 x_2)},$$

$$x_1, x_2 \in \mathcal{R}^1, t_2 > t_1, \quad (13.135)$$

где $\sigma = M\{\xi(t_2)\xi(t_1)\} = R(t_2 - t_1)$, поскольку процесс по условию стационарный. Одномерная плотность

$$p_1(t_1, x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}, \quad -\infty < t_1 < \infty, \quad -\infty < x_1 < \infty, \quad (13.136)$$

ибо $\xi(t) \in N(0, 1)$ при каждом t . Отсюда для плотности переходной вероятности находим

$$q(t_2, x_2|t_1, x_1) = \frac{p_2(t_1, x_1; t_2, x_2)}{p_1(t_1, x_1)} = \frac{1}{\sqrt{2\pi(1-\sigma^2)}} e^{-\frac{1}{2} \frac{(\sigma x_1 - x_2)^2}{1-\sigma^2}}, \quad t_1 < t_2. \quad (13.137)$$

Плотность переходной вероятности удовлетворяет уравнению Смолуховского (13.56)

$$q(t_2, x_2|t_1, x_1) = \int_{-\infty}^{\infty} q(t_2, x_2|\tau, y) q(\tau, y|t_1, x_1) dy, \quad t_1 < \tau < t_2. \quad (13.138)$$

Подставляя в (13.138) выражение (13.137), получим

$$\begin{aligned} & \frac{1}{\sqrt{2\pi(1-\sigma_2^2(t_2-t_0))}} e^{-\frac{1}{2} \frac{(\sigma_2(t_2-t_0)x_1 - x_2)^2}{1-2\sigma_2^2(t_2-t_0)}} = \\ & = \frac{1}{2\pi\sqrt{(1-\sigma_1^2)(1-\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \frac{(\sigma_1 y - x_2)^2}{1-\sigma_1^2}} e^{-\frac{1}{2} \frac{(\sigma_0 x_1 - y)^2}{1-\sigma_0^2}} dy = \\ & = \frac{1}{\sqrt{2\pi(1-\sigma_1^2\sigma_0^2)}} e^{-\frac{1}{2} \frac{(\sigma_1\sigma_0 x_1 - x_2)^2}{1-\sigma_1^2\sigma_0^2}}, \end{aligned} \quad (13.139)$$

где $\sigma_2 = \sigma(t_2 - t_1)$, $\sigma_1 = \sigma(t_2 - \tau)$, $\sigma_0 = \sigma(\tau - t_1)$. Сравнивая левую и правую части формулы (13.139), заключаем, что $\sigma_2 = \sigma_1\sigma_0$, или, подробнее, $R(t_2 - t_1) = R(t_2 - \tau)R(\tau - t_1)$, $t_1 < \tau < t_2$. Решением этого функционального уравнения служит функция $R(u) = e^{-\alpha u}$, $u \neq 0$, α — любое комплексное. Поскольку $\xi(t) \in N(0, 1)$, то $R(\cdot)$ вещественна и, кроме того, $|R(u)| \leq 1$, ибо $|R(u)| = |M\xi(u)\xi(0)| \leq \sqrt{M\xi^2(u)M\xi^2(0)} = 1$. Отсюда $\alpha \geq 0$, и теорема Дуба доказана. ■

Английский ботаник Р. Броун обнаружил (1828), что частицы пыльцы, взвешенные в воде, совершают непрерывное беспорядочное движение, названное впоследствии «броуновским». Физическая теория броуновского движения впервые была предложена А. Эйнштейном и Э. Смолуховским (1905). Л. Бachelье (1900) нестрого вывел закон движения частицы, совершающей одномерное броуновское движение.

Понятие марковской зависимости было введено А. А. Марковым (1906) для случая конечных однородных цепей. А. Н. Колмогоров в работе 1931 г. дал строгое определение марковской зависимости для непрерывной схемы (время меняется непрерывно, а множество состояний конечно, счетно или непрерывно) и вывел основные дифференциальные уравнения для переходных вероятностей. Колмогоров назвал рассмотренные им процессы стохастически определенными, но вскоре для них возник термин «процессы без последствия», а затем по предложению А. Я. Хинчина указанные процессы стали называть «марковскими». Дальнейшее развитие теории марковских процессов получила в работах В. Феллера (1934), П. Леви (1939) и ряда других математиков.

Основы теории стационарных процессов были заложены А. Я. Хинчиным (1934). В настоящее время теория случайных процессов — бурно развивающаяся область математики с широкими приложениями в физике и технике.

Глава 2

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Предмет теории вероятностей составляют стохастические эксперименты, математическими моделями которых являются некоторые вероятностные пространства. К математической статистике принято относить все вопросы, касающиеся сравнения этих моделей с реальностью, например, — вопрос о состоятельности стохастической модели реального эксперимента (явления и т. п.). Последний обычно формулируется как задача проверки гипотезы о том, что результаты, получаемые на основе математической модели эксперимента, не противоречат опытным данным. Поскольку вывод о непротиворечивости должен быть основан на опытных данных, содержащих элемент случайности, принимаемое решение следует формулировать в теоретико-вероятностных терминах.

Начало периода интенсивного развития статистических методов определяется работами К. Пирсона, выполненными в конце девятнадцатого столетия. Раздел математической статистики, в котором изучаются задачи проверки гипотез, был создан Ю. Нейманом, Э. С. Пирсоном и Р. А. Фишером в конце двадцатых — начале тридцатых годов прошлого столетия и с тех пор получил значительное развитие.

В других задачах математической статистики опытные данные привлекаются лишь для уточнения математической модели явления. К этому кругу вопросов относятся задачи оценивания параметров модели, составляющие один из центральных разделов математической статистики. Первые результаты в этой области получены К. Ф. Гауссом (1809) и А. А. Марковым (1900). Методы теории статистического оценивания получили глубокое развитие в работах Р. А. Фишера.

С более общих позиций оба упомянутых раздела математической статистики объединены А. Вальдом (1939). Ему принадлежит единая исчерпывающая формулировка задач общей теории статистических решающих процедур.

2.1. Распределения ортогональных проекций нормального случайного вектора и функций от них

2.1.1. Введение. Пусть ξ — случайный вектор в n -мерном евклидовом пространстве $\mathcal{R}^n$. Если $e_1, \dots, e_n$ — ортонормированный базис

$\mathcal{R}^n$, то

$$\xi = \sum_{j=1}^n \xi_j e_j,$$

где ξ_j , $j = 1, \dots, n$, суть координаты вектора ξ , в базисе $e_1, \dots, e_n$, причем $\xi_j = (\xi, e_j)$, $j = 1, \dots, n$. Скалярные произведения (ξ, e_j) , $j = 1, \dots, n$, являются случайными величинами. Скалярное произведение двух случайных векторов ξ и η также является случайной величиной и в базисе $e_1, \dots, e_n$ определяется равенством

$$(\xi, \eta) = \sum_{j=1}^n \xi_j \eta_j,$$

где $\eta_j = (\eta, e_j)$, $j = 1, \dots, n$.

Если A — линейный оператор, действующий из $\mathcal{R}^n$ в $\mathcal{R}^n$, $A: \mathcal{R}^n \rightarrow \mathcal{R}^n$, то $\zeta = A\xi$ — случайный вектор, причем, как известно,

$$\zeta_i = (A\xi)_i = (A\xi, e_i) = \sum_{j=1}^n a_{ij} \xi_j, \quad i = 1, \dots, n,$$

где

$$a_{ij} = (Ae_j, e_i), \quad i, j = 1, \dots, n, \quad -$$

матричные элементы матрицы оператора A в ортонормированном базисе. Матрицу оператора A будем обозначать $\check{A}$.

Напомним, что переход к новому ортонормированному базису $e'_1, \dots, e'_n$ может быть задан с помощью оператора ортогонального преобразования $U: \mathcal{R}^n \rightarrow \mathcal{R}^n$, или — ортогонального оператора U :

$$e'_i = Ue_i, \quad i = 1, \dots, n.$$

Матрица $\check{U}$ оператора U в базисе $e_1, \dots, e_n$ обладает свойствами

$$\det \check{U} = \pm 1, \quad \check{U}^{\text{TP}} = \check{U}^{-1} \quad (1.1)$$

и называется *ортогональной*.

В новом базисе $e'_1, \dots, e'_n$ матричные элементы a'_{ij} , $i, j = 1, \dots, n$, матрицы оператора A вычисляются по формуле

$$a'_{ij} = (Ae'_j, e'_i) = (AUe_j, Ue_i) = (U^*AUe_j, e_i)$$

и равны

$$a'_{ij} = \sum_{p,q=1}^n \bar{u}_{ip} a_{pq} u_{qi}, \quad i, j = 1, \dots, n,$$

где u_{ij} — матричные элементы матрицы $\check{U}$, $\bar{u}_{ij}$, $i, j = 1, \dots, n$, — матричные элементы матрицы $\check{U}^{\text{TP}} = \check{U}^{-1}$ оператора U^* , сопряженного с U , $U^* = U^{-1}$, в базисе $e_1, \dots, e_n$.

Для любых двух векторов $x, y \in \mathcal{R}^n$:

$$(x, y) = (Ux, Uy) \quad (1.2)$$

и, следовательно, $\|Ux\| = \|x\|$, где $\|x\|$ — обозначение для нормы (длины) вектора x , $\|x\|^2 = (x, x) = \sum_{j=1}^n x_j^2$.

Пусть $\mathcal{L}$ — линейное подпространство $\mathcal{R}^n$ и $\mathcal{L}^\perp$ — ортогональное дополнение $\mathcal{L}$ в $\mathcal{R}^n$, т.е. множество векторов $x \in \mathcal{R}^n$, ортогональных всем векторам из $\mathcal{L}$:

$$\mathcal{L}^\perp = \{x \in \mathcal{R}^n, (x, y) = 0, y \in \mathcal{L}\}.$$

Очевидно, $\mathcal{L}^\perp$ — также линейное подпространство $\mathcal{R}^n$. Как известно, всякий вектор $x \in \mathcal{R}^n$ может быть представлен в виде суммы

$$x = x_1 + x_2, \quad x_1 \in \mathcal{L}, \quad x_2 \in \mathcal{L}^\perp. \quad (1.3)$$

Представление (1.3) единственное. Действительно, равенство $x = x'_1 + x'_2$, $x'_1 \in \mathcal{L}$, $x'_2 \in \mathcal{L}^\perp$, совместно с (1.3) влечет равенство $(x_1 - x'_1) + (x_2 - x'_2) = 0$. Слагаемые в последнем равенстве ортогональны, так как $x_1 - x'_1 \in \mathcal{L}$, $x_2 - x'_2 \in \mathcal{L}^\perp$, поэтому $\|x_1 - x'_1\|^2 + \|x_2 - x'_2\|^2 = 0$, т.е. $x_1 = x'_1$, $x_2 = x'_2$. Следовательно каждому $x \in \mathcal{R}^n$ разложением (1.3) ставится в соответствие единственный вектор $x_1 \in \mathcal{L}$, называемый ортогональной проекцией x на $\mathcal{L}$:

$$x_1 = Px. \quad (1.4)$$

P называется оператором ортогонального проецирования на $\mathcal{L}$, или ортогональным проектором на $\mathcal{L}$. Отметим следующие его свойства.

1) Оператор P — линейный оператор. Действительно, пусть

$$x = x_1 + x_2, \quad y = y_1 + y_2, \quad x_1, y_1 \in \mathcal{L}, \quad x_2, y_2 \in \mathcal{L}^\perp. \quad (1.5)$$

Тогда

$$\alpha x + \beta y = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2), \quad \alpha x_1 + \beta y_1 \in \mathcal{L}, \quad \alpha x_2 + \beta y_2 \in \mathcal{L}^\perp.$$

Следовательно, согласно определению (1.4) $P(\alpha x + \beta y) = \alpha x_1 + \beta y_1 = \alpha Px + \beta Py$, так как $Px = x_1$, $Py = y_1$.

2) P — самосопряженный оператор, $P^* = P$, т.е. для любых $x, y \in \mathcal{R}^n$ $(Px, y) = (x, Py)$. Действительно, воспользовавшись разложением (1.5), найдем

$$(Px, y) = (x_1, y) = (x_1, y_1) = (x, y_1) = (x, Py).$$

3) Оператор P удовлетворяет уравнению $P^2 = P$. Действительно, для всякого $x \in \mathcal{R}^n$: $Px = x_1 = Px_1 = P(Px)$, поскольку для $x \in \mathcal{L}$ разложение (1.3) имеет вид $x = x_1 (= Px_1)$.

На самом деле свойства 1)–3) не только необходимы, но и достаточны для того, чтобы обладающий ими оператор P был ортогональным проектором. Для доказательства предположим, согласно

свойству 1), что Π — линейный оператор. Обозначим $\mathcal{L}$ множество решений уравнения $\Pi x = x$, $\mathcal{N}$ — множество решений уравнения $\Pi x = 0$. Легко убедиться, что $\mathcal{L}$ и $\mathcal{N}$ — линейные подпространства $\mathcal{R}^n$, причем ортогональные, если Π удовлетворяет условию 2). В самом деле, если $x \in \mathcal{L}$, $y \in \mathcal{N}$, то $(x, y) = (\Pi x, y) = (x, \Pi y) = (x, 0) = 0$. Для всякого вектора $x \in \mathcal{R}^n$ можно записать тождество

$$x = \Pi x + (I - \Pi)x. \quad (1.6)$$

Если Π удовлетворяет условию 3), то $\Pi(\Pi x) = \Pi x$, т.е. $\Pi x \in \mathcal{L}$, и $\Pi(I - \Pi)x - (\Pi - \Pi^2)x = 0$, т.е. $(I - \Pi)x \in \mathcal{N}$. Следовательно, Π — оператор ортогонального проецирования на $\mathcal{L} = \{x \in \mathcal{R}^n, \Pi x = x\}$.

Из разложения (1.6) следует также, что оператор $I - \Pi$ ортогонально проецирует на $\mathcal{N} = \{x \in \mathcal{R}^n, (I - \Pi)x = x\} = \mathcal{L}^\perp$.

Отметим следующее важное свойство ортогонального проектора. Пусть Π — оператор ортогонального проецирования на линейное подпространство $\mathcal{L}$, и

$$\rho(x, \mathcal{L}) = \inf\{\|x - y\| | y \in \mathcal{L}\} \quad (1.7)$$

— расстояние от x до $\mathcal{L}$. Тогда

$$\rho(x, \mathcal{L}) = \|x - \Pi x\|. \quad (1.8)$$

Действительно, пусть $y \in \mathcal{L}$. Тогда $\Pi x - y \in \mathcal{L}$, $x - \Pi x = (I - \Pi)x \in \mathcal{L}^\perp$, и, следовательно,

$$\begin{aligned} \|x - y\|^2 &= \|x - \Pi x + \Pi x - y\|^2 = \|x - \Pi x\|^2 + \\ &+ \|\Pi x - y\|^2 \geq \|x - \Pi x\|^2, \end{aligned} \quad (1.9)$$

причем равенство здесь выполняется лишь в случае $\Pi x = y$.

2.1.2. Ортогональные преобразования, ортогональные проекции $N(0, \sigma^2 I)$ -случайного вектора, функции ортогональных проекций.

Теорема 2.1.1. Пусть ξ — случайный вектор, координаты которого $\xi_1, \dots, \xi_n$ в некотором ортонормированном базисе независимы в совокупности и нормальны $N(0, \sigma^2)$. Если U — оператор ортогонального преобразования $\mathcal{R}^n$, то распределение вектора $U\xi$ совпадает с распределением вектора ξ , т.е. случайные величины $(U\xi)_1, \dots, (U\xi)_n$, координаты вектора $U\xi$ распределены так же, как $\xi_1, \dots, \xi_n$.

Доказательство. По условию теоремы плотность совместного распределения координат $\xi_1, \dots, \xi_n$ (плотность распределения вектора ξ)

$$\begin{aligned} p_\xi(x) &= p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{(x, x)}{2\sigma^2}\right), \end{aligned} \quad (1.10)$$

где $x_1, \dots, x_n$ суть координаты вектора $x \in \mathcal{R}^n$. Для вектора $\eta = U\xi$ согласно формуле (8.56)

$$p_\eta(x) = p_{U\xi}(x) = p_\xi(U^{-1}x) |\det \check{U}^{-1}| = p_\xi(x), \quad x \in \mathcal{R}^n, \quad (1.11)$$

так как в равенстве (1.11) согласно (1.2) $(U^{-1}(x), U^{-1}(x)) = (x, x)$ и согласно (1.1) $|\det \check{U}^{-1}| = 1$. Из (1.10) и (1.11) следует, что координаты $(U\xi)_1, \dots, (U\xi)_n$ вектора $U\xi$ независимы в совокупности и нормальны $N(0, \sigma^2)$. ■

Замечание 2.1.1. Поскольку $(U\xi)_i = (U\xi, e_i) = (\xi, U^{-1}e_i)$, $i = 1, \dots, n$, и $e'_i = U^{-1}e_i$, $i = 1, \dots, n$, — ортонормированный базис $\mathcal{R}^n$, причем в таком виде может быть представлен любой ортонормированный базис $\mathcal{R}^n$, то утверждение теоремы 2.1.1 можно сформулировать еще и так: *если координаты $\xi_1, \dots, \xi_n$ случайного вектора ξ независимы в совокупности и нормальны $N(0, \sigma^2)$ в некотором ортонормированном базисе $\mathcal{R}^n$, то они независимы в совокупности и нормальны $N(0, \sigma^2)$ в любом ортонормированном базисе $\mathcal{R}^n$.*

Напомним, что сумма m квадратов независимых нормальных $N(0, 1)$ случайных величин имеет χ_m^2 -распределение (хи-квадрат) с m степенями свободы, см. Приложение. В частности, сумма любых $m \leq n$ квадратов координат вектора ξ , удовлетворяющего условиям теоремы 2.1.1, равна $\sigma^2 \chi_m^2$. Но при условиях теоремы 2.1.1 имеет место и более общий факт.

Следствие 2.1.1. Пусть $\mathcal{R}^k$ — k -мерное линейное подпространство $\mathcal{R}^n$, и Π_k — оператор ортогонального проектирования на $\mathcal{R}^k$. Если выполнены условия теоремы 2.1.1, то случайная величина $\|\Pi_k \xi\|^2 = (\Pi_k \xi, \xi) = (\Pi_k \xi, \Pi_k \xi)$ распределена как $\sigma^2 \chi_k^2$, $\|\Pi_k \xi\|^2 = \sigma^2 \chi_k^2$.

Доказательство. Введем в $\mathcal{R}^n$ ортонормированный базис $e'_1, \dots, e'_n$, причем так, чтобы первые k векторов $e'_1, \dots, e'_k$, $k \leq n$, образовывали ортонормированный базис $\mathcal{R}^k$. Согласно замечанию 2.1.1 к теореме 2.1.1 координаты $\xi'_j = (\xi, e'_j)$, $j = 1, \dots, n$, вектора ξ независимы в совокупности и нормальны $N(0, \sigma^2)$. Поэтому $\sum_{j=1}^k \xi_j'^2 = \sigma^2 \chi_k^2$. С другой стороны, если оператор Π_k ортогонально проецирует на $\mathcal{R}^k$, то $\Pi_k \xi = \sum_{j=1}^k \xi'_j e'_j$, так как $\mathcal{R}^k$ — линейная оболочка векторов $e'_1, \dots, e'_k$.

Поэтому $\|\Pi_k \xi\|^2 = (\Pi_k \xi, \xi) = \sum_{j=1}^k (\xi'_j)^2 = \sigma^2 \chi_k^2$. ■

Следствие 2.1.1 определяет геометрическую интерпретацию «числа степеней свободы» χ^2 -распределения.

Как известно, если $\xi_1, \dots, \xi_n$ независимы в совокупности и нормальны $N(0, 1)$, то случайная величина

$$\varphi_{s,t} = \left(\frac{1}{s} \sum_{i=1}^s \xi_i^2 \right) / \left(\frac{1}{m-s} \sum_{j=s+1}^m \xi_j^2 \right)$$

имеет распределение Снедекора–Фишера со степенями свободы s и $m-s$, см. Приложение. Результат не изменится, если $\xi_1, \dots, \xi_n$ нормальны $N(0, \sigma^2)$, и, более того, при условиях теоремы 2.1.1 имеет место **Следствие 2.1.2.** Пусть $\mathcal{R}^s$ и $\mathcal{R}^t$ — линейные подпространства $\mathcal{R}^n$, причем $\mathcal{R}^s$ ортогонально $\mathcal{R}^t$. Обозначим Π_s и Π_t операторы ортогонального проецирования на $\mathcal{R}^s$ и на $\mathcal{R}^t$ соответственно. Если выполнено условие теоремы 2.1.1, то

1) векторы $\Pi_s \xi$ и $\Pi_t \xi$ независимы, т.е. в любом ортонормированном базисе $\mathcal{R}^n$ совокупности случайных величин $(\Pi_s \xi)_1, \dots, (\Pi_s \xi)_n$ и $(\Pi_t \xi)_1, \dots, (\Pi_t \xi)_n$ (координат $\Pi_s \xi$ и $\Pi_t \xi$ соответственно) взаимно независимы;

2) случайная величина

$$\varphi_{s,t} = (\|\Pi_s \xi\|^2 / s) / (\|\Pi_t \xi\|^2 / t) \quad (1.12)$$

имеет распределение Снедекора–Фишера со степенями свободы s и t .

Доказательство. 1) Пусть $e_1, \dots, e_n$ ортонормированный базис $\mathcal{R}^n$, такой, что $e_1, \dots, e_s$ — ортонормированный базис $\mathcal{R}^s$ и $e_{s+1}, \dots, e_{s+t}$ — ортонормированный базис $\mathcal{R}^t$, $s+t \leq n$. Тогда, как было отмечено в следствии 2.1.1,

$$\Pi_s \xi = \sum_{j=1}^s (\xi, e_j) e_j, \quad \Pi_t \xi = \sum_{j=s+1}^{s+t} (\xi, e_j) e_j,$$

и, как следует из замечания 2.1.1, совокупности случайных величин $(\xi, e_1), \dots, (\xi, e_s)$ и $(\xi, e_{s+1}), \dots, (\xi, e_{s+t})$ нормальны и взаимно независимы. Отсюда, в частности, следует независимость случайных величин

$$\|\Pi_s \xi\|^2 = \sigma^2 \chi_s^2, \quad \|\Pi_t \xi\|^2 = \sigma^2 \chi_t^2, \quad (1.13)$$

а также тот факт, что $\|\Pi_s \xi\|^2 / \sigma^2$ и $\|\Pi_t \xi\|^2 / \sigma^2$ контролируются χ^2 -распределениями с s и t степенями свободы соответственно. В любой другой системе координат координаты случайного вектора $\Pi_s \xi$ выражаются лишь через $(\xi, e_1), \dots, (\xi, e_s)$, а координаты $\Pi_t \xi$ — через $(\xi, e_{s+1}), \dots, (\xi, e_{s+t})$. Отсюда следует независимость случайных векторов $\Pi_s \xi$ и $\Pi_t \xi$.

2) Так как случайные величины (1.13) независимы, то случайная величина (1.12) по определению имеет распределение Снедекора–Фишера со степенями свободы s и t . ■

Подчеркнем, что факты, приведенные в следствиях 2.1.1, 2.1.2, носят геометрический характер и не зависят от системы координат в $\mathcal{R}^n$.

В дальнейшем будут полезны следующие простые обобщения результатов, доказанных в следствиях. Очевидно, что как $\sigma^2 \chi_k^2$ распределена не только случайная величина $\|\Pi_k \xi\|^2$, но также и $\|\Pi_k(\xi + a)\|^2$, где a — любой (случайный или неслучайный) вектор, ортогональный $\mathcal{R}^k$. Дело в том, что в этом случае $\Pi_k a = 0$ (если a случайный вектор, но $\Pi_k a = 0$ с вероятностью единица). Аналогично утверждения следствия 2.1.2 справедливы не только для векторов $\Pi_s \xi$ и $\Pi_t \xi$, но и для любых векторов $\Pi_s(\xi + a)$ и $\Pi_t(\xi + b)$, если векторы a и b ортогональны $\mathcal{R}^s$ и $\mathcal{R}^t$ соответственно.

Следствие 2.1.3. Пусть Π_1 — оператор ортогонального проецирования на одномерное подпространство $\mathcal{R}^n$, натянутое на единичный вектор e , т.е. пусть $\Pi_1 \xi = (\xi, e)e$. Если случайный вектор ξ удовлетворяет условиям теоремы 2.1.1, то случайная величина

$$\tau_{n-1} = \frac{(\xi, e)}{[\|(I - \Pi_1)\xi\|^2 / (n-1)]^{1/2}} \quad (1.14)$$

имеет распределение Стьюдента с $n-1$ степенями свободы.

Доказательство. Напомним, что дробь $\tau_k = \eta / [\chi_k^2 / k]^{1/2}$ имеет распределение Стьюдента с k степенями свободы, если случайная величина η нормальна $N(0, 1)$ и не зависит от χ_k^2 , или, что то же самое, если η нормальна $N(0, \sigma^2)$ и $\tau_k = \eta / [\sigma^2 \chi_k^2 / k]^{1/2}$, см. Приложение.

Согласно теореме 2.1.1 случайная величина (ξ, e) имеет распределение $N(0, \sigma^2)$ и не зависит от $\|(I - \Pi_1)\xi\|^2$, как это показано в следствии 2.1.2, поскольку оператор $I - \Pi_1$ ортогонально проецирует на $n-1$ -мерное пространство векторов, ортогональных e . Отсюда же следует, что $\|(I - \Pi_1)\xi\|^2 = \sigma^2 \chi_{n-1}^2$, чем и завершается доказательство. ■

Пример 2.1.1. Пусть $e = \sum_{j=1}^n e_j / \sqrt{n}$, где $e_1, \dots, e_n$ — некоторый ортонормированный базис $\mathcal{R}^n$. Нетрудно проверить, что все координаты вектора $\Pi_1 \xi$ равны $(\Pi_1 \xi)_j = \frac{1}{n} \sum_{i=1}^n \xi_i$, $j = 1, \dots, n$, где $\xi_i = (\xi, e_i)$, $i = 1, \dots, n$, — координаты вектора ξ . Следовательно, вектор $(I - \Pi_1)\xi$ имеет координаты $((I - \Pi_1)\xi)_j = \xi_j - \frac{1}{n} \sum_{i=1}^n \xi_i$, $j = 1, \dots, n$.

Поэтому в данном случае случайная величина τ_{n-1} в (1.14) принимает вид

$$\tau_{n-1} = \frac{(\xi, e)}{[\|(I - \Pi_1)\xi\|^2 / (n-1)]^{1/2}} = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i}{\left[\frac{1}{n-1} \sum_{j=1}^n \left(\xi_j - \frac{1}{n} \sum_{i=1}^n \xi_i \right)^2 \right]^{1/2}}. \quad (1.15)$$

Если ξ_i , $i = 1, \dots, n$, независимы в совокупности и нормальны $N(\mu, \sigma^2)$, то, подставив в (1.15) вместо вектора ξ вектор $\xi - m$, где m — вектор, все координаты которого равны μ , найдем, что случайная величина

$$\tau_{n-1} = \frac{(\xi - m, e)}{[\|(I - \Pi_1)\xi\|^2 / (n-1)]^{1/2}} = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i - \mu)}{\left[\frac{1}{n-1} \sum_{j=1}^n \left(\xi_j - \frac{1}{n} \sum_{i=1}^n \xi_i \right)^2 \right]^{1/2}} \quad (1.16)$$

имеет *распределение Стьюдента с $n-1$ степенями свободы*. При этом использован тот факт, что $(I - \Pi_1)m = 0$.

2.2. Интервальное оценивание параметров нормального распределения

2.2.1. Интервальная оценка математического ожидания при известной дисперсии. Пусть ξ_j , $j = 1, \dots, n$, нормальны $N(\mu, \sigma^2)$, причем дисперсия σ^2 считается известной, но неизвестно математическое ожидание μ . Случайные величины $\xi_1, \dots, \xi_n$ предполагаются независимыми в совокупности. В этом случае последовательность $\xi_1, \dots, \xi_n$ называется *случайной выборкой из нормального распределения $N(\mu, \sigma^2)$ объема n* .

Так как случайная величина $\frac{1}{n} \sum_{j=1}^n \xi_j$ нормальна $N(\mu, \sigma^2/n)$, то случайная величина

$$\eta = \left(\frac{1}{n} \sum_{j=1}^n (\xi_j - \mu) \right) \frac{\sqrt{n}}{\sigma}, \quad (2.1)$$

нормальна $N(0, 1)$, и, следовательно, вероятность

$$P\{|\eta| < \varepsilon\} = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-x^2/2} dx = \Phi(\varepsilon) - \Phi(-\varepsilon),$$

где $\Phi(\cdot)$ — гауссовская функция распределения $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$, $-\infty < z < \infty$. Задав α , $0 < \alpha < 1$, и воспользовавшись таблицей нормального распределения, определим $\varepsilon > 0$, для которого

$$P\left(\left\{-\varepsilon < \left(\frac{1}{n} \sum_{j=1}^n \xi_j - \mu\right) \frac{\sqrt{n}}{\sigma} < \varepsilon\right\}\right) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 1 - \alpha. \quad (2.2)$$

При этом случайная величина η (2.1) может оказаться вне интервала $(-\varepsilon, \varepsilon)$ лишь с вероятностью α . Соотношение (2.2) можно записать иначе:

$$P\left(\left\{\frac{1}{n}\sum_{j=1}^n \xi_j - \frac{\varepsilon\sigma}{\sqrt{n}} < \mu < \frac{1}{n}\sum_{j=1}^n \xi_j + \frac{\varepsilon\sigma}{\sqrt{n}}\right\}\right) = 1 - \alpha. \quad (2.3)$$

Согласно (2.3) истинное значение математического ожидания μ покрывается случайным интервалом ¹⁾

$$\left(\frac{1}{n}\sum_{j=1}^n \xi_j - \frac{\varepsilon\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{j=1}^n \xi_j + \frac{\varepsilon\sigma}{\sqrt{n}}\right) \quad (2.4)$$

с вероятностью $1 - \alpha$. Вне этого интервала μ может оказаться с вероятностью α . Найденный случайный интервал (2.4) называется интервальной оценкой параметра μ , или $100(1 - \alpha)\%$ доверительным интервалом для параметра μ .

Примером задачи интервального оценивания является задача взвешивания со случайной $N(0, \sigma^2)$ ошибкой, μ — интервально оцениваемый вес.

Понятие интервальной оценки впервые встречается у Лапласа (1814) в связи с задачей определения параметра p биномиального распределения. Однако Лаплас рассматривал p как случайную величину. Правильную интерпретацию процедуры интервального оценивания, в которой речь идет о случайном доверительном интервале, предложил Е. Б. Уилсон в 1927 году.

2.2.2. Интервальная оценка дисперсии при известном математическом ожидании. Пусть в аналогичной ситуации известно математическое ожидание μ , но неизвестна дисперсия σ^2 . Тогда случайные величины $\xi_j - \mu$, $j = 1, \dots, n$, нормальны $N(0, \sigma^2)$ и независимы в совокупности. Следовательно, статистика ²⁾ $\sum_{j=1}^n (\xi_j - \mu)^2$ имеет распределение $\sigma^2 \chi_n^2$ с n степенями свободы.

Для заданных $\varepsilon_1 > 0$ и $\varepsilon_2 > \varepsilon_1$ по таблице χ^2 -распределения можно подсчитать вероятность $P\{\varepsilon_1 < \chi_n^2 < \varepsilon_2\} = 1 - \alpha$. Однако при заданном α , $0 < \alpha < 1$, величины ε_1 и ε_2 вычисляются неоднозначно. Поэтому обычно полагают $P\{\chi_n^2 \leq \varepsilon_1\} = P\{\chi_n^2 \geq \varepsilon_2\} = \alpha/2$ так, чтобы $P\{\varepsilon_1 <$

¹⁾ Случайным является центр интервала (2.3) $\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n \xi_j$ — «точечная оценка» μ , см. примеры 2.4.1, 2.4.4.

²⁾ В дальнейшем случайную величину, которая является функцией выборки $\xi_1, \dots, \xi_n$ условимся называть статистикой.

$\chi_n^2 < \varepsilon_2\} = 1 - P\{\chi_n^2 \geq \varepsilon_2\} - P\{\chi_n^2 \leq \varepsilon_1\} = 1 - \alpha$. Подсчитав таким образом ε_1 и ε_2 , получим

$$P\left(\left\{\varepsilon_1 < \frac{1}{\sigma^2} \sum_{j=1}^n (\xi_j - \mu)^2 < \varepsilon_2\right\}\right) = 1 - \alpha. \quad (2.5)$$

Соотношение (2.5) можно переписать в виде

$$P\left(\left\{\frac{1}{\varepsilon_2} \sum_{j=1}^n (\xi_j - \mu)^2 < \sigma^2 < \frac{1}{\varepsilon_1} \sum_{j=1}^n (\xi_j - \mu)^2\right\}\right) = 1 - \alpha,$$

или, если воспользоваться обозначениями предыдущего параграфа, в виде

$$P\left(\left\{\frac{1}{\varepsilon_2} \|\xi - m\|^2 < \sigma^2 < \frac{1}{\varepsilon_1} \|\xi - m\|^2\right\}\right) = 1 - \alpha. \quad (2.6)$$

Соотношение (2.6) показывает, что истинное значение дисперсии σ^2 покрывается случайным интервалом

$$\left(\frac{1}{\varepsilon_2} \|\xi - m\|^2, \frac{1}{\varepsilon_1} \|\xi - m\|^2\right) \quad (2.7)$$

с вероятностью $1 - \alpha$. Соответственно с вероятностью α дисперсия σ^2 может оказаться вне этого материала. Найденный случайный интервал (2.7) называется интервальной оценкой дисперсии σ^2 , или $100(1 - \alpha)\%$ доверительным интервалом для σ^2 .

Примером в данном случае является упомянутая задача взвешивания, в которой требуется интервально оценить дисперсию гауссовской ошибки взвешивания, μ — априори точно известный вес.

2.2.3. Интервальная оценка математического ожидания при неизвестной дисперсии. Как показано в следствии 2.1.3 теоремы 2.1.1, статистика τ_{n-1} (1.16) имеет распределение Стьюдента с $n - 1$ степенями свободы. По таблице распределения Стьюдента по заданному α , $0 < \alpha < 1$, можно определить ε' так, чтобы $P\{|\tau_{n-1}| < \varepsilon'\} = 1 - \alpha$. Иначе это можно переписать в виде

$$P\left(\left\{\frac{1}{n} \sum_{j=1}^n \xi_j - \frac{\varepsilon' \|(I - \Pi_1)\xi\|}{\sqrt{n(n-1)}} < \mu < \frac{1}{n} \sum_{j=1}^n \xi_j + \frac{\varepsilon' \|(I - \Pi_1)\xi\|}{\sqrt{n(n-1)}}\right\}\right) = 1 - \alpha, \quad (2.8)$$

где Π_1 — ортогональный проектор, определенный в примере 2.1.1. Соотношение (2.8) показывает, что истинное значение μ с вероятностью

стью $1 - \alpha$ покрывается случайным интервалом¹⁾

$$\left(\frac{1}{n} \sum_{j=1}^n \xi_j - \frac{\varepsilon' \|(I - \Pi_1)\xi\|}{\sqrt{n(n-1)}}, \frac{1}{n} \sum_{j=1}^n \xi_j + \frac{\varepsilon' \|(I - \Pi_1)\xi\|}{\sqrt{n(n-1)}} \right), \quad (2.9)$$

называемым *интервальной оценкой математического ожидания при неизвестной дисперсии*.

2.2.4. Интервальная оценка дисперсии при неизвестном математическом ожидании. Как отмечалось в следствии 2.1.3 теоремы 2.1.1,

$$\|(I - \Pi_1)(\xi - m)\|^2 = \|(I - \Pi_1)\xi\|^2 = \sigma^2 \chi_{n-1}^2. \quad (2.10)$$

Используя статистику (2.10) так же, как в параграфе 2.2.2 была использована статистика $\|\xi - m\|^2 = \sigma^2 \chi_n^2$, по заданному α , $0 < \alpha < 1$, определим ε_1 и ε_2 из требований

$$P(\{\chi_{n-1}^2 \leq \varepsilon_1'\}) = P(\{\chi_{n-1}^2 \geq \varepsilon_2'\}) = \alpha/2.$$

Тогда

$$P\left(\left\{\varepsilon_1' < \frac{1}{\sigma^2} \|(I - \Pi_1)\xi\|^2 < \varepsilon_2'\right\}\right) = 1 - \alpha,$$

или, что то же самое,

$$P\left(\left\{\frac{1}{\varepsilon_2} \|(I - \Pi_1)\xi\|^2 < \sigma^2 < \frac{1}{\varepsilon_1} \|(I - \Pi_1)\xi\|^2\right\}\right) = 1 - \alpha. \quad (2.11)$$

Случайный интервал

$$\left(\frac{1}{\varepsilon_2} \|(I - \Pi_1)\xi\|^2, \frac{1}{\varepsilon_1} \|(I - \Pi_1)\xi\|^2 \right) \quad (2.12)$$

называется *интервальной оценкой для дисперсии σ^2 при неизвестном математическом ожидании μ* .

В дальнейшем пространство $\mathcal{R}^n$ с некоторым фиксированным ортонормированным базисом будем называть выборочным пространством. При этом векторы из $\mathcal{R}^n$ будем задавать наборами их координат $(\xi_1, \dots, \xi_n)$, образующими выборки.

2.3. Общая задача интервального оценивания

Предположим, что $\xi_1, \dots, \xi_n$ — случайная выборка из распределения $F(x, \vartheta)$, $x \in \mathcal{R}^1$. Параметр ϑ функции распределения $F(\cdot, \vartheta)$ считается неизвестным. Для простоты будем считать, что ϑ изменя-

¹⁾ Случайны как центр $\hat{\mu} = \frac{1}{n} \sum_{j=1}^n \xi_j$ интервала (2.9), так и его полуширина $\frac{\varepsilon'}{\sqrt{n}} \cdot \frac{\|(I - \Pi_1)\xi\|}{\sqrt{n-1}}$, причем $\hat{\sigma}_n^2 = \frac{\|(I - \Pi_1)\xi\|^2}{n-1}$ — точечная оценка (4.6) дисперсии σ^2 в (2.4).

ется на действительной прямой. Пусть функции $\underline{\vartheta} = \underline{\vartheta}(x_1, \dots, x_n)$ и $\overline{\vartheta} = \overline{\vartheta}(x_1, \dots, x_n)$, $(x_1, \dots, x_n) \in \mathcal{R}^n$, таковы, что

$$P(\{\underline{\vartheta}(\xi_1, \dots, \xi_n) < \vartheta < \overline{\vartheta}(\xi_1, \dots, \xi_n)\}, \vartheta) = \gamma,$$

где $P(A, \vartheta)$ — вероятность события $(\xi_1, \dots, \xi_n) \in A \subset \mathcal{R}^n$, причем случайные величины $\xi_1, \dots, \xi_n$ взаимно независимы и имеют совместную функцию распределения $F(x_1, \vartheta) \cdot \dots \cdot F(x_n, \vartheta)$, $x_1, \dots, x_n \in \mathcal{R}^1$. В данном случае событие $A \subset \mathcal{R}^n$ состоит из тех $(\xi_1, \dots, \xi_n)$, для которых $\underline{\vartheta}(\xi_1, \dots, \xi_n) < \vartheta < \overline{\vartheta}(\xi_1, \dots, \xi_n)$.

Пара случайных величин $\underline{\vartheta}$, $\overline{\vartheta}$ называется интервальной оценкой параметра ϑ , или 100%-ным доверительным интервалом для ϑ . Случайные величины $\underline{\vartheta}$ и $\overline{\vartheta}$ называются доверительными границами, верхней и нижней соответственно. Как и в случае нормального распределения, $\underline{\vartheta}$ и $\overline{\vartheta}$ определяют случайный интервал $(\underline{\vartheta}, \overline{\vartheta})$, покрывающий истинное значение параметра ϑ с вероятностью γ .

Рассмотрим один из методов построения интервальных оценок.

Лемма 2.3.1. Пусть функция $g(x_1, \dots, x_n, \vartheta)$ определена для $-\infty < x_i < +\infty$, $i = 1, \dots, n$, и для любых фиксированных $x_1, \dots, x_n$ непрерывна и монотонна по $\vartheta \in \mathcal{R}^1$. Пусть, кроме того, $g(\xi_1, \dots, \xi_n, \vartheta)$ является случайной величиной, функция распределения которой не зависит от ϑ , если $F(x_1, \vartheta) \cdot \dots \cdot F(x_n, \vartheta)$, $-\infty < x_i < \infty$, $i = 1, \dots, n$, — функция распределения $\xi_1, \dots, \xi_n$. Если (g_1, g_2) — интервал, для которого

$$P(\{g_1 < g(\xi_1, \dots, \xi_n, \vartheta) < g_2\}, \vartheta) = \gamma,$$

то

$$P(\{\underline{\vartheta} < \vartheta < \overline{\vartheta}\}, \vartheta) = \gamma, \quad (3.1)$$

где $\underline{\vartheta} < \overline{\vartheta}$ — решения относительно ϑ уравнений $g(\xi_1, \dots, \xi_n, \vartheta) = g_1, g_2$.

Доказательство. Для определенности будем считать, что $g(\cdot, \vartheta)$ монотонно возрастает по ϑ . Поскольку распределение $\zeta = g(\xi_1, \dots, \xi_n, \vartheta)$ от ϑ не зависит, то для всякого γ , $0 < \gamma < 1$, и ϑ_0 , $-\infty < \vartheta_0 < +\infty$, можно указать g_1 и g_2 , $g_1 < g_2$, так чтобы

$$P(\{g_1 < \zeta < g_2\}, \vartheta_0) = \gamma, \quad (3.2)$$

причем g_1 и g_2 можно указать (вообще говоря не единственным способом), если задано лишь γ , а истинное значение ϑ_0 параметра ϑ неизвестно. Неравенства

$$g_1 < g(\xi_1, \dots, \xi_n, \vartheta_0) < g_2 \quad (3.3)$$

для каждой выборки $\xi_1, \dots, \xi_n$ из распределения $F(\cdot, \vartheta_0)$ эквивалентны неравенствам

$$\underline{\vartheta}(\xi_1, \dots, \xi_n) < \vartheta_0 < \overline{\vartheta}(\xi_1, \dots, \xi_n), \quad (3.4)$$

где $\underline{\vartheta}(\xi_1, \dots, \xi_n)$ и $\overline{\vartheta}(\xi_1, \dots, \xi_n)$ — решения относительно ϑ уравнений $g(\xi_1, \dots, \xi_n, \vartheta) = g_1$ и $g(\xi_1, \dots, \xi_n, \vartheta) = g_2$ соответственно. Поэтому события (3.3) и (3.4) совпадают. Следовательно,

$$P(\{\underline{\vartheta}(\xi_1, \dots, \xi_n) < \vartheta_0 < \overline{\vartheta}(\xi_1, \dots, \xi_n)\}, \vartheta_0) = \gamma. \blacksquare$$

Однако всегда ли можно найти функцию $g(\xi_1, \dots, \xi_n, \vartheta)$ с указанными в лемме свойствами? Если функция распределения $F(x, \vartheta)$ непрерывна по x , $-\infty < x < +\infty$, для каждого ϑ , $-\infty < \vartheta < +\infty$, непрерывна и монотонна по ϑ , $-\infty < \vartheta < +\infty$, для каждого x , $-\infty < x < +\infty$, то ответ на поставленный вопрос утвердительный. Для доказательства заметим вначале, что для всякой случайной величины ξ с *непрерывной* функцией распределения $F(\cdot)$ случайная величина $\eta = F(\xi)$ равномерно распределена на $[0, 1]$. Чтобы не прерывать изложение, этот факт докажем в последнюю очередь.

Рассмотрим функцию

$$g(\xi_1, \dots, \xi_n, \vartheta) = F(\xi_1, \vartheta) \cdot \dots \cdot F(\xi_n, \vartheta).$$

Согласно замечанию ее распределение не зависит от параметра ϑ , если $F(\cdot, \vartheta)$ — функция распределения ξ_i , $i = 1, \dots, n$, и случайные величины $\xi_1, \dots, \xi_n$ независимы в совокупности. Кроме того, по условию $g(\xi_1, \dots, \xi_n, \vartheta)$ монотонна и непрерывна по ϑ . Следовательно, $g(\cdot)$ удовлетворяет условиям леммы.

Докажем теперь, что *случайная величина $\eta = F(\xi)$, где $F(\cdot)$ — непрерывная функция распределения ξ , равномерно распределена на $[0, 1]$* . Действительно,

$$F(x) = P(\{\xi < x\}) = P(\{\xi \leq x\}), \quad -\infty < x < +\infty, \quad (3.5)$$

поскольку функция распределения $F(\cdot)$ непрерывна, и, следовательно, $P\{\xi = x\} = 0$. Событие $\xi \leq x$, очевидно, влечет событие $F(\xi) \leq F(x)$, поэтому в согласии с (3.5)

$$F(x) = P(\{\xi \leq x\}) \leq P(\{F(\xi) \leq F(x)\}). \quad (3.6)$$

С другой стороны,

$$\{F(\xi) \leq F(x)\} = \{\xi \leq x\} \cup \{F(\xi) = F(x)\},$$

и

$$P(\{F(\xi) \leq F(x)\}) \leq P(\{\xi \leq x\}) + P(\{F(\xi) = F(x)\}) = P(\{\xi \leq x\}), \quad (3.7)$$

так как функция распределения $F(\cdot)$ непрерывна и, следовательно, $P(\{F(\xi) = \text{const}\}) = 0$. Из (3.6) и (3.7) получаем $P(\{F(\xi) < F(x)\}) =$

$= P(\{F(\xi) \leq F(x)\}) = P(\{\xi < x\}) = F(x)$, $-\infty < x < +\infty$, что иначе можно записать следующим образом:

$$P(\{\eta < z\}) = \begin{cases} 0, & z < 0, \\ z, & 0 \leq z \leq 1, \\ 1, & z > 1, \end{cases} \quad \eta = F(\xi).$$

Равенства (3.5) означают, что $\eta = F(\xi)$ равномерно распределена на $[0, 1]$.

В заключение заметим, что если $g(\xi_1, \dots, \xi_n, \vartheta)$ обладает всеми свойствами, перечисленными в лемме, кроме монотонности по ϑ , то неравенства (3.3), будучи разрешенными относительно ϑ , эквивалентны включению

$$\vartheta_0 \in A(\xi_1, \dots, \xi_n),$$

где случайное множество $A(\xi_1, \dots, \xi_n)$ определяется выборкой и в данном случае может и не быть интервалом. Вполне аналогично рассмотренному случаю равенство (3.2) эквивалентно равенству

$$P(\{\vartheta_0 \in A(\xi_1, \dots, \xi_n) | \vartheta_0\}) = \gamma.$$

Множество $A(\xi_1, \dots, \xi_n)$ называется $100\gamma\%$ доверительным множеством для ϑ , или, иначе, доверительным множеством уровня γ .

Пример 2.3.1. Пусть $\xi_1, \dots, \xi_n$ — выборка из распределения $N(\mu, \sigma^2)$, дисперсия σ^2 известна. Положим $g(x_1, \dots, x_n, \mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right\}$, $(x_1, \dots, x_n) \in \mathcal{R}^n$. В данном случае свойство монотонности g по μ не выполнено. Поскольку статистика

$$-2 \ln g(\xi_1, \dots, \xi_n, \mu) - 2n \ln(\sqrt{2\pi}\sigma) = \frac{1}{\sigma^2} \sum_{j=1}^n (\xi_j - \mu)^2$$

имеет распределение χ^2 с n степенями свободы, определим по таблице распределения χ_n^2 постоянную c из условия $P\{\chi_n^2 < c\} = \gamma$. Тогда неравенство $\frac{1}{\sigma^2} \sum_{j=1}^n (\xi_j - \mu)^2 < c$ определяет доверительное множество для μ уровня γ .

2.4. Точечные оценки

Рассмотрим типичную ситуацию, в которой возникают задачи точечного оценивания. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайная выборка из распределения $F(x, \vartheta)$, $-\infty < x < +\infty$, зависящего от параметра $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Theta$ и $\tau(\vartheta)$ — известная функция, определенная на Θ . Требуется вычислить значение $\tau(\vartheta)$, но аргумент $\vartheta \in \Theta$ неизвестен.

В этом случае единственная доступная наблюдению информация о $\tau(\vartheta)$ заключена в выборке $\xi = (\xi_1, \dots, \xi_n)$ (и в функции $\tau(\cdot)$), и лучшее,

что можно сделать в такой ситуации, это построить так называемую оценку значения $\tau(\vartheta)$, основанную на выборке ξ . Иными словами, речь идет о построении функции $t(\cdot)$, определенной на выборочном пространстве $\mathcal{R}^n$, с целью использовать статистику $t(\xi)$ вместо $\tau(\vartheta)$. Статистика $t(\xi)$ называется (*точечной*) *оценкой* $\tau(\vartheta)$, и при этом, естественно, предполагается, что распределение $t(\xi)$ в известном смысле концентрируется около значения $\tau(\vartheta)$, где ϑ — параметр распределения выборки ξ .

Прежде чем сформулировать условия, определяющие оценку, рассмотрим несколько характерных примеров.

Пример 2.4.1. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка из нормального распределения $N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, $\Theta = \{-\infty < \mu < +\infty, 0 < \sigma^2\}$. Рассмотрим статистику

$$\hat{\mu}_n = t(\xi) = \frac{1}{n} \sum_{j=1}^n \xi_j. \quad (4.1)$$

Так как ¹⁾

$$M_{\vartheta} \hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n M_{\vartheta} \xi_j = \mu, \quad (4.2)$$

то, используя $\hat{\mu}_n$ в качестве оценки математического ожидания $\mu = \tau(\vartheta)$, мы не будем совершать *систематической ошибки* в том смысле, что $M_{\vartheta}(\hat{\mu}_n - \mu) = 0$. Оценки, обладающие свойством (4.2), называются *несмещенными*.

Поскольку

$$M_{\vartheta}(\hat{\mu}_n - \mu)^2 = D_{\vartheta} \hat{\mu}_n = \frac{1}{n^2} \sum_{j=1}^n D_{\vartheta} \xi_j = \frac{1}{n} \sigma^2, \quad (4.3)$$

то согласно лемме 1.10.1 $\hat{\mu}_n$ при $n \rightarrow \infty$ сходится по вероятности к μ . Это означает, что распределение $\hat{\mu}_n$ при $n \rightarrow \infty$ концентрируется около μ , поскольку при $n \rightarrow \infty$ вероятность любого отклонения $\hat{\mu}_n$ от μ стремится к нулю:

$$P\{|\hat{\mu}_n - \mu| > \varepsilon, \vartheta\} < \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty, \quad (4.4)$$

для любого $\varepsilon > 0$.

Оценки, обладающие свойством (4.4), т. е. сходящиеся по вероятности к оцениваемому параметру (функции параметра) при объеме n выборки, стремящемся к бесконечности, называются *состоятельными*. Это, безусловно, желательное свойство оценок, но оно не определяет качества оценки при фиксированном n , если качество определить, на-

¹⁾ M_{ϑ} и D_{ϑ} — обозначения математического ожидания и дисперсии, отвечающих распределению $F(\cdot, \vartheta)$.

пример, как величину уклонения $\hat{\mu}_n$ от μ и задать числом $M_{\vartheta}(\hat{\mu}_n - \mu)^2$. Разумеется, чем меньше $M_{\vartheta}(\hat{\mu}_n - \mu)^2$, тем лучше оценка. В данном случае речь идет о несмещенной оценке, и ее качество определяется величиной дисперсии. Поэтому наилучшую оценку μ естественно определить как оценку с минимальной дисперсией среди всех несмещенных оценок. Далее мы покажем, что в этом смысле $\hat{\mu}_n$ (4.1) действительно наилучшая, см. пример 2.4.4.

Какова в данном случае роль несмещенности? Пусть на основании независимых выборок $\xi^1 = (\xi_1^1, \dots, \xi_n^1)$, $\xi^2 = (\xi_1^2, \dots, \xi_n^2), \dots, \xi^k = (\xi_1^k, \dots, \xi_n^k)$ из распределения $F(\cdot, \vartheta)$ построены оценки $t_1(\xi^1), t_2(\xi^2), \dots, t_k(\xi^k)$ значения $\tau(\vartheta)$, причем $M_{\vartheta}t_i(\xi^i) = \tau(\vartheta) + \varepsilon_i$, $D_{\vartheta}t_i(\xi^i) = \sigma_i^2 \leq \sigma^2$, $i = 1, 2, \dots, k$. Тогда для статистики $T_k = \frac{1}{k} \sum_{j=1}^k t_j(\xi^j)$ найдем $M_{\vartheta}T_k = \frac{1}{k} \sum_{j=1}^k M_{\vartheta}t_j(\xi^j) = \tau(\vartheta) + \frac{1}{k} \sum_{j=1}^k \varepsilon_j$ и $D_{\vartheta}T_k = \frac{1}{k^2} \sum_{j=1}^k \sigma_j^2 \leq \sigma^2/k \rightarrow 0$ при $k \rightarrow \infty$. Если смещение $\frac{1}{k} \sum_{j=1}^k \varepsilon_j$ не стремится к нулю при $k \rightarrow \infty$, то T_k при $k \rightarrow \infty$ не сходится к истинному значению $\tau(\vartheta)$. Например, при $\varepsilon_j = \varepsilon$, $j = 1, 2, \dots, k$, T_k сходится по вероятности к $\tau(\vartheta) + \varepsilon$, что, разумеется, нежелательно. С другой стороны, если $\varepsilon_j = 0$, $j = 1, 2, \dots, k$, то можно говорить о суммировании информации о $\tau(\vartheta)$, содержащейся в независимых несмещенных оценках $t_j(\xi^j)$, $j = 1, 2, \dots, k$.

Вместе с тем, очевидно, что если речь идет об оценке $\overset{\circ}{t}$ с минимальным уклонением

$$M_{\vartheta}(\overset{\circ}{t} - \tau(\vartheta))^2 = \min_t M_{\vartheta}(t - \tau(\vartheta))^2, \quad (4.5)$$

то при прочих равных условиях *дополнительное требование несмещенности может лишь увеличить уклонение (4.5)*, так как условие $M_{\vartheta}t = \tau(\vartheta)$ сужает класс оценок, на котором ищется минимум в (4.5). Более того, *класс несмещенных оценок может оказаться даже пустым*.

Пример 2.4.2. Пусть ξ — число успехов в последовательности n испытаний Бернулли с неизвестной вероятностью успеха $p = \vartheta$, $0 < \vartheta \leq 1$, и требуется несмещенно оценить $\tau(\vartheta) = \vartheta^{-1}$. Для всякой такой оценки $t(\xi)$ должно выполняться равенство $M_{\vartheta}t(\xi) = \sum_{i=0}^n C_n^i \vartheta^i (1 - \vartheta)^{n-i} t(i) = \vartheta^{-1}$ для всех $\vartheta \in (0, 1]$. Но, это, очевидно, невозможно, так как для любой функции $t(\cdot)$ при $\vartheta \rightarrow 0$ $M_{\vartheta}t(\xi) \rightarrow C_n^0 t(0)$, а $\vartheta^{-1} \rightarrow \infty$.

Пример 2.4.3. Рассмотрим статистику

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (\xi_j - \hat{\mu}_n)^2 = \frac{\|(I - \Pi_1)\xi\|^2}{n-1}, \quad (4.6)$$

где $\xi = (\xi_1, \dots, \xi_n)$ — выборка из $N(\mu, \sigma^2)$, $\hat{\mu}_n$ определено в (4.1) и Π_1 — оператор ортогонального проецирования в $\mathcal{R}^n$, введенный в примере 2.1.1. Согласно следствию 2.1.1 теоремы 2.1.1 статистика (4.6) имеет распределение $(\sigma^2/(n-1))\chi_{n-1}^2$. Поэтому, см. Приложение, $M_\vartheta \hat{\sigma}_n^2 = \frac{\sigma^2}{n-1} M\chi_{n-1}^2 = \sigma^2$, $D_\vartheta \hat{\sigma}_n^2 = \frac{\sigma^4}{(n-1)^2} D_\vartheta \chi_{n-1}^2 = \frac{2\sigma^4}{n-1} \rightarrow 0$, $n \rightarrow \infty$. Следовательно, статистика $\hat{\sigma}_n^2$ является несмещенной состоятельной оценкой $\sigma^2 = \tau(\vartheta)$, $\vartheta = (\mu, \sigma^2)$.

Однако $M_\vartheta \left(\frac{1}{n} \|(I - \Pi_1)\xi\|^2 - \sigma^2 \right)^2 = \frac{\sigma^4}{n^2} [2(n-1) + (n-1)^2] - 2\frac{\sigma^4}{n}(n-1) + \sigma^4 = \left(\frac{2}{n} - \frac{1}{n^2} \right) \sigma^4 < D_\vartheta \hat{\sigma}_n^2$, $n = 1, 2, \dots$. Поэтому для каждого $n = 1, 2, \dots$ статистика $\tilde{\sigma}_n^2 = \frac{1}{n} \|(I - \Pi_1)\xi\|^2$ меньше уклоняется от σ^2 , чем $\hat{\sigma}_n^2$, хотя и имеет смещение, ибо $M_\vartheta \tilde{\sigma}_n^2 = \sigma^2 - \sigma^2/n$. Так как $M_\vartheta (\tilde{\sigma}_n^2 - \sigma^2) = \sigma^2/n \rightarrow 0$, $n \rightarrow \infty$, $D_\vartheta (\tilde{\sigma}_n^2 - \sigma^2) = M_\vartheta (\tilde{\sigma}_n^2 - \sigma^2)^2 - \frac{\sigma^4}{n^2} = \frac{2n-2}{n^2} \sigma^4 \rightarrow 0$, $n \rightarrow \infty$, то оценка $\tilde{\sigma}_n^2$ при $n \rightarrow \infty$ сходится по вероятности к σ^2 , т. е. состоятельна. В данном случае смещенная оценка $\tilde{\sigma}_n^2$ дисперсии σ^2 оказывается предпочтительнее несмещенной $\hat{\sigma}_n^2$. Впрочем, заметим, что статистика $\check{\sigma}_n^2 = \frac{\|(I - \Pi_1)\xi\|^2}{n+1}$ обладает еще меньшим отклонением от σ^2 , более того, минимальным среди всех статистик вида $\text{const} \|(I - \Pi_1)\xi\|^2$, $M_\vartheta \left(\frac{1}{n+1} \|(I - \Pi_1)\xi\|^2 - \sigma^2 \right)^2 = \frac{2\sigma^4}{n+1}$.

Статистики $\hat{\mu}_n$ и $\hat{\sigma}_n^2$ как оценки для математического ожидания и дисперсии сохраняют ряд свойств и без предположения о нормальности выборки. Действительно, если $\xi_1, \dots, \xi_n$ — выборка из произвольного распределения, удовлетворяющего условиям

$$M\xi_j = \mu, \quad D\xi_j = \sigma^2, \quad j = 1, 2, \dots, n,$$

то согласно (4.2), (4.3)

$$M\hat{\mu}_n = \mu, \quad D\hat{\mu}_n = \sigma^2/n,$$

и тем самым $\hat{\mu}_n$ — несмещенная и состоятельная оценка μ . Так как согласно определению оператора проецирования Π_1 в примере 2.1.1 $\|(I - \Pi_1)\xi\|^2 = \|(\xi - m) - \Pi_1(\xi - m)\|^2 = \|\xi - m\|^2 - \|\Pi_1(\xi - m)\|^2$

$$\| - m \|^2 = \sum_{j=1}^n (\xi_j - \mu)^2 - n \left(\frac{1}{n} \sum_{j=1}^n (\xi_j - \mu) \right)^2, \text{ где } m = (\mu, \dots, \mu), \text{ то}$$

$$M \| (I - \Pi_1) \xi \|^2 = n\sigma^2 - \sigma^2.$$

Следовательно, и без предположения о нормальности статистика $\hat{\sigma}_n^2$ является несмещенной оценкой для σ^2 . Однако заключения о состоятельности $\hat{\sigma}_n^2$ без дополнительных предположений о распределении на сей раз сделать не удастся. Что касается оценок $\tilde{\sigma}_n^2$ и $\check{\sigma}_n^2$, то в данном случае они утрачивают преимущества перед $\hat{\sigma}_n^2$.

Итак, для дальнейшего обсуждения мы выделяем следующие свойства оценок:

1) *несмещенность*: $M_{\vartheta} t(\xi) = \tau(\vartheta)$;

2) *минимальность уклонения*: $M_{\vartheta} (t(\xi) - \tau(\vartheta))^2 \sim \min$ или, в частности, минимальность дисперсии, если речь идет о несмещенных оценках;

3) *Состоятельность*: $t_n(\xi) \xrightarrow{P} \tau(\vartheta)$, $n \rightarrow \infty$.

Каждое из этих условий является ограничением на $\tau(\xi)$, и далеко не всегда можно гарантировать существование оценки с такими свойствами.

2.4.1. Несмещенные оценки минимальной дисперсии. Как было отмечено, не всегда можно удовлетворить требованию несмещенности. Однако если существует несмещенная оценка минимальной дисперсии (Н.О.М.Д.), то она единственная.

Лемма 2.4.1. Пусть t_1 и t_2 — Н.О.М.Д. Тогда $t_1 = t_2$ с вероятностью единица.

Доказательство. Обозначим $M_{\vartheta} t_1 = M_{\vartheta} t_2 = \tau(\vartheta)$. Тогда и для оценки $t_3 = 1/2(t_1 + t_2)$: $M_{\vartheta} t_3 = \tau(\vartheta)$. Кроме того, в силу неравенства Коши–Буняковского,

$$\begin{aligned} D_{\vartheta} t_3 &= 1/4(D_{\vartheta} t_1 + D_{\vartheta} t_2 + 2 \operatorname{cov} t_1 t_2) \leq \\ &\leq 1/4(D_{\vartheta} t_1 + D_{\vartheta} t_2 + 2(D_{\vartheta} t_1 D_{\vartheta} t_2)^{1/2}) = \\ &= 1/4((D_{\vartheta} t_1)^{1/2} + (D_{\vartheta} t_2)^{1/2})^2. \end{aligned} \quad (4.7)$$

Равенство в (4.7) выполняется тогда и только тогда, когда с вероятностью единица $t_1 - \tau(\vartheta) = k(\vartheta)(t_2 - \tau(\vartheta))$, $k(\vartheta) \geq 0$. Обозначим $D_{\vartheta} t_1 = D_{\vartheta} t_2 = \delta$. Из (4.7) следует, что $D_{\vartheta} t_3 \leq 1/4(\delta + \delta + 2\delta) = \delta$. Поскольку t_1 и t_2 — оценки с минимальной дисперсией, должно быть $D_{\vartheta} t_3 \geq \delta$, следовательно, $D_{\vartheta} t_3 = \delta$. Поэтому в (4.7) выполнено равенство и, следовательно, с вероятностью единица $t_1 - \tau(\vartheta) = k(\vartheta)(t_2 - \tau(\vartheta))$. А так как $M_{\vartheta}(t_1 - \tau(\vartheta))(t_2 - \tau(\vartheta)) = k(\vartheta) \cdot \delta = \delta$, то $k(\vartheta) = 1$. ■

Далее статистики, имеющие конечный момент второго (а следовательно, и первого) порядка, будем называть гильбертовыми¹⁾.

¹⁾ Гильбертовы обычно комплексные статистики (случайные величины), мы будем называть так любые статистики с конечным моментом второго порядка.

Теорема 2.4.1. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка из распределения $F(\cdot, \vartheta)$, $\vartheta \in \Theta$. Для того чтобы гильбертова статистика $t(\xi)$ была Н.О.М.Д. для $\tau(\vartheta)$, необходимо и достаточно, чтобы для всякой гильбертовой статистики $\eta = s(\xi)$, такой, что $M_\vartheta \eta = 0$,

$$M_\vartheta t(\xi) \eta = 0. \quad (4.8)$$

Доказательство. Пусть $t(\xi)$ — гильбертова несмещенная оценка для $\tau(\vartheta)$. Если гильбертова статистика $\eta = s(\xi)$ такова, что $M_\vartheta \eta = 0$, то $t(\xi) + \lambda \eta$ также несмещенная гильбертова оценка $\tau(\vartheta)$ для всех λ . При этом $\varphi_\lambda = M_\vartheta(t(\xi) - \tau(\vartheta) + \lambda \eta)^2 = M_\vartheta(t(\xi) - \tau(\vartheta))^2 + 2\lambda M_\vartheta(t(\xi) - \tau(\vartheta))\eta + \lambda^2 M_\vartheta \eta^2$, и, дифференцируя по λ , найдем $\min_\lambda \varphi_\lambda = M_\vartheta(t(\xi) - \tau(\vartheta))^2 + \frac{[M_\vartheta(t(\xi) - \tau(\vartheta))\eta]^2}{M_\vartheta \eta^2}$ при $\lambda = -\frac{M_\vartheta(t(\xi) - \tau(\vartheta))\eta}{M_\vartheta \eta^2}$.

Пусть $t(\xi)$ — Н.О.М.Д. для $\tau(\vartheta)$. Тогда для всякой гильбертовой статистики $\eta = s(\xi)$, $M_\vartheta \eta = 0$: $M_\vartheta(t(\xi) - \tau(\vartheta))^2 \leq \min_\lambda M_\vartheta(t(\xi) - \tau(\vartheta) + \lambda \eta)^2 = M_\vartheta(t(\xi) - \tau(\vartheta))^2 - \frac{[M_\vartheta(t(\xi) - \tau(\vartheta))\eta]^2}{M_\vartheta \eta^2}$. Отсюда следует (4.8): $0 = M_\vartheta(t(\xi) - \tau(\vartheta))\eta = M_\vartheta t(\xi) \eta$. Наоборот, если условие (4.8) выполнено для всякой гильбертовой статистики $\eta = s(\xi)$, $M_\vartheta \eta = 0$, то $M_\vartheta(t(\xi) + \lambda \eta - \tau(\vartheta))^2 \geq \min_\lambda M_\vartheta(t(\xi) + \lambda \eta - \tau(\vartheta))^2 = M_\vartheta(t(\xi) - \tau(\vartheta))^2$. Следовательно, $t(\xi)$ — Н.О.М.Д. для $\tau(\vartheta)$, так как в виде $t(\xi) + \lambda \eta$ можно представить любую гильбертову несмещенную оценку $\tau(\vartheta)$. ■

Замечание 2.4.1. Множество гильбертовых случайных величин, т.е. случайных величин с конечными моментами второго порядка, образует линейное пространство. Действительно, если $M\zeta^2 < \infty$, $M\eta^2 < \infty$, то и $M(a\zeta + b\eta)^2 = a^2 M\zeta^2 + 2abM\zeta\eta + b^2 M\eta^2 \leq a^2 M\zeta^2 + 2|ab|(M\zeta^2 M\eta^2)^{1/2} + b^2 M\eta^2 < \infty$. Это линейное пространство превращается в гильбертово пространство, если в нем задать скалярное произведение $(\eta, \zeta) = M\eta\zeta$, а элементами считать классы равных с вероятностью единица случайных величин. Норма случайной величины η определяется как $(\eta, \eta)^{1/2}$ и отождествляется с нормой класса случайных величин, совпадающих с η с вероятностью единица.

Возвращаясь к доказанной теореме, обозначим $\mathcal{H}_\vartheta$ гильбертово пространство гильбертовых статистик $\zeta = t(\xi)$, $M_\vartheta \zeta^2 < \infty$. Подмножество $\mathcal{H}_\vartheta^0$ статистик из $\mathcal{H}_\vartheta$, для которых $M_\vartheta \eta = 0$, является линейным подпространством $\mathcal{H}_\vartheta$. Рассмотрим гиперплоскость $\mathcal{H}_\vartheta^0 + t(\xi)$, где $t(\xi) \in \mathcal{H}_\vartheta$ и $M_\vartheta t(\xi) = \tau(\vartheta)$, т.е. множество всех статистик вида $t(\xi) + \eta$, $\eta \in \mathcal{H}_\vartheta^0$. Н.О.М.Д. для $\tau(\vartheta)$ есть элемент $t(\xi) + \mathcal{H}_\vartheta^0$ с минимальной нормой, т.е. ортогональная проекция $t^0(\xi)$ нуля на $t(\xi) + \mathcal{H}_\vartheta^0$, определяемая условием: $(t^0(\xi), \eta) = 0$ для любой статистики $\eta \in \mathcal{H}_\vartheta^0$. Таков геометрический смысл теоремы 2.4.1, см. рис. 2.1.

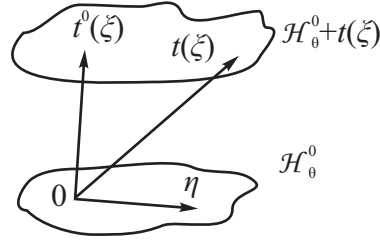


Рис. 2.1. Гиперплоскость $\mathcal{H}_\vartheta^0 = \{\eta, M_\vartheta \eta^2 < \infty, M_\vartheta \eta = 0\}$ и ортогональная проекция нуля на $\mathcal{H}_\vartheta^0 + t(\xi)$.

2.4.2. Эффективные оценки. При некоторых предположениях о функциях распределения $F(\cdot, \vartheta)$, $\vartheta \in \Theta$, может быть получена удобная априорная оценка качества оценивания, известная как *неравенство Рао–Крамера*. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайная выборка объема n из распределения с плотностью вероятности $f(x, \vartheta)$, $x \in \mathcal{R}^1$, $\vartheta \in \Theta \subset \mathcal{R}^k$. Поскольку ξ_j , $j = 1, \dots, n$, независимы в совокупности, плотность $L(x, \vartheta)$, $x \in \mathcal{R}^n$, $\vartheta \in \Theta$, совместного распределения выборки $\xi_1, \dots, \xi_n$ при каждом $\vartheta \in \Theta$ определяется равенством

$$L(x, \vartheta) = f(x_1, \vartheta) \dots f(x_n, \vartheta), \quad x \in \mathcal{R}^n, \quad x_1, \dots, x_n \in \mathcal{R}^1.$$

В статистических задачах $L(x, \vartheta)$ рассматривается как функция двух аргументов $x \in \mathcal{R}^n$, $\vartheta \in \Theta$ и, как функция $\vartheta \in \Theta$, называется *функцией правдоподобия*.

Если выборку ξ_i , $i = 1, \dots, n$, образуют дискретные случайные величины, то функция правдоподобия определяется как произведение вероятностей

$$L(x, \vartheta) = P(\{\xi_1 = x_1\}, \vartheta) \dots P(\{\xi_n = x_n\}, \vartheta), \quad \vartheta \in \Theta, \quad x = (x_1, \dots, x_n).$$

Теорема 2.4.2. (Рао–Крамер). Пусть $\Theta = \mathcal{R}^1$,

1) статистика $t(\xi)$, $\xi = (\xi_1, \dots, \xi_n)$, является несмещенной оценкой $\tau(\vartheta)$: $M_\vartheta t(\xi) = \tau(\vartheta)$, $\vartheta \in \Theta$;

2) функции $L(x, \vartheta)$ при каждом $x \in \mathcal{R}^n$ и $\tau(\vartheta)$ дифференцируемы по $\vartheta \in \Theta$;

3) множество тех $x \in \mathcal{R}^n$, при которых $L(x, \vartheta) > 0$, не зависит от $\vartheta \in \Theta$ и $\frac{d}{d\vartheta} \int_{\mathcal{R}^n} L(x, \vartheta) dx = \int_{\mathcal{R}^n} \frac{\partial L(x, \vartheta)}{\partial \vartheta} dx$, $\frac{d}{d\vartheta} \int_{\mathcal{R}^n} t(x) L(x, \vartheta) dx = \int_{\mathcal{R}^n} t(x) \frac{\partial L(x, \vartheta)}{\partial \vartheta} dx$. Тогда

$$D_\vartheta t(\xi) \geq \frac{[\tau'(\vartheta)]^2}{M_\vartheta \left(\frac{\partial \ln L(\xi, \vartheta)}{\partial \vartheta} \right)^2}, \quad \vartheta \in \Theta, \quad (4.9)$$

причем знак равенства в (4.9) имеет место тогда и только тогда, когда

$$\frac{\partial \ln L(\xi, \vartheta)}{\partial \vartheta} = a(\vartheta)(t(\xi) - \tau(\vartheta)) \quad (4.10)$$

с вероятностью единица для некоторой функции $a(\vartheta)$, $\vartheta \in \Theta$.

Доказательство. Дифференцируя равенства $\int L(x, \vartheta) dx = 1$,

$$\int t(x)L(x, \vartheta) dx = \tau(\vartheta) \text{ по } \vartheta, \text{ найдем } \int \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} L(x, \vartheta) dx = 0,$$

$$\int t(x) \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} L(x, \vartheta) dx = \tau'(\vartheta), \vartheta \in \mathcal{R}^1. \text{ Отсюда следует, что}$$

$$\tau'(\vartheta) = \int (t(x) - \tau(\vartheta)) \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} L(x, \vartheta) dx, \vartheta \in \mathcal{R}^1, \quad (4.11)$$

и неравенство (4.9) получается из (4.11) как неравенство Коши-Буняковского

$$[\tau'(\vartheta)]^2 \leq \int (t(x) - \tau(\vartheta))^2 L(x, \vartheta) dx \int \left(\frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} \right)^2 L(x, \vartheta) dx. \quad (4.12)$$

Равенство в (4.12), а следовательно, и в (4.9), имеет место лишь тогда, когда для некоторого множителя пропорциональности $a(\vartheta)$

$$\sqrt{L(x, \vartheta)} \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} = \sqrt{L(x, \vartheta)} a(\vartheta)(t(x) - \tau(\vartheta)).$$

Это условие эквивалентно равенству (4.10), верному с вероятностью единица, поскольку множество тех ξ , для которых $L(\xi, \vartheta) = 0$, имеет вероятность, равную нулю. ■

Определение 2.4.1. Несмещенная оценка $t(\xi)$, $M_{\vartheta}t(\xi) = \tau(\vartheta)$, называется эффективной, если

$$D_{\vartheta}t(\xi) = [\tau'(\vartheta)]^2 / M_{\vartheta} \left(\frac{\partial \ln L(\xi, \vartheta)}{\partial \vartheta} \right)^2.$$

В условиях теоремы 2.4.2 для того, чтобы оценка $t(\xi)$ была эффективной, необходимо и достаточно, чтобы она удовлетворяла условию (4.10). Так как согласно (4.10) $M_{\vartheta} \left(\frac{\partial \ln L(\xi, \vartheta)}{\partial \vartheta} \right) = a^2(\vartheta) D_{\vartheta}t(\xi)$, то для эффективной оценки

$$D_{\vartheta}t(\xi) = \left| \frac{\tau'(\vartheta)}{a(\vartheta)} \right|. \quad (4.13)$$

Эффективная оценка является, очевидно, Н.О.М.Д., но обратное неверно. Заметим, что согласно (4.10) эффективная оценка может существовать лишь для вполне определенной функции $\tau(\cdot)$. Если такая оценка существует для $\tau(\cdot)$, то она не существует ни для какой другой функции, отличной от $a\tau(\cdot) + b$, где a и b - некоторые числа.

Рассмотрим примеры эффективных оценок.

Пример 2.4.4. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\vartheta, \sigma^2)$ с известной дисперсией σ^2 . В таком случае функция правдоподобия имеет вид

$$L(x, \vartheta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \vartheta)^2 \right\}, \quad x \in \mathcal{R}^n, \quad \vartheta \in \mathcal{R}^1. \quad (4.14)$$

Следовательно,

$$\frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} = \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \vartheta) = \frac{n}{\sigma^2} \left[\frac{1}{n} \sum_{j=1}^n x_j - \vartheta \right], \quad \vartheta \in \mathcal{R}^1.$$

Поэтому согласно (4.10) статистика $t(\xi) = \frac{1}{n} \sum_{j=1}^n \xi_j$ является эффективной оценкой своего математического ожидания ϑ . Ее дисперсия в согласии с (4.13) равна σ^2/n .

Пример 2.4.5. Если $\xi = (\xi_1, \dots, \xi_n)$ — выборка из $N(\mu, \vartheta^2)$ с известным математическим ожиданием μ , то функция правдоподобия

$$L(x, \vartheta) = \left(\frac{1}{\sqrt{2\pi}\vartheta} \right)^n \exp \left\{ -\frac{1}{2\vartheta^2} \sum_{j=1}^n (x_j - \mu)^2 \right\}, \quad \vartheta \in (0, \infty),$$

$x = (x_1, \dots, x_n)$, и

$$\frac{\partial \ln L}{\partial \vartheta} = -\frac{n}{\vartheta} + \frac{1}{\vartheta^3} \sum_{j=1}^n (x_j - \mu)^2 = \frac{n}{\vartheta^3} \left\{ \frac{1}{n} \sum_{j=1}^n (x_j - \mu)^2 - \vartheta^2 \right\}.$$

Поэтому статистика $t(\xi) = \frac{1}{n} \sum_{j=1}^n (\xi_j - \mu)^2$ является эффективной оценкой дисперсии ϑ^2 . Ее дисперсия равна $|\tau'(\vartheta)/a(\vartheta)| = 2\vartheta^4/n$.

Теорема (2.4.2) остается верной, если плотность $f(x, \vartheta)$, $x \in \mathcal{R}^1$, $\vartheta \in \Theta$, заменить на вероятность, интегрирование — на суммирование. Рассмотрим пример эффективного оценивания параметра дискретного распределения.

Пример 2.4.6. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка из распределения Пуассона с неизвестным параметром $\vartheta \in (0, \infty)$. В таком случае логарифм функции правдоподобия будет иметь вид $\ln L(x, \vartheta) = \ln \prod_{j=1}^n \frac{\vartheta^{x_j}}{x_j!} e^{-\vartheta} = \sum_{j=1}^n (-\vartheta + x_j \ln \vartheta - \ln(x_j!))$, $\vartheta \in (0, \infty)$, $x_j \in \{0, 1, \dots\}$,

$$j = 1, \dots, n, \text{ и } \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} = \sum_{j=1}^n \left(-1 + \frac{x_j}{\vartheta} \right) = \frac{n}{\vartheta} \left\{ \frac{1}{n} \sum_{j=1}^n x_j - \vartheta \right\}. \text{ Поэто-}$$

му $t(\xi) = \frac{1}{n} \sum_{j=1}^n \xi_j$ — эффективная оценка ϑ с дисперсией ϑ/n .

Рассмотрим так называемое *экспоненциальное семейство распределений*, для которого семейство плотностей (или вероятностей) определяется следующим образом:

$$f(x, \vartheta) = \exp\{a(\vartheta)b(x) + c(\vartheta) + d(x)\}, \quad x \in \mathcal{R}^1, \quad \vartheta \in \mathcal{R}^1.$$

Соответственно функцию правдоподобия запишем как

$$L(x, \vartheta) = \exp\{a(\vartheta) \sum_{j=1}^n b(x_j) + nc(\vartheta) + \sum_{j=1}^n d(x_j)\},$$

где $x = (x_1, \dots, x_n) \in \mathcal{R}^n$. Если $a(\vartheta)$ и $c(\vartheta)$ — дифференцируемые функции $\vartheta \in \mathcal{R}^1$, то

$$\frac{\partial \ln L}{\partial \vartheta} = a'(\vartheta)n \left\{ \frac{1}{n} \sum_{j=1}^n b(x_j) + \frac{c'(\vartheta)}{a'(\vartheta)} \right\}.$$

Если для $L(x, \vartheta)$, $\tau(\vartheta) = -c'(\vartheta)/a'(\vartheta)$ и $t(\xi) = \frac{1}{n} \sum_{j=1}^n b(\xi_j)$ выполнены условия теоремы 2.4.2, то $\frac{1}{n} \sum_{j=1}^n b(\xi_j)$ — эффективная оценка $\tau(\vartheta) = -c'(\vartheta)/a'(\vartheta)$ с дисперсией, равной $|\tau'(\vartheta)/(na'(\vartheta))|$.

Экспоненциальному семейству принадлежат многие важные для практики распределения. Примеры некоторых из них рассмотрены выше.

2.4.3. Достаточные статистики. Изучая точечные оценки, следует обратить внимание на тот факт, что при их построении чаще всего не используется полная информация, содержащаяся в выборке. Так, для оценивания математического ожидания по выборке $\xi_1, \dots, \xi_n$ из нормального распределения достаточно знать лишь значение суммы $\xi_1 + \dots + \xi_n$, а не каждой случайной величины ξ_j , $j = 1, \dots, n$, в отдельности. Если μ известно, и по выборке $\xi_1, \dots, \xi_n$ из $N(\mu, \sigma^2)$ надлежит оценить дисперсию σ^2 , то для этого достаточно знать лишь $\sum_{j=1}^n (\xi_j - \mu)^2$ и т. п.

Оказывается, что во многих случаях для оценивания параметра ϑ распределения $F(\cdot, \vartheta)$ по выборке $\xi_1, \dots, \xi_n$ можно указать статистику

$$T(\xi) = (T_1(\xi), \dots, T_m(\xi)), \quad T_i : \mathcal{R}^n \rightarrow \mathcal{R}^1, \quad i = 1, \dots, m, \quad m < n, \quad (4.15)$$

в известном смысле содержащую всю информацию о параметре ϑ .

Определение 2.4.2. Статистика $T(\xi)$ (4.15) называется *достаточной* для семейства плотностей (вероятностей) $L(x, \vartheta)$, $x \in \mathcal{R}^n$, $\vartheta \in \Theta$, если условное распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ при условии $T(\xi) = t = \text{const}$ не зависит от ϑ .

На практике пользоваться этим определением для установления факта достаточности не очень удобно. Более удобно следующее *необходимое и достаточное условие достаточности статистики* $T(\xi)$.
Теорема 2.4.3. (о факторизации). $T(\xi)$ — достаточная статистика тогда и только тогда, когда функция правдоподобия может быть представлена в виде

$$L(x, \vartheta) = g(T(x), \vartheta)h(x), \quad x \in \mathcal{R}^n, \quad \vartheta \in \Theta. \quad (4.16)$$

Доказательство. Ограничимся случаем дискретного распределения. Пусть

$$p_{\vartheta}(x, t) = P(\{\xi = x, T(\xi) = t\}, \vartheta) = \begin{cases} L(x, \vartheta), & \text{если } T(x) = t, \\ 0, & \text{если } T(x) \neq t, \end{cases}$$

совместное распределение ξ и $T(\xi)$. Тогда распределение $T(\xi)$ определится равенством

$$p_{\vartheta}(t) = P(\{T(\xi) = t\}, \vartheta) = \sum_{x \in \mathcal{R}^n} p_{\vartheta}(x, t) = \sum_{x: T(x)=t} L(x, \vartheta),$$

и соответственно для условного распределения $P\{\xi = x | T(\xi) = t, \vartheta\} = p_{\vartheta}(x|t)$ найдем $p_{\vartheta}(x|t) = p_{\vartheta}(x, t)/p_{\vartheta}(t) = L(x, \vartheta) / \sum_{x: T(x)=t} L(x, \vartheta) = h(x) / \sum_{x: T(x)=t} h(x)$. Следовательно, статистика $T(\xi)$ — достаточная,

поскольку $p_{\vartheta}(x|t)$ от ϑ не зависит. Наоборот, пусть $p_{\vartheta}(x|t) = P\{\xi = x | T(\xi) = t, \vartheta\} = f(x, t)$, $x \in \mathcal{R}^n$, не зависит от ϑ . Тогда $p_{\vartheta}(x, t) = p_{\vartheta}(x|t)p_{\vartheta}(t)$ и для $t = T(x)$ $p_{\vartheta}(x, t) = L(x, \vartheta) = f(x, T(x))p_{\vartheta}(T(x))$, что совпадает с выражением (4.16), если $p_{\vartheta}(T(x)) = g(T(x), \vartheta)$. ■

Пример 2.4.7. В случае нормального распределения $N(\vartheta, \sigma^2)$ функция правдоподобия выборки $\xi_1, \dots, \xi_n$ имеет вид (4.14), или, иначе,

$$L(x, \vartheta) = g(T(x), \vartheta)h(x), \quad \text{где } g(T(x), \vartheta) = \exp \left\{ \frac{\vartheta}{\sigma^2} \sum_{j=1}^n x_j - \frac{n}{2\sigma^2} \vartheta^2 \right\} \text{ и}$$

$$h(x) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n x_j^2 \right\}. \text{ Следовательно, } T(x) = \sum_{j=1}^n x_j, \quad x \in \mathcal{R}^n, \\ \text{— достаточная статистика.}$$

На самом деле всякая эффективная оценка при условии (4.10) является достаточной статистикой, так как

$$\ln L(x, \vartheta) = \int a(\vartheta)(t(x) - \tau(\vartheta)) d\vartheta + l(x), \quad x \in \mathcal{R}^n,$$

где $l(\cdot)$ — некоторая функция, и для $L(x, \vartheta)$ выполнено условие факторизации (4.16), где

$$g(T(x), \vartheta) = \exp \left(\int a(\vartheta)(T(x) - \tau(\vartheta)) d\vartheta \right)$$

и

$$h(x) = \exp l(x), \quad x \in \mathcal{R}^n.$$

Более того, оказывается, что если Н.О.М.Д. существует, то ее можно найти как функцию достаточной статистики. Этот факт является следствием следующей теоремы.

Теорема 2.4.4. Пусть $T(x)$ — достаточная статистика для семейства $L(x, \vartheta)$, $\vartheta \in \Theta$, $x \in \mathcal{R}^n$. Если $t(\xi)$ — гильбертова несмещенная оценка $\tau(\vartheta)$, то $f(T(\xi))$ также гильбертова несмещенная оценка $\tau(\vartheta)$, причем

$$M_{\vartheta}(t(\xi) - \tau(\vartheta))^2 \geq M_{\vartheta}(f(T(\xi)) - \tau(\vartheta))^2; \quad (4.17)$$

здесь $f(s) = M_{\vartheta}(t(\xi)|T(\xi) = s)$ — условное математическое ожидание $t(\xi)$ при условии $T(\xi) = s$, $s \in \mathcal{R}^1$.

Доказательство. Отметим вначале, что $f(s)$ не зависит от ϑ (как математическое ожидание по условному распределению, не зависящему от ϑ). Поскольку согласно свойствам условного математического ожидания, рассмотренным в § 9,

$$M_{\vartheta}t(\xi) = M_T(M(t(\xi)|T)) = M_{\vartheta}f(T), \quad (4.18)$$

где M_T — математическое ожидание по распределению статистики $T(\xi)$ (зависящему от ϑ). Равенство (4.18) означает, что $f(T(\xi))$ — несмещенная оценка $\tau(\vartheta)$. Далее,

$$M_{\vartheta}f(T)(t - f(T)) = M_T\{f(T)M(t - f(T)|T)\} = 0, \quad (4.19)$$

так как

$$M_{\vartheta}(f(T)|T) = f(T), \quad M_{\vartheta}(t|T) = f(T).$$

Отсюда следует, что

$$\begin{aligned} M_{\vartheta}(t - \tau(\vartheta))^2 &= M_{\vartheta}(t - f(T) + f(T) - \tau(\vartheta))^2 = \\ &= M_{\vartheta}(t - f(T))^2 + M_{\vartheta}(f(T) - \tau(\vartheta))^2 \geq M_{\vartheta}(f(T) - \tau(\vartheta))^2, \end{aligned}$$

поскольку для математического ожидания произведения согласно (4.19), (4.18)

$$M_{\vartheta}(t - f(T))(f(T) - \tau(\vartheta)) = -M_{\vartheta}(t - f(T))\tau(\vartheta) = 0.$$

Если $t(\xi)$ — Н.О.М.Д. для $\tau(\vartheta)$ и $T(\xi)$ — достаточная статистика, то $f(T) = M_{\vartheta}(t(\xi)|T)$ согласно (4.19) также несмещенная оценка для $\tau(\vartheta)$, а согласно (4.17) ее дисперсия не превосходит дисперсии $t(\xi)$. Следовательно, $f(T)$ также Н.О.М.Д. для $\tau(\vartheta)$. ■

Определение 2.4.3. Достаточная статистика T называется *полной*, если всякая функция от нее с нулевым для всех $\vartheta \in \Theta$ математическим ожиданием равна нулю с вероятностью (для любого $\vartheta \in \Theta$) единица.

Покажем, что всякая функция $f(T)$ полной достаточной статистики является Н.О.М.Д. своего математического ожидания $\tau(\vartheta) = M_{\vartheta}f(T)$

для всех $\vartheta \in \Theta$. Действительно, как было отмечено, Н.О.М.Д. $\tau(\vartheta)$ следует искать в классе функций от T , а поскольку статистика T полная, то не существует двух различных функций от T , несмещенно оценивающих $\tau(\vartheta)$.

Если существует полная достаточная статистика T , то для получения Н.О.М.Д. $\tau(\vartheta)$ можно вначале построить любую несмещенную оценку $t(\xi)$, а затем найти ее условное математическое ожидание $f(s) = M_{\vartheta}(t(\xi)|T(\xi) = s)$. Тогда функция $f(s)$, $s \in \mathcal{R}^1$, не зависит от ϑ и $f(T(\xi))$ — Н.О.М.Д. для $\tau(\vartheta)$.

Пример 2.4.8. Рассмотрим схему испытаний Бернулли. Пусть ξ — число успехов в серии из n испытаний, ϑ — вероятность успеха в отдельном испытании, $0 < \vartheta < 1$. Поскольку $L(x, \vartheta) = C_n^x \vartheta^x (1 - \vartheta)^{n-x}$, $x = 0, 1, \dots, n$, то $T(\xi) = \xi$ — достаточная статистика. Пусть $f(\xi)$ — любая статистика, такая, что $M_{\vartheta} f(\xi) = \sum_{k=0}^n f(k) C_n^k \vartheta^k (1 - \vartheta)^{n-k} = 0$, $\vartheta \in (0, 1)$. То же самое можно

записать в виде $\sum_{k=0}^n f(k) C_n^k z^k = 0$, $0 < z < \infty$. Если полином степени n равен нулю для любого $z \in (0, \infty)$, то его коэффициенты — нули: $f(k) = 0$, $k = 0, \dots, n$. Следовательно, ξ — полная достаточная статистика, и тем самым любая функция $f(\xi)$ является Н.О.М.Д. для $M_{\vartheta} f(\xi)$. Поскольку

$$M_{\vartheta} \left(\frac{\xi}{n} \right) = \vartheta, \quad M_{\vartheta} \left(\frac{\xi(\xi-1)}{n(n-1)} \right) = \vartheta^2, \quad M_{\vartheta} \left(\frac{\xi(n-\xi)}{n(n-1)} \right) = \vartheta(1-\vartheta),$$

то ξ/n , $\xi(\xi-1)/n$ и $\xi(n-\xi)/(n(n-1))$ — Н.О.М.Д. для ϑ , ϑ^2 и $\vartheta(1-\vartheta)$ соответственно.

2.4.4. Оценки максимального правдоподобия. Пусть $L(x, \vartheta)$, $\vartheta \in \Theta$, $x \in \mathcal{R}^n$, — функция правдоподобия. *Оценкой максимального правдоподобия называется статистика $\hat{\vartheta} = \hat{\vartheta}(\xi)$, удовлетворяющая условию: $L(x, \hat{\vartheta}) \geq L(x, \vartheta)$ для всех $x \in \mathcal{R}^n$ и $\vartheta \in \Theta$, или, что то же самое,*

$$L(x, \hat{\vartheta}) = \max_{\vartheta \in \Theta} L(x, \vartheta), \quad x \in \mathcal{R}^n.$$

Если $L(x, \vartheta)$, $x \in \mathcal{R}^n$, не достигает максимума по $\vartheta \in \Theta$, то оценка максимального правдоподобия не существует. Название «оценка максимального правдоподобия» мы сохраним и за функцией $\hat{\vartheta}(x)$, $x \in \mathcal{R}^n$.

Пусть Θ — подмножество k -мерного евклидова пространства $\mathcal{R}^k$, функция правдоподобия $L(x, \vartheta)$ дифференцируема по $\vartheta \in \Theta$ и достигает максимума по ϑ во внутренней точке Θ для каждого $x \in \mathcal{R}^n$. В таком случае оценка максимального правдоподобия удовлетворяет уравнениям

$$\partial \ln L(x, \vartheta) / \partial \vartheta_j |_{\vartheta = \hat{\vartheta}(x)} = 0, \quad j = 1, \dots, k. \quad (4.20)$$

Уравнения (4.20) являются необходимыми условиями максимума и называются *уравнениями максимального правдоподобия*.

Если $x = (x_1, \dots, x_n)$ — наблюдаемое значение выборки $\xi = (\xi_1, \dots, \xi_n)$ из дискретного распределения с параметром ϑ , то $L(x, \vartheta) = P(\{\xi_1 = x_1, \dots, \xi_n = x_n, \vartheta\})$ — вероятность того, что $\xi_i = x_i$, $i = 1, \dots, n$. За оценку максимального правдоподобия $\hat{\vartheta}(x)$ принимается то значение параметра ϑ , при котором вероятность получить наблюдаемое значение x выборки ξ максимальна. Это замечание поясняет смысл принципа максимального правдоподобия: *в качестве значения неизвестного параметра предлагается использовать то, при котором вероятность наблюдаемой реализации выборки максимальна*.

Пример 2.4.9. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$. Поскольку

$$L(x, \vartheta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu^2) \right\},$$

то уравнения (4.20) максимального правдоподобия принимают вид $\frac{\partial \ln L(x, \vartheta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu) = 0$, $\frac{\partial \ln L(x, \vartheta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^n (x_j - \mu)^2 = 0$, откуда следует оценка максимального правдоподобия $\hat{\vartheta} = (\tilde{\mu}, \tilde{\sigma}^2)$, $\tilde{\mu} = \frac{1}{n} \sum_{j=1}^n x_j$, $\tilde{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \tilde{\mu})^2$.

Принцип максимального правдоподобия не основывается на каких-либо соображениях оптимальности. Вместе с тем, если, например, существует эффективная оценка числового параметра, то согласно (4.10) она является оценкой максимального правдоподобия, так как равенство $\partial \ln L(x, \vartheta) / \partial \vartheta = a(\vartheta)(t(x) - \vartheta) = 0$ влечет равенство $\hat{\vartheta}(x) = t(x)$, $x \in \mathcal{R}^n$. Более того, оценка максимального правдоподобия является функцией достаточной статистики, поскольку из условия $L(x, \vartheta) = g(T(x), \vartheta)h(x)$ следует, что $\hat{\vartheta} = \varphi(T)$. Но действительно замечательные свойства оценок максимального правдоподобия могут быть доказаны лишь как асимптотические для достаточно больших объемов выборки.

Пример 2.4.10. Остановимся подробнее на анализе оценки максимального правдоподобия параметров полиномиального распределения. Напомним, что полиномиальное распределение задается набором вероятностей $\vartheta = (p_1, \dots, p_r)$ каждого из r возможных исходов испытаний. Если в серии из n испытаний i -й исход наблюдается n_i раз, $i = 1, \dots, r$, $n_1 + \dots + n_r = n$, и ν_i — частота i -го исхода, $\nu_1 + \dots + \nu_r = 1$, то

$$P\left(\left\{\nu_1 = \frac{n_1}{n}, \dots, \nu_r = \frac{n_r}{n}\right\}, \vartheta\right) = L(\nu, \vartheta) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}. \quad (4.21)$$

Здесь $n_1 + \dots + n_r = n$, $\nu = (\nu_1, \dots, \nu_r)$, $\vartheta = (p_1, \dots, p_r)$. Пусть ν — набор наблюдаемых частот элементарных событий в серии из n испытаний, набор вероятностей $\vartheta = (p_1, \dots, p_r)$ неизвестен. Оценка $\hat{\vartheta}$ максимального правдоподобия ϑ определяется условиями

$$L(\nu, \hat{\vartheta}) = \max_{\vartheta \in \Theta} L(\nu, \vartheta). \quad (4.22)$$

Поскольку множество

$$\Theta = \{\vartheta = (p_1, \dots, p_r), 0 \leq p_i \leq 1, i = 1, \dots, r, p_1 + \dots + p_r = 1\}$$

ограничено и замкнуто в $\mathcal{R}^r$, а $L(\nu, \vartheta)$ при каждом $\nu \in \Theta$ как функция $\vartheta \in \Theta$ непрерывна, то максимум в (4.22) достигается в некоторой точке Θ . Следовательно, оценка максимального правдоподобия существует. Более того, множество Θ не только замкнуто и ограничено, но и выпукло в $\mathcal{R}^r$. Докажем, что следующая функция $f(\cdot)$ строго вогнута на Θ

$$\ln L(\nu, \vartheta) - \ln \frac{n!}{n_1! \dots n_r!} = \sum_{i=1}^r n_i \ln p_i = f(\vartheta), \quad (4.23)$$

откуда будет следовать единственность оценки максимального правдоподобия. Воспользуемся строгой вогнутостью $\ln x$, $x > 0$,

$$\ln(\alpha x + (1 - \alpha)y) > \alpha \ln x + (1 - \alpha) \ln y, \quad 0 < \alpha < 1, x, y > 0.$$

Пусть

$$\alpha \vartheta + (1 - \alpha) \bar{\vartheta} = (\alpha p_1 + (1 - \alpha) \bar{p}_1, \dots, \alpha p_r + (1 - \alpha) \bar{p}_r).$$

Тогда

$$\begin{aligned} f(\alpha \vartheta + (1 - \alpha) \bar{\vartheta}) &= \ln L(\nu, \alpha \vartheta + (1 - \alpha) \bar{\vartheta}) - \ln \frac{n!}{n_1! \dots n_r!} = \\ &= \sum_{i=1}^r n_i \ln(\alpha p_i + (1 - \alpha) \bar{p}_i) > \sum_{i=1}^r n_i (\alpha \ln p_i + (1 - \alpha) \ln \bar{p}_i) = \\ &= \alpha f(\vartheta) + (1 - \alpha) f(\bar{\vartheta}), \end{aligned}$$

что и означает строгую вогнутость функции (4.23).

Итак, оценка максимального правдоподобия ϑ существует, причем и единственная. Пусть $\vartheta_0 = (p_1^0, \dots, p_r^0)$ — истинные значения вероятностей исходов и $\hat{\vartheta} = (\hat{p}_1, \dots, \hat{p}_r)$. Покажем, что $\hat{\vartheta} \xrightarrow{P} \vartheta_0$ при числе испытаний n , стремящемся к бесконечности. Для этого нам потребуется

Лемма 2.4.2. Пусть $0 < p_i < 1$, $0 < q_i < 1$, $i = 1, \dots, r$, и $\sum_{i=1}^r p_i = \sum_{i=1}^r q_i = 1$. Тогда

$$\sum_{i=1}^r p_i \ln \frac{p_i}{q_i} \geq \frac{1}{2} \sum_{i=1}^r p_i (q_i - p_i)^2. \quad (4.24)$$

Доказательство. По формуле Тейлора

$$\ln x = \ln[1 + (x - 1)] = (x - 1) - \frac{(x - 1)^2}{2z^2}$$

для некоторого $z \in [1, x]$, если $x \geq 1$, или $z \in [x, 1]$, если $0 < x < 1$. Следовательно,

$$\begin{aligned} \sum_{i=1}^r p_i \ln \frac{q_i}{p_i} &= \sum_{i=1}^r p_i \left(\frac{q_i}{p_i} - 1 \right) - \frac{1}{2} \sum_{i=1}^r p_i \frac{(q_i/p_i - 1)^2}{z_i^2} = \\ &= -\frac{1}{2} \sum_{i=1}^r p_i (q_i - p_i)^2 \frac{1}{(p_i z_i)^2} \leq -\frac{1}{2} \sum_{i=1}^r p_i (q_i - p_i)^2, \end{aligned}$$

так как $p_i z_i \in [p_i, q_i]$, если $p_i \leq q_i$, или $p_i z_i \in [q_i, p_i]$ в противном случае. ■

Возвращаясь к поставленной задаче, найдем, воспользовавшись условием (4.22),

$$\sum_{i=1}^r \nu_i \ln \hat{p}_i = \max_p \sum_{i=1}^r \nu_i \ln p_i \geq \sum_{i=1}^r \nu_i \ln p_i^0.$$

Вместе с (4.24) это означает, что

$$\sum_{i=1}^r \nu_i \ln \frac{\nu_i}{p_i^0} \geq \sum_{i=1}^r \nu_i \ln \frac{\nu_i}{\hat{p}_i} \geq \frac{1}{2} \sum_{i=1}^r \nu_i (\nu_i - \hat{p}_i)^2. \quad (4.25)$$

Пусть теперь $n \rightarrow \infty$. Тогда в силу закона больших чисел $\nu_i \xrightarrow{P} p_i^0$, $i = 1, \dots, r$, а поскольку $\sum_{i=1}^r \nu_i \ln \frac{\nu_i}{p_i^0}$ — непрерывная функция $\nu = (\nu_1, \dots, \nu_r)$, левая, а, следовательно, и правая части неравенства (4.25) стремятся по вероятности к нулю. Отсюда, в свою очередь, следует, что $p_i \xrightarrow{P} p_i^0$, $i = 1, \dots, r$. Таким образом, оценка максимального правдоподобия параметров полиномиального распределения состоятельна.

2.5. Линейная модель измерений

Рассмотрим задачу, характерную для экспериментальных исследований, когда требуется измерить вектор параметров $\alpha = (\alpha_1, \dots, \alpha_k)$, но непосредственно измерены, причем с ошибками, могут быть лишь линейные комбинации его координат

$$a_{j1}\alpha_1 + \dots + a_{jk}\alpha_k, \quad j = 1, \dots, n,$$

где $n \geq k$. Коэффициенты a_{ji} , $j = 1, \dots, n$, $i = 1, \dots, k$, считаются известными.

Как правило, достаточно реалистичной является следующая схема измерений:

$$\xi_j = a_{j1}\alpha_1 + \dots + a_{jk}\alpha_k + \nu_j, \quad j = 1, \dots, n, \quad (5.1)$$

где $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор результатов измерений, $\nu = (\nu_1, \dots, \nu_n)$ — случайный вектор ошибок. Поскольку систематические ошибки измерений всегда могут быть учтены, разумно считать, что $M\nu_j = 0$, $j = 1, \dots, n$. Для простоты примем, что измерения (5.1) попарно независимы и имеют одинаковую точность, которую охарактеризуем дисперсией

$$M\nu_j^2 = \sigma^2, \quad j = 1, \dots, n, \quad (5.2)$$

которая, однако, не предполагается известной. Наконец, векторы $a_i = (a_{i1}, \dots, a_{in})$, $i = 1, \dots, k$ в (5.1) будем считать *линейно независимыми*. Задача состоит в том, чтобы по результатам измерений $\xi_1, \dots, \xi_n$ оценить $\alpha_1, \dots, \alpha_k$ и σ^2 .

Речь идет о типичной так называемой обратной задаче, когда по данным измерений требуется определить непосредственно ненаблюдаемые параметры объекта или явления. В математической статистике рассматриваемую задачу принято называть задачей анализа *линейной регрессии*.

2.5.1. Несмещенные оценки минимальной дисперсии значений $\alpha_1, \dots, \alpha_k$. Запишем схему измерений (5.1) в векторном виде

$$\xi = \sum_{i=1}^k a_i \alpha_i + \nu, \quad \xi \in \mathcal{R}^n. \quad (5.3)$$

Не предполагая известным распределение вектора ξ , покажем, что при сформулированных условиях *существуют линейные несмещенные оценки параметров $\alpha_1, \dots, \alpha_k$ с минимальной дисперсией, а при $n > k$ и несмещенная оценка σ^2* .

Линейную несмещенную оценку параметра α_i будем искать в виде

$$\hat{\alpha}_i = \sum_{j=1}^n b_{ij} \xi_j = (b_i, \xi), \quad b_i = (b_{i1}, \dots, b_{in}), \quad i = 1, \dots, k,$$

где $(\cdot, \cdot)$ — обозначение скалярного произведения в $\mathcal{R}^n$. Требование несмещенности $\hat{\alpha}_i$ приводит к условию $M\hat{\alpha}_i = (b_i, M\xi) = \sum_{i'=1}^k (b_i, a_{i'}) \alpha_{i'} = \alpha_i$, $i = 1, \dots, k$. Но поскольку α_i , $i = 1, \dots, k$, априори любые, это означает, что

$$(b_i, a_{i'}) = \delta_{ii'} = \begin{cases} 1, & i = i', \\ 0, & i \neq i', \end{cases} \quad i' = 1, \dots, k. \quad (5.4)$$

Подсчитаем теперь дисперсию $\hat{\alpha}_i$

$$D\hat{\alpha}_i = \sum_{j=1}^n (b_{ij})^2 D\xi_j = \sigma^2 \sum_{j=1}^n (b_{ij})^2 = \sigma^2(b_i, b_i). \quad (5.5)$$

Таким образом, задача определения несмещенной оценки минимальной дисперсии $\hat{\alpha}_i$, сводится к задаче на минимум дисперсии (5.5) при ограничениях (5.4)

$$\min\{(b_i, b_i) | b_i \in \mathcal{R}^n, (b_i, a_{i'}) = \delta_{ii'}, i' = 1, \dots, k\}, i = 1, \dots, k. \quad (5.6)$$

Речь идет о задаче на условный экстремум, которую можно решить методом множителей Лагранжа. Введем функцию Лагранжа i -ой задачи (5.6)

$$L_i = (b_i, b_i) - 2 \sum_{i'=1}^k \lambda_{ii'} (b_i, a_{i'}).$$

Здесь $\lambda_{ii'}$, $i' = 1, \dots, k$, множители Лагранжа. Дифференцируя L_i по координатам b_{ij} , $j = 1, \dots, n$, вектора, $b_i = (b_{i1}, \dots, b_{in})$, найдем

$$\frac{\partial L_i}{\partial b_{ij}} = 2b_{ij} - 2 \sum_{i'=1}^k \lambda_{ii'} a_{ji'}, j = 1, \dots, n,$$

где $(a_{1i'}, \dots, a_{ni'}) = a_{i'}$, $i' = 1, \dots, k$. Следовательно, уравнения Лагранжа задачи (5.6) в векторной форме будут иметь вид

$$b_i = \sum_{i'=1}^k \lambda_{ii'} a_{i'}, i = 1, \dots, k,$$

и совместно с условиями (5.4) это дает следующие уравнения для определения множителей Лагранжа

$$\delta_{ii''} = \sum_{i'=1}^k \lambda_{ii'} (a_{i'}, a_{i''}) = \begin{cases} 1, & i = i'', \\ 0, & i \neq i'', \end{cases} i, i'' = 1, \dots, k. \quad (5.7)$$

Так как векторы a_i , $i = 1, \dots, k$, линейно независимы, матрица $k \times k$ скалярных произведений $\{(a_i, a_{i'})\}$ невырождена. Поэтому из уравнений (5.7) следует, что

$$\lambda_{ii'} = (a_i, a_{i'})^-, i' = 1, \dots, k,$$

где $(\cdot, \cdot)^-$ — обозначение для матричных элементов матрицы, обратной $\{(a_i, a_{i'})\} : (a_i, a_{i'})^-$, — i, i' -матричный элемент матрицы $\{(a_i, a_{i'})\}^{-1}$, обратной $\{(a_i, a_{i'})\}$. Следовательно,

$$b_i = \sum_{i'=1}^k (a_i, a_{i'})^- a_{i'},$$

и линейная несмещенная оценка минимальной дисперсии параметра α_i имеет вид

$$\hat{\alpha}_i = (b_i, \xi) = \sum_{i'=1}^k (a_i, a_{i'})^{-1} (a_{i'}, \xi), \quad i = 1, \dots, k. \quad (5.8)$$

2.5.2. Метод наименьших квадратов оценивания значений $\alpha_1, \dots, \alpha_k$. Определим теперь оценку $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_k)$ параметров $\alpha = (\alpha_1, \dots, \alpha_k)$, из условия

$$\begin{aligned} & \left\| \xi - \sum_{i=1}^k \tilde{\alpha}_i a_i \right\| = \\ & = \min \left\{ \left\| \xi - \sum_{i=1}^k \alpha_i a_i \right\| \mid -\infty < \alpha_i < \infty, \quad i = 1, \dots, k \right\}. \end{aligned} \quad (5.9)$$

Смысл задачи (5.9) состоит в том, что в качестве оценок $\tilde{\alpha}_i, i = 1, \dots, k$ выбираются те значения $\alpha_i, i = 1, \dots, k$, при которых вектор $\sum_{i=1}^k \alpha_i a_i$ наименее уклоняется в $\mathcal{R}^n$ от результата измерений — вектора ξ . Вероятностная природа измерений в (5.1), точность измерения $\sum_{i=1}^k \alpha_i a_i$ и точность оценивания истинных значений $\alpha_i, i = 1, \dots, k$, при этом не учитывается. Значения $\tilde{\alpha}_i, i = 1, \dots, k$ называются *оценками наименьших квадратов* значений $\alpha_i, i = 1, \dots, k$.

Для решения задачи (5.9) введем оператор Π_a ортогонального проецирования в $\mathcal{R}^n$ на линейную оболочку $\mathcal{L}(a_1, \dots, a_k)$ векторов $a_1, \dots, a_k$. Тогда искомые оценки $\tilde{\alpha}_i, i = 1, \dots, k$, будут найдены из условия

$$\sum_{i=1}^k \tilde{\alpha}_i a_i = \Pi_a \xi. \quad (5.10)$$

Действительно,

$$\begin{aligned} \min \{ \|\xi - \eta\|^2 \mid \eta \in \mathcal{L}_a \} &= \min \{ \|\xi - \Pi_a \xi + \Pi_a \xi - \eta\|^2 \mid \eta \in \mathcal{L}_a \} \\ &= \|\xi - \Pi_a \xi\|^2 + \min \{ \|\Pi_a \xi - \eta\|^2 \mid \eta \in \mathcal{L}_a \} \end{aligned} \quad (5.11)$$

так как векторы $\xi - \Pi_a \xi \in \mathcal{L}_a^\perp$ и $\Pi_a \xi - \eta \in \mathcal{L}_a$ ортогональны. Поскольку $\Pi_a \xi \in \mathcal{L}_a$, из (5.11) следует, что $\min \|\xi - \eta\|$ достигается на $\eta = \Pi_a \xi$.

Равенство (5.10) выполняется тогда и только тогда, когда вектор $\xi - \sum_{i=1}^k \tilde{\alpha}_i a_i = \xi - \Pi_a \xi$ ортогонален $\mathcal{L}_a$, т. е. когда

$$\left(\xi - \sum_{i=1}^k \tilde{\alpha}_i a_i, a_{i'} \right) = 0, \quad i' = 1, \dots, k,$$

или, что то же самое,

$$\sum_{i'=1}^k (a_i, a_{i'}) \tilde{\alpha}_i = (\xi, a_i), \quad i = 1, \dots, k,$$

Отсюда следует, что оценки наименьших квадратов

$$\tilde{\alpha}_i = \sum_{i'=1}^k (a_i, a_{i'})^{-} (a_{i'}, \xi), \quad i = 1, \dots, k, \quad (5.12)$$

совпадают с оценками (5.8), полученными из условия их несмещенности и минимальности дисперсии и, следовательно, обладают этим экстремальным свойством.

Метод линейного оценивания, основанный на минимизации дисперсии, принадлежит Маркову (1900). Метод наименьших квадратов значительно раньше предложен Гауссом (1809).

Сформулируем полученные результаты.

Теорема 2.5.1. (Гаусса–Маркова). Пусть случайные координаты $\xi, \dots, \xi_n$ случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ попарно независимы и имеют одинаковую дисперсию. Если

$$M\xi = \sum_{i=1}^k \alpha_i a_i,$$

где $k \leq n$, векторы $a_i, i = 1, \dots, k$, известны и линейно независимы, то линейные несмещенные оценки минимальной дисперсии коэффициентов $\alpha_i, i = 1, \dots, k$, даются равенствами (5.8) и совпадают с оценками наименьших квадратов (5.12). Если значение σ^2 известно, то матричные элементы матрицы ковариаций оценок

$$\text{cov}(\hat{\alpha}_i \hat{\alpha}_{i'}) = \sigma^2 (a_i, a_{i'})^{-}, \quad i, i' = 1, \dots, k. \quad (5.13)$$

Доказательство. Следует проверить лишь последние равенства. Поскольку согласно (5.8) и (5.3)

$$\begin{aligned} \hat{\alpha}_i &= \sum_{i'=1}^k (a_i, a_{i'})^{-} \left(a_{i'}, \sum_{i''=1}^k a_{i''} \alpha_{i''} + \nu \right) = \\ &= \alpha_i + \sum_{i'=1}^k (a_i, a_{i'})^{-} (a_{i'}, \nu), \end{aligned} \quad (5.14)$$

то

$$\begin{aligned} \text{cov} \hat{\alpha}_i \hat{\alpha}_{i'} &= M(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_{i'} - \alpha_{i'}) = \\ &= \sum_{i'', i'''=1}^k (a_i, a_{i''})^{-} \sigma^2 (a_{i''}, a_{i'''}) (a_{i'''}, a_{i'})^{-} = \sigma^2 (a_i, a_{i'})^{-}. \end{aligned}$$

Здесь использовано соотношение

$$\begin{aligned} M(a_{i''}, \nu)(a_{i'''}, \nu) &= M \sum_{j=1}^n a_{ji''} \nu_j \sum_{j'=1}^n a_{ji''' } \nu_{j'} = \\ &= \sigma^2 \sum_{j=1}^n a_{ji''} a_{ji''' } = \sigma^2(a_{i''}, a_{i'''}). \blacksquare \end{aligned}$$

2.5.3. Несмещенная оценка дисперсии σ^2 , $n > k$. Рассмотрим статистики

$$s^2 = \left\| \xi - \sum_{i=1}^n \alpha_i a_i \right\|^2 = \|\nu\|^2, \quad (5.15)$$

$$s_1^2 = \left\| \xi - \Pi_a \xi \right\|^2 = \left\| \xi - \sum_{i=1}^k \hat{\alpha}_i a_i \right\|^2 = \|\nu - \Pi_a \nu\|^2,$$

$$s_2^2 = \left\| \Pi_a \xi - \sum_{i=1}^k \alpha_i a_i \right\|^2 = \left\| \sum_{i=1}^k (\hat{\alpha}_i - \alpha_i) a_i \right\|^2 = \|\Pi_a \nu\|^2.$$

Здесь ξ имеет выражение (5.3), $\hat{\alpha}_i$, $i = 1, \dots, k$, определены в (5.8) и учтены равенства (5.10), (5.12). Поскольку векторы $\xi - \Pi_a \xi \in \mathcal{L}_a^\perp$ и $\Pi_a \xi - \sum_{i=1}^k \hat{\alpha}_i a_i \in \mathcal{L}_a$ ортогональны, то

$$\begin{aligned} s^2 &= \left\| \xi - \Pi_a \xi + \Pi_a \xi - \sum_{i=1}^k \alpha_i a_i \right\|^2 = \\ &= \left\| \xi - \Pi_a \xi \right\|^2 + \left\| \Pi_a \xi - \sum_{i=1}^k \alpha_i a_i \right\|^2 = s_1^2 + s_2^2. \end{aligned}$$

Далее,

$$Ms^2 = M\|\nu\|^2 = M \sum_{j=1}^n \nu_j^2 = n\sigma^2,$$

и, следовательно,

$$\begin{aligned} Ms_2^2 &= M \left\| \sum_{i=1}^k (\hat{\alpha}_i - \alpha_i) a_i \right\|^2 = \sum_{i,i'=1}^k (a_i, a_{i'}) M(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_{i'} - \alpha_{i'}) = \\ &= \sum_{i,i'=1}^k (a_i, a_{i'})(a_i, a_{i'})^{-} \sigma^2 = k\sigma^2. \end{aligned}$$

Поэтому $Ms_1^2 = M(s^2 - s_2^2) = (n - k)\sigma^2$ и, следовательно, статистика

$$\hat{\sigma}^2 = \frac{\|\xi - \Pi_a \xi\|^2}{n - k} \quad (5.16)$$

является несмещенной оценкой дисперсии σ^2 .

Таким образом, составляющая $\Pi_a \xi \in \mathcal{L}_a$ вектора измерений определяет оценки $\hat{\alpha}_i$, $i = 1, \dots, k$, а оставшаяся информация в $\xi - \Pi_a \xi \in \mathcal{L}_a^\perp$ позволяет несмещенно оценить дисперсию измерений σ^2 .

2.5.4. Доверительные множества для значений $\alpha_1, \dots, \alpha_k$ в случае нормального распределения ошибок измерений. Предположим дополнительно, что в равенствах (5.1) ν_j , $j = 1, \dots, n$, нормально распределены $N(0, \sigma^2)$ и независимы в совокупности. В таком случае, воспользовавшись результатами §1.9 и равенствами (5.15), найдем

$$\begin{aligned} \|\xi - \sum_{i=1}^k \alpha_i a_i\|^2 &= \sigma^2 \chi_n^2, \\ \|\xi - \Pi_a \xi\|^2 &= \sigma^2 \chi_{n-k}^2, \\ \|\Pi_a \xi - \sum_{i=1}^k \alpha_i a_i\|^2 &= \sigma^2 \chi_k^2. \end{aligned} \quad (5.17)$$

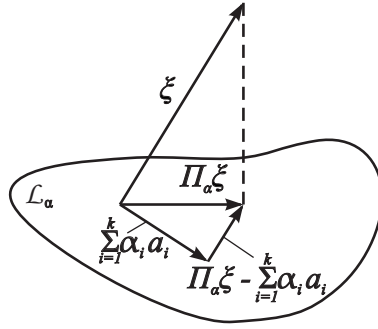


Рис. 2.2. Геометрическая иллюстрация равенств (5.17).

При этом согласно следствию 2.1.2 теоремы 2.1.1 случайные векторы $\xi - \Pi_a \xi = \nu - \Pi_a \nu$ и $\Pi_a \xi - \sum_{i=1}^k \alpha_i a_i = \Pi_a \nu$ независимы как ортогональные проекции вектора ν на ортогональные подпространства $\mathcal{L}_a^\perp$ и $\mathcal{L}_a$ соответственно. По этой причине статистика

$$\varphi_{k,n-k} = \frac{\|\Pi_a \xi - \sum_{i=1}^k \alpha_i a_i\|^2 / k}{\|\xi - \Pi_a \xi\|^2 / (n-k)} = \frac{\|\sum_{i=1}^k (\hat{\alpha}_i - \alpha_i) a_i\|^2 / k}{\|\xi - \Pi_a \xi\|^2 / (n-k)}$$

контролируется распределением Фишера с k и $n-k$ степенями свободы, см. рис. 2.2.

Если коэффициенты $\alpha_1, \dots, \alpha_k$ рассматривать как вектор $\alpha = (\alpha_1, \dots, \alpha_k)$ в k -мерном евклидовом пространстве $\mathcal{R}^k$, то статистика $\varphi_{k,n-k}$, может быть использована для получения доверительных множеств в

R^k для α . Действительно, для заданного γ , $0 \leq \gamma \leq 1$, определим по таблице распределения Фишера $\varepsilon > 0$ так, чтобы

$$P(\{\varphi_{k,n-k} < \varepsilon\}) = \gamma.$$

Тогда с вероятностью γ вектор истинных коэффициентов $(\alpha_1, \dots, \alpha_k)$ покрывается случайным подмножеством $\mathcal{R}^k$:

$$\left\{ \alpha = (\alpha_1, \dots, \alpha_k) : \left\| \sum_{i=1}^k (\hat{\alpha}_i - \alpha_i) a_i \right\|^2 = \sum_{i,i'=1}^k (a_i, a_{i'}) (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_{i'} - \alpha_{i'}) < \frac{\varepsilon k}{n-k} \|\xi - \Pi_a \xi\|^2 \right\}. \quad (5.18)$$

Доверительное подмножество (5.18) уровня γ является случайным эллипсоидом с центром $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_k)$. Понятие доверительного эллипсоида введено Хоттelingом в 1931 г.

Можно получить интервальные оценки для каждой координаты вектора $(\alpha_1, \dots, \alpha_k)$. Для этого заметим, что оценка $\hat{\alpha}_i$, как линейная функция нормальных случайных величин, также распределена нормально. Параметры этого распределения определяются равенствами

$$M\hat{\alpha}_i = \alpha_i, \quad D\hat{\alpha}_i = \text{cov } \hat{\alpha}_i \hat{\alpha}_i = \sigma^2(a_i, a_i)^-, \quad i = 1, \dots, k. \quad (5.19)$$

Заметим, в частности, что вектор $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_k)$ нормально распределен $N(\alpha, \sigma^2 \{(a_i, a_j)^-\})$. Так как случайный вектор $\Pi_a \xi$ и статистика $\|\xi - \Pi_a \xi\|^2$ независимы, то соответственно независимы $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_k)$ и $\|\xi - \Pi_a \xi\|^2$, поскольку $\hat{\alpha}_i = \sum_{i'=1}^k (a_i, a_{i'})^-(\xi, a_{i'}) = \sum_{i'=1}^k (a_i, a_{i'})^-(\Pi_a \xi, a_{i'})$, $i = 1, \dots, k$. А так как согласно (5.19) $(\hat{\alpha}_i - \alpha_i)/[(a_i, a_i)^-]^{1/2}$ имеет распределение $N(0, \sigma^2)$ и $\|\xi - \Pi_a \xi\|^2 = \sigma^2 \chi_{n-k}^2$, то статистика $\tau_{n-k} = \frac{\hat{\alpha}_i - \alpha_i}{[(a_i, a_i)^- \|\xi - \Pi_a \xi\|^2 / (n-k)]^{1/2}}$ имеет распределение Стюдента с $n-k$ степенями свободы.

Этот факт позволяет построить доверительный интервал для α_i , $i = 1, \dots, k$. Действительно, определим для заданного γ , $0 \leq \gamma \leq 1$, $\varepsilon \geq 0$ так, чтобы $P\{|\tau_{n-k}| < \varepsilon\} = \gamma$. Тогда случайный интервал $(\hat{\alpha}_i - \varepsilon \delta_i(\xi), \hat{\alpha}_i + \varepsilon \delta_i(\xi))$, где $\delta_i(\xi) = [(a_i, a_i)^- \|\xi - \Pi_a \xi\|^2 / (n-k)]^{1/2} = [(a_i, a_i)^- \hat{\sigma}^2]^{1/2}$ является доверительным интервалом уровня γ для α_i , $i = 1, \dots, k$.

Воспользовавшись выражением (5.16) для оценки дисперсии, легко получить доверительный интервал для σ^2 . Действительно, в рассматриваемом случае $\hat{\sigma}^2 = \|\xi - \Pi_a \xi\|^2 / (n-k) = \sigma^2 \chi_{n-k}^2$. Определим для

заданного γ , $0 \leq \gamma \leq 1$, величины ε_1 и ε_2 , $0 < \varepsilon_1 < \varepsilon_2$ так, чтобы

$$P\{\chi_{n-k}^2 \leq \varepsilon_1\} = P\{\chi_{n-k}^2 \geq \varepsilon_2\} = \frac{(1-\gamma)}{2}.$$

Тогда

$$P\{\varepsilon_1 < \hat{\sigma}^2/\sigma^2 < \varepsilon_2\} = P\{\hat{\sigma}^2/\varepsilon_2 < \sigma^2 < \hat{\sigma}^2/\varepsilon_1\} = \gamma.$$

Следовательно, $(\hat{\sigma}^2/\varepsilon_2, \hat{\sigma}^2/\varepsilon_1)$ — *искомый доверительный интервал уровня γ для σ^2* .

Пример 2.5.1. Проиллюстрируем полученные результаты на примере задач о взвешивании двух предметов. Пусть α_1 и α_2 — истинные веса двух предметов. Рассмотрим следующие три схемы взвешивания:

I	II	III
$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_2 + \nu_2; \end{cases}$	$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 + \alpha_2 + \nu_3; \end{cases}$	$\begin{cases} \xi_1 = \alpha_1 + \alpha_2 + \nu_1, \\ \xi_2 = \alpha_1 - \alpha_2 + \nu_2. \end{cases}$

Согласно первой схеме каждый предмет взвешивается один раз. Во второй схеме при дополнительном взвешивании на чашку весов кладутся оба предмета. Согласно третьей схеме при первом взвешивании предметы кладутся на одну чашку весов, а при втором — на разные. Будем считать, что в любой схеме ошибки взвешивания независимы, причем $M\nu_i = 0$, $D\nu_i = \sigma^2$, $i = 1, 2, 3$.

В схеме I: $a_1 = (1, 0)$, $a_2 = (0, 1)$, $\{(a_i, a_{i'})\} = \{(a_i, a_{i'})^-\} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Следовательно, $\hat{\alpha}_1 = \xi_1$, $\hat{\alpha}_2 = \xi_2$, $\{\text{cov}(\hat{\alpha}_i \hat{\alpha}_{i'})\} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. В этой схеме $n = k$, и измерения не содержат информации, достаточной для оценивания σ^2 . В схеме II: $a_1 = (1, 0, 1)$, $a_2 = (0, 1, 1)$,

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix},$$

и по формулам (5.8)

$$\begin{aligned} \hat{\alpha}_1 &= \frac{2}{3}\xi_1 - \frac{1}{3}\xi_2 + \frac{1}{3}\xi_3, \\ \hat{\alpha}_2 &= \frac{2}{3}\xi_2 - \frac{1}{3}\xi_1 + \frac{1}{3}\xi_3, \end{aligned} \tag{5.20}$$

причем

$$\{\text{cov} \hat{\alpha}_i \hat{\alpha}_{i'}\} = \sigma^2 \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}.$$

Отсюда, в частности, следует, что дополнительное взвешивание привело к уменьшению дисперсий оценок $\hat{\alpha}_i$, $i = 1, 2$: σ^2 в схеме I, $2\sigma^2/3$ в схеме II.

В схеме II возможно несмещенное оценивание дисперсии σ^2 . Задав, что $\hat{\alpha}_1 a_1 + \hat{\alpha}_2 a_2 = \frac{1}{3}((2\xi_1 - \xi_2 + \xi_3), (2\xi_2 - \xi_1 + \xi_3), (\xi_1 + \xi_2 +$

+ $2\xi_3$)), найдем $\xi - \Pi_a \xi = \xi - a_1 \hat{\alpha}_1 - a_2 \hat{\alpha}_2 = \frac{1}{3}((\xi_1 + \xi_2 - \xi_3), (\xi_2 + \xi_1 - \xi_3), (-\xi_1 - \xi_2 + \xi_3))$, Отсюда следует, что $\hat{\sigma}^2 = \frac{1}{3}(\xi_1 + \xi_2 - \xi_3)^2$. Доверительный эллипсоид для этого случая определяется неравенством $2(\alpha_1 - \hat{\alpha}_1)^2 + 2(\alpha_1 - \hat{\alpha}_1)(\alpha_2 - \hat{\alpha}_2) + 2(\alpha_2 - \hat{\alpha}_2)^2 \leq 2\varepsilon \hat{\sigma}^2$. Малая его полусось характеризует точность измерения $\alpha_1 + \alpha_2$ и в $\sqrt{3}$ раз короче большой.

Доверительный интервал для α_i , записывается в виде $(\hat{\alpha}_i - \varepsilon \delta(\xi), \hat{\alpha}_i + \varepsilon \delta(\xi))$, где $\delta(\xi) = \frac{\sqrt{2}}{3}|\xi_1 + \xi_2 - \xi_3|$ и $\hat{\alpha}_i$ определено в (5.20), $i = 1, 2$.

Наконец, в случае схемы III

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

и

$$\hat{\alpha}_1 = \frac{1}{2}(\xi_1 + \xi_2), \quad \hat{\alpha}_2 = \frac{1}{2}(\xi_1 - \xi_2).$$

В этом случае разумно организованное взвешивание позволило уменьшить дисперсию оценок вдвое по сравнению с результатом схемы I и в $4/3$ раза — по сравнению с результатом схемы II. В схеме III $\text{cov } \hat{\alpha}_1 \hat{\alpha}_2 = \text{cov } \hat{\alpha}_2 \hat{\alpha}_1 = \sigma^2/2$.

Как следует изменить стратегию взвешиваний в схеме II, чтобы, не увеличивая число взвешиваний, повысить точность результатов? Для этого, очевидно, следует стремиться к тому, чтобы при каждом взвешивании на чашках весов находились оба предмета.

Рассмотрим еще две схемы:

IV	V
$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_1 + \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 - \alpha_2 + \nu_3; \end{cases}$	$\begin{cases} \xi_1 = \alpha_1 + \alpha_2 + \nu_1, \\ \xi_2 = \alpha_1 + \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 - \alpha_2 + \nu_3; \end{cases}$
$a_1 = (1, 1, 1), \quad a_2 = (0, 1, -1);$	$a_1 = (1, 1, 1), \quad a_2 = (1, 1, -1);$
$\{(a_i, a_{i'})\} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix},$	$\{(a_i, a_{i'})\} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} =$
$\{(a_i, a_{i'})^-\} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \end{pmatrix};$	$= \begin{pmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{pmatrix};$
$\text{cov } \hat{\alpha}_1 \hat{\alpha}_1 = \sigma^2/3,$	$\text{cov } \hat{\alpha}_1 \hat{\alpha}_1 = 3\sigma^2/8,$
$\text{cov } \hat{\alpha}_2 \hat{\alpha}_2 = \sigma^2/2.$	$\text{cov } \hat{\alpha}_2 \hat{\alpha}_2 = 3\sigma^2/8.$

Каждая из этих схем дает результаты лучшие, чем схемы I, II и III. Схема V предпочтительнее схемы IV, если в качестве критерия выбрать суммарную ошибку оценивания: $\frac{3}{4}\sigma^2 = \frac{3}{8}\sigma^2 + \frac{3}{8}\sigma^2 < \frac{1}{3}\sigma^2 + \frac{1}{2}\sigma^2 = \frac{5}{6}\sigma^2$.

2.6. Методы анализа и интерпретации данных измерительного эксперимента

В этом параграфе рассмотрены математические методы анализа и интерпретации данных измерительного эксперимента, основанные на идеологии редукции данных измерений, представленной в монографиях [15, 16], и на «приборном» обобщении результатов §2.5. Под интерпретацией данных понимается редукция данных к виду, определенному исследователем и характерному для предметной области, а под анализом данных понимается независимое от предметной области исследование состоятельности математической модели эксперимента и ее характеристик, определяющих возможность получения новых знаний об объекте исследования, основанное на результатах измерений и другой доступной информации. Изложение ведется в терминах теории измерительно-вычислительных преобразователей (ИВП) как средств измерений [15, 18].

Рассмотрены понятия качества ИВП как средства измерения и качества измерительного преобразователя (ИП) как компоненты класса ИВП, понятия собственных базисов ИП и собственных базисов ИВП как инструментов исчерпывающего анализа свойств ИП и ИВП соответственно.

Проиллюстрированы актуальные для технических приложений следствия теории ИВП, важнейшим из которых является требование к характеристикам ИП, обеспечивающее максимальное качество ИВП как средства измерения, существенно отличающееся от требования к характеристикам ИП, обеспечивающим его высокое качество как средства измерений того же назначения.

2.6.1. Измерительно-вычислительный преобразователь как средство измерения. На рис. 2.3 представлена схема измерительного эксперимента, средством измерения в котором является измерительно-вычислительный преобразователь (ИВП). Данные измерений, сформированные (в системе I) в процессе (акте) взаимодействия измеряемого объекта, среды и измерительного преобразователя (ИП), представлены воздействием g на ИП (тепловым, механическим, радиационным и т. п.) и преобразованы им в электрический сигнал $\xi = Ag + \nu$, где A — оператор, моделирующий физические процессы в ИП (взаимодействующем с измеряемым объектом и со средой), определяющие преобразование воздействия g в сигнал Ag , ν — шум, погрешность преобразования.

Заметим, что до этого момента все процессы контролируются *физическими законами*, определяющими хорошо известные ограничения и запреты — термодинамические, дифракционные, квантовые и т. п. Но далее, когда сигнал ξ поступает в вычислительный преобразователь (ВП), решающую роль играют *математические свойства* физической модели измерительного эксперимента, т. е. взаимодействующих ИП,

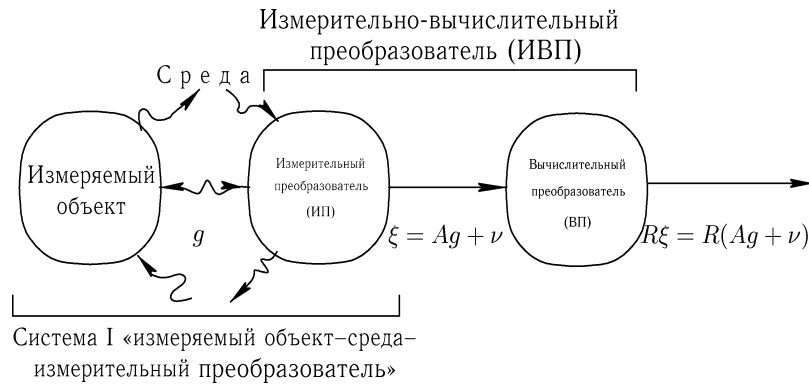


Рис. 2.3. g — входной сигнал ИП, сформированный в системе I при взаимодействии измеряемого объекта, среды и ИП, $\xi = Ag + \nu$ — выходной сигнал ИП, поступивший в ВП, $R\xi$ — выходной сигнал ИВП.

среды и измеряемого объекта, и *математические свойства* реализуемого ВП алгоритма анализа и интерпретации данных измерений, *доступных* в виде сигнала $\xi = Ag + \nu$. Алгоритм "должен извлечь из ξ " *максимально точные* значения представляющих интерес *характеристик исследуемого объекта*, подчеркнем, не *измеряемого*, характеристики которого искажены взаимодействием с ИП, а в некоторых случаях и со средой, а *исследуемого*, какие он имел бы, будучи в своем естественном состоянии, см. рис. 2.4.

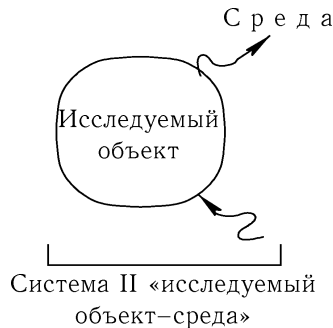


Рис. 2.4. Исследуемый объект, "естественное состояние" которого определено как взаимодействующего со средой.

Иначе говоря, на выходе ИВП должны быть *максимально точные версии характеристик исследуемого объекта*, т.е. характеристик "виртуальной реальности", представленной на рис.2.4.

На рис. 2.5 представлена схема этого же измерительного эксперимента с "идеальным ИП", моделируемым оператором U , который взаимодействует с измеряемым объектом и со средой точно так же, как ИП в системе I на рис. 2.3 и, следовательно, как ИП в системе I,

подвержен воздействию g со стороны исследуемого объекта и среды, но в отличие от выходного сигнала ИП на рис. 2.3 *выходной сигнал Ug идеального ИП определяет точные значения характеристик исследуемого объекта* в системе II, см. рис. 2.5.

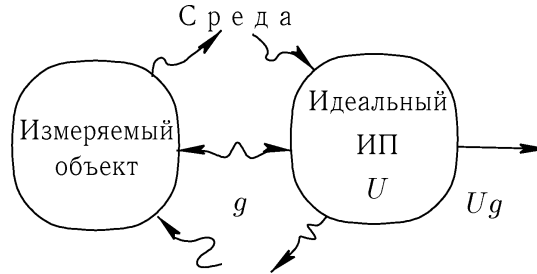


Рис. 2.5. Идеальный ИП взаимодействует с измеряемым объектом и со средой, как ИП в системе I на рис. 2.3, но на его выходе — характеристики Ug исследуемого объекта в системе II на рис. 2.4, неискаженные измерением.

Разумеется, идеальный ИП, как правило, не может быть реализован материально («в железе»), поскольку его «функционирование» обычно противоречит физическим законам, но его выходной сигнал Ug при известных условиях может быть синтезирован ИВП.

Иначе говоря, при известных требованиях к математическим моделям физических процессов, описывающих взаимодействия, представленные на рис. 2.3 и 2.4, *сигнал Ug может быть максимально точно*¹⁾ воспроизведен ИВП, в котором ВП математически моделирует «виртуальную реальность», представленную на рис. 2.4, на основе доступных в виде $\xi = Ag + \nu$ данных измерений.

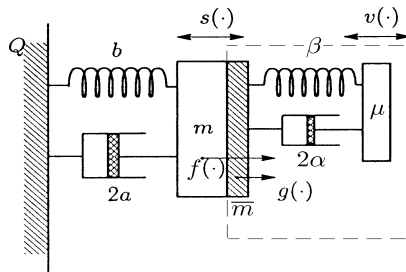


Рис. 2.6. Колебательная система, модель которой определена задачей (6.2), (6.3). Штриховой линией выделен ИП второго порядка.

¹⁾ Максимально точно в выбранном классе алгоритмов, реализуемых ВП. Далее, как правило, рассматриваются линейные ИП, ВП и ИВП. Нелинейные ИП и ИВП рассмотрены в [16].

Пример 2.6.1. Пусть исследуемым объектом в среде является механическая колебательная система второго порядка, динамика которой описывается решением задачи Коши

$$m\ddot{y}(t) + 2a\dot{y}(t) + by(t) = f(t), \quad 0 < t \leq T, \quad y(0) = 0, \quad \dot{y}(0) = 0. \quad (6.1)$$

Здесь $y(t)$ — смещение в момент времени t , m — масса подвижной части, b — коэффициент упругости, $2a$ — коэффициент вязкого трения, $f(t)$ — (неизвестная) сила, действующая на систему в момент времени t .

Для наблюдения за системой к ней жестко прикрепляется ИП, который также является колебательной системой второго порядка, у которой μ — масса подвижной части, 2α и β — коэффициенты вязкого трения и упругости соответственно (рис. 2.6). ИП «закреплен» на измеряемом объекте, и у последнего вследствие этого увеличена масса $m \rightarrow m + \bar{m}$, где $\bar{m}$ — масса «корпуса» ИП.

Динамика полученной системы описывается решением задачи Коши

$$(m + \bar{m})\ddot{s}(t) + 2a\dot{s}(t) + bs(t) = f(t) + \beta(v(t) - s(t)) + 2\alpha(\dot{v}(t) - \dot{s}(t)), \quad 0 < t \leq T, \quad s(0) = \dot{s}(0) = 0, \quad (6.2)$$

$$\mu\ddot{v}(t) + 2\alpha\dot{v}(t) + \beta v(t) = \beta s(t) + 2\alpha\dot{s}(t) \equiv g(t), \quad 0 < t \leq T, \quad v(0) = \dot{v}(0) = 0, \quad (6.3)$$

где $s(t)$, $v(t)$ — смещения *измеряемого объекта* и соответственно подвижной части ИП (относительно Q) в момент t , второе и третье слагаемые в правой части и слагаемое $\bar{m}\ddot{s}(t)$ в левой части (6.2) описывают *искажающее влияние* ИП на *динамику измеряемого объекта* (по сравнению с *динамикой исследуемого объекта* (6.1)), правая часть (6.3) определяет силу $g(\cdot)$, действующую на ИП.

Обычно в таких системах регистрируется сигнал, пропорциональный скорости движения подвижной части ИП относительно его корпуса,

$$\xi(t) = k(\dot{v}(t) - \dot{s}(t)) + \nu(t), \quad 0 \leq t \leq T, \quad (6.4)$$

здесь $\nu(t)$ — ошибка регистрации в момент $t \in [0, T]$.

Решив задачу Коши (6.2), (6.3), получим выражения

$$s(t) = \int_0^t p(t - \tau) f(\tau) d\tau, \quad v(t) = \int_0^t q(t - \tau) f(\tau) d\tau, \quad 0 \leq t \leq T,$$

где $p(\cdot)$ и $q(\cdot)$ — известные функции. Отсюда следует, что в (6.4)

$$k(\dot{v}(t) - \dot{s}(t)) = k \int_0^t (\dot{q}(t - \tau) - \dot{p}(t - \tau)) f(\tau) d\tau \stackrel{\text{def}}{=} Bf(t), \quad 0 \leq t \leq T, \quad (6.5)$$

ибо $p(0) = q(0) = 0$, и

$$g(t) = \int_0^t (2\alpha \dot{p}(t - \tau) + \beta p(t - \tau)) f(\tau) d\tau \stackrel{\text{def}}{=} C f(t) \quad (6.6)$$

— сила, с которой *измеряемый* объект действует на ИП в момент $t \in [0, T]$; B и C — известные линейные операторы, причем C имеет обратный C^{-1} . Согласно (6.4) — (6.6), $\xi(t) = Ag(t) + \nu(t) \equiv Bf(t) + \nu(t)$, $0 \leq t \leq T$, где $A = BC^{-1}$ — известный линейный оператор.

Если представляющей интерес характеристикой *исследуемого* объекта является его смещение $y(\cdot)$ в (6.1), то выходной сигнал идеального ИП, измеряющего это смещение,

$$y(t) = Ug(t) \equiv UCf(t) = \frac{1}{m\Omega} \int_0^t \exp(-\gamma(t - \tau)) \sin(\Omega(t - \tau)) f(\tau) d\tau, \quad (6.7)$$

$$0 \leq t \leq T,$$

является решением задачи Коши (6.1), $\gamma = a/m$, $\Omega^2 = (mb - a^2)/m^2 > 0$.

Если в (6.1) интерес представляет сила $f(\cdot)$, то $Ug(t) = f(t)$, $0 \leq t \leq T$, т. е. $U = C^{-1}$.

Одним из нетривиальных следствий рассматриваемой теории ИВП как средства измерения является *принцип измерений*, согласно которому при измерениях на ИВП значения характеристик измеряемого объекта могут быть существенно искажены по сравнению с их значениями, свойственными исследуемому объекту. Требование максимальной точности интерпретации измерений на ИВП как измерений на идеальном ИП приводит к тому, что при измерении взаимодействие ИП с измеряемым объектом должно быть достаточно сильным, в то время как при традиционных измерениях, как правило, требуется, чтобы ИП как можно меньше искажал характеристики измеряемого объекта. Важность нового принципа измерений проиллюстрирована в [16].

Другим нетривиальным следствием теории ИВП являются *требования к характеристикам* ИП, обеспечивающим максимальную точность синтеза на ИВП выходного сигнала идеального ИП, существенно отличающиеся от требований к характеристикам ИП, обеспечивающим его высокое качество как средства измерения того же назначения. Парадоксально, но во многих случаях улучшение ИП как такового приводит к ухудшению ИВП как средства измерения того же назначения, что и ИП. Характерные примеры приведены в [16].

Важным следствием в этой связи является тот факт, что распространенные методы "обработки" данных измерений типа методов наименьших квадратов [23] и их регуляризованных вариантов [22], максимальной энтропии [21] и т. п. не могут служить основой теории ИВП как средства измерения, поскольку не позволяют синтезировать сигнал

Ug , не гарантируют максимальной точности оценивания сигнала g и не позволяют определить параметры ИП, при которых максимальная точность ИВП как средства измерений достигается.

Наконец, важным следствием теории ИВП является осознание того, что оценка погрешности синтеза на ИВП выходного сигнала идеального ИП не вполне характеризует ИВП как средство измерений. Поскольку как модели ИП и взаимодействий в системе I, формирующих воздействие g на ИП, так и модель идеального ИП непременно содержат неконтролируемые погрешности, на выходе ИВП, кроме синтезированного выходного сигнала идеального ИП и оценки сопутствующей погрешности, должны быть данные, характеризующие непротиворечивость упомянутых моделей, возможность их использования для синтеза идеального ИП, и (желательно) данные, характеризующие адекватность найденных характеристик исследуемого объекта [16].

2.6.2. ИВП $[A, \Sigma_\nu, U]$ как максимально точная версия идеального ИП $[U, 0]$. Пусть в равенстве

$$\xi = Ag + \nu, \quad (6.8)$$

определяющем *схему измерения* в измерительном эксперименте на рис. 2.3, заданы линейный оператор $A : \mathcal{R}^m \rightarrow \mathcal{R}^n$, $n \geq m$, моделирующий ИП, взаимодействующий с измеряемым объектом и со средой, и ковариационный оператор $\Sigma_\nu : \mathcal{R}^n \rightarrow \mathcal{R}^n$ шума ν как случайного вектора $\mathcal{R}^n$

$$\Sigma_\nu x = M\nu(x, \nu)_n, \quad x \in \mathcal{R}^n, \quad (6.9)$$

где $(\cdot, \cdot)_n$ — символ скалярного произведения в $\mathcal{R}^n$, M — символ математического ожидания и предполагается, что $M\nu = 0$. Пусть, наконец, g в (6.8) — априори произвольный вектор $\mathcal{R}^m$, что вместе со сказанным об операторах A и Σ_ν будет означать, что *задана модель $[A, \Sigma_\nu]$ схемы измерения* (6.8) и тем самым *модель ИП в системе I* на рис. 2.3. Аналогично идеальный ИП, представленный на рис. 2.5, охарактеризуем моделью $[U, 0]$, где линейный оператор $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$. В операторе U "представлены" эффекты взаимодействия с измеряемым объектом и со средой, "представленные" в операторе A модели $[A, \Sigma_\nu]$, но в $[U, 0]$ отсутствует шум. Заметим, что если $e_1, \dots, e_n$ — ортонормированный базис $\mathcal{R}^n$, то $(e_i, \Sigma_\nu e_j)$, $i, j = 1, \dots, n$, — матричные элементы ковариационной матрицы ν ; модель $[A, \Sigma_\nu]$ *определена независимо от базисов $\mathcal{R}^m$ и $\mathcal{R}^n$* .

Рассмотрим задачу **редукции измерения** [16], выполненного по схеме (6.8) на ИП $[A, \Sigma_\nu]$, к виду $R_*\xi = R_*Ag + R_*\nu$, свойственному измерению на ¹⁾ ИП $[R_*A, R_*\Sigma_\nu R_*^*]$, *максимально точно аппроксимирующему выходной сигнал идеального ИП $[U, 0]$ на рис. 2.5, или, в иных терминах, — задачу редукции ИП $[A, \Sigma_\nu]$ к ИП $[R_*A, R_*\Sigma_\nu R_*^*]$,*

¹⁾ $R\Sigma_\nu R^* \stackrel{\text{def}}{=} \Sigma_{R\nu}$ — ковариационный оператор $R\nu$.

$[A, \Sigma_\nu] \rightarrow [R_*A, R_*\Sigma_\nu R_*^*]$, как задачу на минимум (максимальной) среднеквадратичной (с.к.) погрешности *интерпретации преобразования* $R\xi = RA g + R\nu$ измерения ξ как *выходного сигнала* Ug *идеального* ИП $[U, 0]$,

$$\text{Er}(A, \Sigma_\nu, U) \stackrel{\text{def}}{=} \sup_{g \in \mathcal{R}^m} M \|R_*\xi - Ug\|_k^2 = \min_{R: \mathcal{R}^m \rightarrow \mathcal{R}^k} \sup_{g \in \mathcal{R}^m} M \|R\xi - Ug\|_k^2. \quad (6.10)$$

В задаче (6.10) требуется найти *линейный оператор* $R_*: \mathcal{R}^n \rightarrow \mathcal{R}^k$, преобразующий в ВП выходной сигнал ξ (6.8) ИП $[A, \Sigma_\nu]$ в наиболее точную в с.к. версию $R_*\xi = R_*Ag + R_*\nu$ выходного сигнала Ug идеального ИП $[U, 0]$ (равномерно по $g \in \mathcal{R}^m$), и *определить ковариационный оператор* $\Sigma_U \stackrel{\text{def}}{=} R_*\Sigma_\nu R_*^*$, характеризующий погрешность, сопутствующую использованию ИП $[R_*A, R_*\Sigma_\nu R_*^*]$ в качестве идеального ИП $[U, 0]$. Тот факт, что цель редукции $[A, \Sigma_\nu] \rightarrow [R_*A, R_*\Sigma_\nu R_*^*]$ — *приблизить в смысле (6.10) выходной сигнал* идеального ИП $[U, 0]$, выразим символом $[A, \Sigma_\nu] \sim > [U, 0]$ приближенной, но самой точной в с.к., (линейной минимаксной) редукции ИП $[A, \Sigma_\nu]$ к ИП $[U, 0]$.

Выходной сигнал $R_*\xi = R_*Ag + R_*\nu$ ИВП, модель которого обозначим $[A, \Sigma_\nu, U]$, будем интерпретировать как *выходной сигнал* ИП $[R_*A, R_*\Sigma_\nu R_*^*]$, *наближайший в с.к. к выходному сигналу* идеального ИП $[U, 0]$, значение левой части в (6.10) назовем *погрешностью редукции* $[A, \Sigma_\nu] \sim > [U, 0]$. ИВП $[A, \Sigma_\nu, U]$ назовем *решающим задачу редукции* $[A, \Sigma_\nu] \sim > [U, 0]$.

Заметим, что в (6.10) согласно равенству (6.8)

$$M \|R\xi - Ug\|_k^2 = \|(RA - U)g\|_k^2 + \text{tr}(R\Sigma_\nu R^*), \quad (6.11)$$

где учтено, что $MR\nu = 0$ и $M(R\nu, R\nu)_k = \text{tr}(R\Sigma_\nu R^*)$. Если в (6.11) при любом $R: \mathcal{R}^n \rightarrow \mathcal{R}^k$ найдется $g_0 \in \mathcal{R}^m$, при котором $(RA - U)g_0 \neq 0$, то задача (6.10) неразрешима, поскольку $\sup_{g \in \mathcal{R}^m} M \|R\xi - Ug\|_k^2 \geq \sup_{g \in \mathcal{R}^m} \|(RA - U)g\|_k^2 = \infty$. Поэтому для разрешимости задачи (6.10) необходимо, чтобы

$$RA = U, \quad (6.12)$$

а задача (6.10) должна быть сформулирована как задача на условный минимум, в которой требуется найти оператор $R_*: \mathcal{R}^n \rightarrow \mathcal{R}^k$, удовлетворяющий условию

$$\text{tr}(R_*\Sigma_\nu R_*^*) = \min\{\text{tr}(R\Sigma_\nu R^*) | R: \mathcal{R}^n \rightarrow \mathcal{R}^k, RA = U\}. \quad (6.13)$$

Рассмотрим *вариант задачи* (6.13), в котором операторы Σ и $A^*\Sigma^{-1}A$ невырожденные. В таком случае условие (6.12) тождественно условию

$$(R - U(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma^{-1})A = 0, \quad (6.14)$$

согласно которому для любого $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$

$$R = U(A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma^{-1} + Z, \quad (6.15)$$

где $Z : \mathcal{R}^n \rightarrow \mathcal{R}^k$ — любой линейный оператор, удовлетворяющий условию

$$ZA = 0. \quad (6.16)$$

Согласно равенствам (6.15), (6.16), обозначив S оператор $A^* \Sigma_\nu^{-1} A$, найдем: $\text{tr}(R \Sigma_\nu R^*) = \text{tr}(U S^{-1} A^* \Sigma_\nu^{-1} + Z) \Sigma_\nu (U S^{-1} A^* \Sigma_\nu^{-1} + Z)^* = \text{tr}(U S^{-1} U^* + Z \Sigma_\nu Z^*)$, а так как $\text{tr}(Z \Sigma_\nu Z^*) \geq 0$ при любом $Z : \mathcal{R}^n \rightarrow \mathcal{R}^k$, то минимум в (6.13) достигается при

$$R = R_* \stackrel{\text{def}}{=} U(A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \quad (6.17)$$

и равен с.к. погрешности

$$\text{Er}(A, \Sigma_\nu, U) = \text{tr}(R_* \Sigma_\nu R_*^*) = \text{tr}(U(A^* \Sigma_\nu^{-1} A)^{-1} U^*) \quad (6.18)$$

редукции $[A, \Sigma_\nu] \sim > [U, 0]$ ИП $[A, \Sigma_\nu]$ к идеальному ИП $[U, 0]$.

Поэтому

$$[R_* A, R_* \Sigma_\nu R_*^*] = [U, \Sigma_U], \quad \Sigma_U = U(A^* \Sigma_\nu^{-1} A)^{-1} U^*. \quad (6.19)$$

Суммируем полученные результаты.

Теорема 2.6.1. Пусть операторы Σ_ν и $A^* \Sigma_\nu^{-1} A$ невырожденные. Тогда

- задача (6.13) при любом $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$ имеет единственное решение R_* (6.17),
 - минимальная с.к. ошибка интерпретации $R_* \xi = R_* A g + R_* \nu = U g + R_* \nu \stackrel{\text{def}}{=} \widehat{U} g$ как $U g$, $g \in \mathcal{R}^m$, равна погрешности редукции $[A, \Sigma_\nu] \sim > [U, 0]$ $\text{Er}(A, \Sigma_\nu, U)$ (6.18),
 - ковариационный оператор ошибки интерпретации $\widehat{U} g$ как $U g$ равен $\Sigma_U = U(A^* \Sigma_\nu^{-1} A)^{-1} U^*$ (6.19), где учтено, что $M \widehat{U} g = U g$.
- Поскольку при $U = I$ $\widehat{U} g = \widehat{g} = (A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \xi$, то
- $\widehat{U} g = U \widehat{g}$, где $\widehat{g}$ — наилучшая в с.к. оценка g , $\Sigma_I = (A^* \Sigma_\nu^{-1} A)^{-1}$ — ковариационный оператор $\widehat{g}$, $\text{Er}(A, \Sigma_\nu, I) = \text{tr}(A^* \Sigma_\nu^{-1} A)^{-1}$ — с.к. погрешность редукции $[A, \Sigma_\nu] \sim > [I, 0]$.

Согласно теореме 2.6.1 ИП $[A, \Sigma_\nu, U]$ моделирует ИП $[U, U(A^* \Sigma_\nu^{-1} A)^{-1} U^*]$, выходной сигнал которого — наиболее точная версия выходного сигнала идеального ИП $[U, 0]$.

Замечание 2.6.1. Поскольку в (6.8) g — произвольный вектор $\mathcal{R}^m$, исследователь может предпочесть априори конкретизировать модель шума редукции $[A, \Sigma_\nu] \sim > [U, 0]$, на фоне которого ему предстоит наблюдать значение $U g$ характеристик исследуемого объекта. В этой связи представляет интерес редукция схемы измерения (6.8) на ИП $[A, \Sigma_\nu]$ к виду

$$R \xi = U g + V \nu, \quad (6.20)$$

свойственному измерению на ИП $[U, V\Sigma_\nu V^*]$ с априори заданным ковариационным оператором шума $V\Sigma_\nu V^*$, где $U: \mathcal{R}^m \rightarrow \mathcal{R}^k$, $V: \mathcal{R}^n \rightarrow \mathcal{R}^k$.

Наиболее точную в с.к. версию $R_*\xi$ выходного сигнала ИП $[U, V\Sigma_\nu V^*]$ определим как решение задачи редукции $[A, \Sigma_\nu] \sim > [U, V\Sigma_\nu V^*]$

$$\begin{aligned} \sup_{g \in \mathcal{R}^m} M \|R_*\xi - Ug - V\nu\|^2 &= \min_{R: \mathcal{R}^n \rightarrow \mathcal{R}^k} \sup_{g \in \mathcal{R}^m} M \|R\xi - Ug - V\nu\|^2 = \\ &= \min_{R: \mathcal{R}^n \rightarrow \mathcal{R}^k} \left(\sup_{g \in \mathcal{R}^m} \|(RA - U)g\|^2 + \text{tr}((R - V)\Sigma_\nu(R - V)^*) \right), \end{aligned}$$

эквивалентной задаче на условный минимум

$$\sup_{g \in \mathcal{R}^m} M \|R_*\xi - Ug - V\nu\|^2 = \min_{\substack{R: \mathcal{R}^n \rightarrow \mathcal{R}^k, \\ RA = U}} \text{tr}((R - V)\Sigma_\nu(R - V)^*). \quad (6.21)$$

Так как при условиях теоремы 2.6.1 равенство $RA = U$ эквивалентно равенству $[R - V - (U - VA)(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1}]A = 0$, то подобно (6.15) получим

$$R - V = (U - VA)(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1} + Z, \quad ZA = 0$$

и далее подобно (6.17) – (6.19) найдем, что минимум в (6.21) достигается при

$$R_* = V + (U - VA)(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1}$$

и равен с.к. погрешности редукции $[A, \Sigma_\nu] \sim > [U, V\Sigma_\nu V^*]$

$$\text{Er}(A, \Sigma_\nu, U, V) = \text{tr}((U - VA)(A^*\Sigma_\nu^{-1}A)^{-1}(U - VA)^*). \quad (6.22)$$

При этом $[A, \Sigma_\nu] \rightarrow [R_*A, R_*\Sigma_\nu R_*^*] = [U, \Sigma_{U,V}]$, где

$$\Sigma_{U,V} = (U - VA)(A^*\Sigma_\nu^{-1}A)^{-1}(U - VA)^*$$

Замечание 2.6.2. Поскольку $\Pi = \Sigma_\nu^{-1/2}A(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1/2}: \mathcal{R}^n \rightarrow \mathcal{R}^n$ ортогонально проецирует на пространство $\mathcal{R}(\Sigma_\nu^{-1/2}A)$ значений $\Sigma_\nu^{-1/2}A$, то $(I - \Pi)\Sigma_\nu^{-1/2}\xi = (I - \Pi)\Sigma_\nu^{-1/2}\nu$, и если измерение в (6.8) контролируется моделью $[A, \Sigma_\nu]$, то статистика $\zeta = \text{tr}((I - \Pi)\Sigma_\nu^{-1/2}\xi\xi^*\Sigma_\nu^{-1/2})$ не зависит от $g \in \mathcal{R}^m$, $M\zeta = \text{tr}(I - \Pi) = n - m$ и $\text{Pr}(\zeta/(n - m) \geq \delta) \leq \delta^{-1}$. Если же, на самом деле, $\xi = Ag + \nu$, то $M\zeta = \text{tr}((I - \Pi)\Sigma_\nu^{-1/2}(\tilde{A} - A)gg^*(\tilde{A} - A)^*) + (n - m) \geq n - m$. Поэтому если, например, $\zeta/(n - m) \geq 10^2$, то модель $[A, \Sigma_\nu]$ может быть отвергнута как противоречащая результату измерения ξ , поскольку вероятность ошибки, сопутствующей такому решению, не больше 0,01.

2.6.3. О методе «наименьших квадратов». Как известно метод наименьших квадратов (МНК), введенный в практику интерпретации измерительного эксперимента К. Гауссом, состоит в следующем. Если $\xi = Ag + \nu$ – результат измерения сигнала $g \in \mathcal{R}^m$, то в качестве его

оценки $\hat{g}$ предлагается решение задачи на минимум для так называемой ¹⁾ *невязки*

$$\|\xi - Ag\|_n \sim \min_{f \in \mathcal{R}^m} \quad (6.23)$$

Проверим, что если $\|\xi - Ag\|_n^2 = (\xi - Ag, \xi - Ag)_n$, то $\tilde{g} = (A^*A)^{-1}A^*\xi$ — решение задачи (6.23). В самом деле $\Pi = A(A^*A)^{-1}A^*$ — ортогональный проектор в $\mathcal{R}^m$, поскольку $\Pi^2 = \Pi$ и $\Pi^* = \Pi$ причем в равенстве $\|\xi - Ag\|_n^2 = \|\xi - \Pi\xi + \Pi\xi - Ag\|_n^2 = \|\xi - \Pi\xi\|_n^2 + 2(\xi - \Pi\xi, \Pi\xi - Ag)_n + \|\Pi\xi - Ag\|_n^2$ скалярное произведение равно нулю, ибо $\xi - \Pi\xi \in R^\perp(\Pi)$, $\Pi\xi - Ag = \Pi\xi - \Pi Ag \in R(\Pi)$. Поэтому

$$\|\xi - Ag\|_n^2 = \|\xi - \Pi\xi\|_n^2 + \|\Pi\xi - Ag\|_n^2 \geq \|\xi - \Pi\xi\|_n^2, \quad (6.24)$$

и левая часть неравенства достигает минимума при $\Pi\xi - Ag = A((A^*A)^{-1}A^*\xi - g) = 0$, т.е. на единственном $g = \tilde{g} = (A^*A)^{-1}A^*\xi$, поскольку $\text{rang } A = m$, и следовательно ²⁾, если $Af = 0$, то $f = 0$.

Понятно, что найденной МНК оценке $\tilde{g} = (A^*A)^{-1}A^*\xi$ при $\Sigma_\nu \neq \sigma^2 I$, $\sigma^2 > 0$, сопутствует бо́льшая с.к. погрешность, чем оценке $\hat{g}$. Но если вместо евклидовой нормы в (6.23) выбрать $\|\Sigma_\nu^{-1/2}(\xi - Ag)\|_n$, то $\tilde{g} = (A^*A)^{-1}A^*\xi$ «превратится» в $\hat{g} = (A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1/2}(\Sigma_\nu^{-1/2}\xi)$, т.е. в наилучшую в среднеквадратичном оценку. Но этот акт не имеет отношения к методу наименьших квадратов.

Заметим также, что МНК не позволяет поставить задачу наилучшего в с.к. оценивания Uf , $f \in \mathcal{R}^m$, где $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$.

2.6.4. Качество ИВП $[A, \Sigma_\nu, U]$ и ИП $[A, \Sigma_\nu]$ как компоненты ИВП. Качество ИВП $[A, \Sigma_\nu, U]$ тем выше, чем меньше с.к. погрешность $\text{Er}(A, \Sigma_\nu, U)$ (6.18) интерпретации его выходного сигнала $R_*\xi$ как Ug , $g \in \mathcal{R}^k$.

Определение 2.6.1. Пусть операторы $A : \mathcal{R}^m \rightarrow \mathcal{R}^n$, $\tilde{A} : \mathcal{R}^m \rightarrow \mathcal{R}^{\tilde{n}}$, $\Sigma_\nu : \mathcal{R}^n \rightarrow \mathcal{R}^n$, $\tilde{\Sigma}_{\tilde{\nu}} : \mathcal{R}^{\tilde{n}} \rightarrow \mathcal{R}^{\tilde{n}}$, $m \leq \min(n, \tilde{n})$, удовлетворяют условиям теоремы 2.6.1. Тогда будем говорить, что *при фиксированном* $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$ *качество* ИВП $[A, \Sigma_\nu, U]$ *как* ИП $[U, 0]$ *выше, чем качество* ИВП $[\tilde{A}, \tilde{\Sigma}_{\tilde{\nu}}, U]$ *как* ИП $[U, 0]$, и писать $[A, \Sigma_\nu, U] \prec [\tilde{A}, \tilde{\Sigma}_{\tilde{\nu}}, U]$, если

$$\text{tr}(U(A^*\Sigma_\nu^{-1}A)^{-1}U^*) < \text{tr}(U(\tilde{A}^*\tilde{\Sigma}_{\tilde{\nu}}^{-1}\tilde{A})^{-1}U^*), \quad (6.25)$$

а если $\text{tr}(U(A^*\Sigma_\nu^{-1}A)^{-1}U^*) = \text{tr}(U(\tilde{A}^*\tilde{\Sigma}_{\tilde{\nu}}^{-1}\tilde{A})^{-1}U^*)$, то будем говорить, что ИВП $[A, \Sigma_\nu, U]$ и $[\tilde{A}, \tilde{\Sigma}_{\tilde{\nu}}, U]$ *эквивалентны*, и писать $[A, \Sigma_\nu, U] \sim [\tilde{A}, \tilde{\Sigma}_{\tilde{\nu}}, U]$.

¹⁾ Подчеркнем, что минимизируется не ошибка оценивания g , а «невязка», т.е. расстояние между ξ и Ag .

²⁾ Отсюда следует, что $A(A^*A)^{-1}A^*$ — ортогонально проецирует на $\mathcal{R}(A)$.

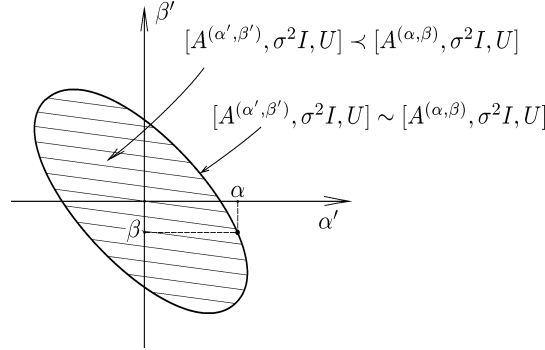


Рис. 2.7. Множества, определяющие классы ИВП, эквивалентных $[A^{(\alpha, \beta)}, \sigma^2 I, U]$ и более высокого качества.

Пример 2.6.2. Пусть в ИП $[A, \Sigma_\nu]$ оператор $A = A^{(\alpha, \beta)}$ определен как решение задачи Коши $\alpha \dot{u}(t) + \beta u(t) = g(t)$, $0 < t < T$, $u(0) = 0$, $\dot{u} \stackrel{\text{def}}{=} du/dt$, $\alpha \neq 0$, т.е. пусть $A^{(\alpha, \beta)} g(t) = \frac{1}{\alpha} \int_0^t \exp(-\beta(t-\tau)/\alpha) g(\tau) d\tau$, $0 \leq t \leq T$, и пусть $\Sigma_\nu = \sigma^2 I$. Тогда для любого U , удовлетворяющего условию, обеспечивающему разрешимость задачи редукции $[A^{(\alpha, \beta)}, \sigma^2 I] \rightarrow [U, \Sigma_U]$, [16], класс ИВП, эквивалентных ИВП $[A^{(\alpha, \beta)}, \sigma^2 I, U]$ на плоскости (α', β') , $-\infty < \alpha' < \infty$, $-\infty < \beta' < \infty$, есть эллипс с центром в $(0, 0)$, содержащий точку (α, β) см. рис.2.7. Любая точка (α', β') , $\alpha' \neq 0$, внутри (вне) этого эллипса определяет ИВП $[A^{(\alpha', \beta')}, \sigma^2 I, U]$, качество которого выше (ниже), чем качество $[A^{(\alpha, \beta)}, \sigma^2 I, U]$, [16].

На практике ИВП $[A, \Sigma_\nu, U]$ чаще используется для решения класса задач редукции $[A, \Sigma_\nu] \sim [U, 0]$, $U \in \mathcal{U}$. В этих случаях актуально **Определение 2.6.2.** Пусть операторы A , $\tilde{A}$, Σ_ν и $\tilde{\Sigma}_\nu$ удовлетворяют условиям, сформулированным в определении 2.6.1. Тогда будем говорить, что для класса задач редукции $[A, \Sigma_\nu] \sim [U, 0]$, $U \in \mathcal{U}$, качество ИП $[A, \Sigma_\nu]$ как компоненты ИВП равномерно не ниже, чем качество ИП $[\tilde{A}, \tilde{\Sigma}_\nu]$, и писать $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\preceq} [\tilde{A}, \tilde{\Sigma}_\nu]$, если для любого $U \in \mathcal{U}$

$$\text{tr}(U(A^* \Sigma_\nu^{-1} A)^{-1} U^*) \leq \text{tr}(U(\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1} U^*). \quad (6.26)$$

Если $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\preceq} [\tilde{A}, \tilde{\Sigma}_\nu]$ и $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\succ} [\tilde{A}, \tilde{\Sigma}_\nu]$, то будем говорить, что ИП $[A, \Sigma_\nu]$ и $[\tilde{A}, \tilde{\Sigma}_\nu]$ как компоненты ИВП эквивалентны, и писать $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\sim} [\tilde{A}, \tilde{\Sigma}_\nu]$, а если $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\preceq} [\tilde{A}, \tilde{\Sigma}_\nu]$ и ИП неэквивалентны, то будем говорить, что качество $[A, \Sigma_\nu]$ как компоненты ИВП равномерно выше, чем качество $[\tilde{A}, \tilde{\Sigma}_\nu]$, и писать $[A, \Sigma_\nu] \stackrel{\mathcal{U}}{\prec} [\tilde{A}, \tilde{\Sigma}_\nu]$.

Если $\mathcal{U}$ — класс всех линейных операторов $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$, $k = 1, \dots, m$, то вместо символов $\overset{\mathcal{U}}{\preceq}$ и $\overset{\mathcal{U}}{\succeq}$ будем использовать $\preceq$ и $\succeq$.

Лемма 2.6.1. Если класс $\mathcal{U}$ в определении 2.6.2 содержит операторы

$$U : \mathcal{R}^m \rightarrow \mathcal{R}^1, Ug \stackrel{\text{def}}{=} (u, g)_m, g \in \mathcal{R}^m, \quad (6.27)$$

при любом $u \in \mathcal{R}^m$, то $[A, \Sigma_\nu] \preceq [\tilde{A}, \tilde{\Sigma}_\nu]$, $[A, \Sigma_\nu] \prec [\tilde{A}, \tilde{\Sigma}_\nu]$, $[A, \Sigma_\nu] \sim [\tilde{A}, \tilde{\Sigma}_\nu]$, если и только если соответственно

$$\begin{aligned} (A^* \Sigma_\nu^{-1} A)^{-1} &\leq (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1}, \quad (A^* \Sigma_\nu^{-1} A)^{-1} < (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1}, \\ (A^* \Sigma_\nu^{-1} A)^{-1} &= (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1}, \end{aligned} \quad (6.28)$$

или, что в силу невырожденности оператора $\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A} (> 0)$ эквивалентно условиям (6.28),

$$A^* \Sigma_\nu^{-1} A \geq \tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A}, \quad A^* \Sigma_\nu^{-1} A > \tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A}, \quad A^* \Sigma_\nu^{-1} A = \tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A}. \quad (6.28^*)$$

Действительно, выбрав в определении 2.6.2 в качестве U оператор (6.27), получим условия (6.26) в виде верного для любого $u \in \mathcal{R}^m$ неравенства $(u, (A^* \Sigma_\nu^{-1} A)^{-1} u)_m \leq (u, (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1} u)_m$, означающего, что выполнено первое операторное неравенство в (6.28), а если для любого $u \in \mathcal{R}^m$, $\|u\|_m > 0$, $(u, (A^* \Sigma_\nu^{-1} A)^{-1} u)_m > (u, (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1} u)_m$, то выполнено второе неравенство в (6.28), наконец, равенство в (6.28) эквивалентно равенству $(u, (A^* \Sigma_\nu^{-1} A)^{-1} u)_m = (u, (\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1} u)_m$. Наоборот, если выполнено одно из условий в (6.28), например второе, то для любого $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$, $k = 1, \dots, m$, $\text{tr}(U(A^* \Sigma_\nu^{-1} A)^{-1} U^*) < \text{tr}(U(\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A})^{-1} U^*)$, если $\text{tr}(UU^*) > 0$.

Следующий пример показывает, что требования к характеристикам ИП, определяющие его высокое качество как средства измерения, могут значительно отличаться от требований к характеристикам ИП как компоненты ИВП, определяющих высокое качество ИВП как средства измерения того же назначения.

Пример 2.6.3. Рассмотрим ИВП $[A, \Sigma_\nu, U]$, в котором $A = \begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix}$,

$\Sigma_\nu = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_\nu$, $U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$. Поскольку оператор U идеального ИП определен как единичный I , отличие A от идеального U тем больше, чем больше $|b|$ (тем значительнее искажения, вносимые A). В схеме измерения $\xi = Ag + \nu$ (6.8) g можно считать выходным сигналом идеального ИП $[I, 0]$, а оператор A — моделирующим "аппаратные" искажения, свойственные ИП $[A, I_\nu]$. Так как в рассматриваемом случае редукции $[A, I_\nu] \rightarrow [I, \Sigma_I]$ $\Sigma_I = (A^* A)^{-1} = \frac{1}{(1-b^2)^2} \begin{pmatrix} 1+b^2 & -2b \\ -2b & 1+b^2 \end{pmatrix}$, то погрешность редукции

$$[A, I_\nu] \sim > [I, 0]$$

$$\text{Er}(A, I_\nu, I) = 2(1 + b^2)/(1 - b^2)^2 \stackrel{\text{def}}{=} q(b). \quad (6.29)$$

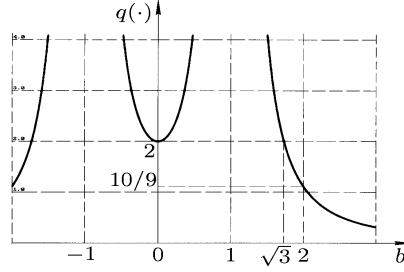


Рис. 2.8. График с. к. погрешности $\text{Er}(A, I_\nu, I)$ (6.29) редукции $[A, I_\nu] \sim > [I, 0]$ как функции $b \in (-\infty, \infty)$.

Если $b = 0$, то A в ИП $[A, I_\nu]$ совпадает с I в идеальном ИП $[I, 0]$, не вносящим искажений в сигнал g , и в этом случае с. к. погрешность редукции $\text{Er}(I, I_\nu, I) = 2$, см. рис. 2.8. Для $|b| > 1$ $q(b)$ (6.29) монотонно убывает, при $|b| = \sqrt{3}$ с. к. погрешность редукции $[A, I_\nu] \sim > [I, 0]$ равна 2, т. е. совпадает с погрешностью редукции $[I, I_\nu] \sim > [I, 0]$, хотя при $|b| = \sqrt{3}$ оператор A существенно искажает сигнал g , и далее с ростом $|b|$ погрешность редукции $[A, I_\nu] \sim > [I, 0]$ убывает, в то время как искажения, вносимые A , растут. Для $|b| > \sqrt{3}$ ИВП $[A, I_\nu, I]$ "представляет" идеальный ИП $[I, 0]$ лучше, чем ИВП $[I, I_\nu, I]$ с идеальным I "на входе". Наконец, согласно (6.28*) $[A, I_\nu] \preceq [I, I_\nu]$, если $A^*A - I = \begin{pmatrix} b^2 & 2b \\ 2b & b^2 \end{pmatrix} \geq 0$, т. е. если $|b| \geq 2$.

Что касается последнего результата, то при $|b| > 2$ "энергия" сигнала Ag больше, чем "энергия" сигнала Ig , ибо $\forall g \in \mathcal{R}^2$ $\|Ag\|^2 - \|g\|^2 = ((A^*A - I)g, g) > 0$, поскольку при $|b| > 2$ $A^*A - I > 0$. Поэтому при $|b| > 2$ $[A, I_\nu, U] \prec [I, I_\nu, U]$ для любого U . Но если $\sqrt{3} < b < 2$, то $\|Ag\|^2$ при некоторых $g \in \mathcal{R}^2$ меньше, чем $\|g\|^2$, но ИВП $[A, I_\nu, I]$ моделирует идеальный ИП $[I, 0]$ точнее, чем ИВП $[I, I_\nu, I]$ с идеальным ИП «на входе».

Более того, если в ИВП $[A, \Sigma_\nu, U]$ несколько увеличить уровень шума, т. е. $\Sigma_\nu = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ заменить на $\tilde{\Sigma}_\nu = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$, $\sigma^2 > 1$, то соответственно возрастет с. к. ошибка (6.29): $\text{Er}(A, I_\nu, I) \rightarrow 2\sigma^2(1 + b^2)/(1 - b^2)^2 = \text{Er}(A, \tilde{\Sigma}_\nu, I)$, но увеличив искажения, вносимые A , $|b| \rightarrow |b(\sigma)|$, так, чтобы $2\sigma^2(1 + b^2(\sigma))/(1 - b^2(\sigma))^2 = 2$, получим, что для $|b| > |b(\sigma)|$ $[A, \tilde{\Sigma}_\nu, I] \prec [I, I_\nu, I]$. Этот результат с инженерной точки зрения выглядит парадоксально: в ИП $[A, \tilde{\Sigma}_\nu]$ A вносит больше искажений, чем I в ИП $[I, I_\nu]$, и уровень шума в ИП $[A, \tilde{\Sigma}_\nu]$ выше, чем в ИП $[I, I_\nu]$, но $[A, \tilde{\Sigma}_\nu, I] \prec [I, I_\nu, I]$!

2.6.5. Влияние уточнения модели ИП на его качество. Если модель взаимодействия ИП, среды и измеряемого объекта, см. рис. 2.3, позволяет *априори охарактеризовать сигнал* $g \in \mathcal{R}^m$, который *может воздействовать* на ИП, то это следует отразить в модели ИП и тем самым, возможно, повысить его качество.

2.6.5.1 Модель $[A, \mathcal{G}, \Sigma_\nu]$.

Если априори известно, что g содержится в области $\mathcal{G} \subset \mathcal{R}^m$, то речь идет о модели $[A, \mathcal{G}, \Sigma_\nu]$ схемы измерения (6.8) и соответствующего ИП. В таком случае задачу редукции следует поставить как следующую задачу на минимум с.к. погрешности редукции $[A, \mathcal{G}, \Sigma_\nu] \sim> [U, 0]$:

$$\text{Er}(A, \mathcal{G}, \Sigma_\nu, U) = \sup_{g \in \mathcal{G}} M \|R_* \xi - Ug\|_k^2 = \min_{R: \mathcal{R}^n \rightarrow \mathcal{G}} \sup_{g \in \mathcal{G}} M \|R\xi - Ug\|_k^2. \quad (6.30)$$

Учет условия $g \in \mathcal{G}$ может повлечь уменьшение с.к. погрешности редукции $[A, \mathcal{G}, \Sigma_\nu] \sim> [U, 0]$, ибо для любого $R: \mathcal{R}^n \rightarrow \mathcal{R}^m$ и для любого $U: \mathcal{R}^m \rightarrow \mathcal{R}^k$ $\sup_{g \in \mathcal{G}} M \|R\xi - Ug\|_k^2 \leq \sup_{g \in \mathcal{R}^m} M \|R\xi - Ug\|_k^2$, откуда согласно (6.10) и (6.30) следует, что для любого $U: \mathcal{R}^m \rightarrow \mathcal{R}^k$ $\text{Er}(A, \mathcal{G}, \Sigma_\nu, U) \leq \text{Er}(A, \Sigma_\nu, U)$, т.е. что в смысле определения 2.6.2 $[A, \mathcal{G}, \Sigma_\nu] \preceq [A, \Sigma_\nu]$, а если $\mathcal{G}$ — ограниченная область в $\mathcal{R}^m$, то в отличие от задачи (6.10) в этом случае для любого $R: \mathcal{R}^n \rightarrow \mathcal{R}^k$ $\sup_{g \in \mathcal{G}} \|(RA - U)g\|_k^2 < \infty$ и условие (6.12) не требуется.

Методы решения задачи (6.30) для некоторых $\mathcal{G} \subset \mathcal{R}^m$ см. в [16].

2.6.5.2 Модели $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu]$ и $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu]$.

Если сигнал $g \in \mathcal{R}^n$ в системе I на рис. 2.3 можно моделировать как случайный вектор γ с известным корреляционным оператором $\tilde{\Sigma}_\gamma$ (и неизвестным $\gamma_0 = M\gamma$), то речь идет о модели $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu]$ схемы измерения $\xi = A\gamma + \nu$ и соответствующего ИП, в которой предполагается, что случайные векторы $\gamma \in \mathcal{R}^m$ и $\nu \in \mathcal{R}^n$ независимы. Рассмотрим задачу редукции $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu] \rightarrow [U, \Sigma_U]$ как задачу на минимум

$$\text{Er}(A, \tilde{\Sigma}_\gamma, \Sigma_\nu, U) \stackrel{\text{def}}{=} M \|\tilde{R}_* \xi - U\gamma\|_k^2 = \min_{R: \mathcal{R}^n \rightarrow \mathcal{R}^k} M \|R\xi - U\gamma\|_k^2 \quad (6.31)$$

для с.к. погрешности редукции $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu] \sim> [U, 0]$ к идеальному ИП $[U, 0]$. Если оператор $A\tilde{\Sigma}_\gamma A^* + \Sigma_\nu$ невырожденный, то

$$\tilde{R}_* = U\tilde{\Sigma}_\gamma A^* (A\tilde{\Sigma}_\gamma A^* + \Sigma_\nu)^{-1}, \quad (6.32)$$

$$\begin{aligned} \text{Er}(A, \tilde{\Sigma}_\gamma, \Sigma_\nu, U) &= \text{tr} [U(\tilde{\Sigma}_\gamma - \tilde{\Sigma}_\gamma A^* (A\tilde{\Sigma}_\gamma A^* + \Sigma_\nu)^{-1} A\tilde{\Sigma}_\gamma) U^*], \\ \Sigma_U &= U(\tilde{\Sigma}_\gamma - \tilde{\Sigma}_\gamma A^* (A\tilde{\Sigma}_\gamma A^* + \Sigma_\nu)^{-1} A\tilde{\Sigma}_\gamma) U^*. \end{aligned}$$

Что касается модели $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu]$, в которой известны математическое ожидание $\gamma_0 = M\gamma$ и ковариационный оператор Σ_γ , то в этом

случае задача редукции $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu] \rightarrow [U, \Sigma_U]$ ставится как задача на минимум

$$\text{Er}(A, \gamma_0, \Sigma_\gamma, \Sigma_\nu, U) \stackrel{\text{def}}{=} M \|R_* \xi + r_* - U\gamma\|_k^2 = \min_{\substack{R: \mathcal{R}^n \rightarrow \mathcal{R}^k \\ r \in \mathcal{R}^k}} M \|R\xi + r - U\gamma\|_k^2 \quad (6.33)$$

для с.к. погрешности редукции $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu] \sim > [U, 0]$ к идеальному ИП $[U, 0]$, в которой требуется найти $R_* : \mathcal{R}^n \rightarrow \mathcal{R}^k$ и вектор $r_* \in \mathcal{R}^k$, при которых выходной сигнал $R_* \xi + r_*$ ИП $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu, U]$ — ближайший в с.к. к выходному сигналу $U\gamma$ идеального ИП $[U, 0]$.

Решение задачи (6.33) дается выражениями

$$\begin{aligned} R_* &= U\Sigma_\gamma A^* (A\Sigma_\gamma A^* + \Sigma_\nu)^{-1}, \quad r_* = -(R_* A - U)\gamma_0, \\ \text{Er}(A, \gamma_0, \Sigma_\gamma, \Sigma_\nu, U) &= \text{tr} [U(\Sigma_\gamma - \Sigma_\gamma A^* (A\Sigma_\gamma A^* + \Sigma_\nu)^{-1} A\Sigma_\gamma) U^*], \\ \Sigma_U &= U(\Sigma_\gamma - \Sigma_\gamma A^* (A\Sigma_\gamma A^* + \Sigma_\nu)^{-1} A\Sigma_\gamma) U^* \end{aligned}$$

и, соответственно,

$$R_* \xi + r_* = U\gamma_0 + U\Sigma_\gamma A^* (A\Sigma_\gamma A^* + \Sigma_\nu)^{-1} (\xi - A\gamma_0).$$

Нетрудно показать, что $[A, r_0, \Sigma_\gamma, \Sigma_\nu] \preceq [A, \tilde{\Sigma}_\gamma, \Sigma_\nu] \preceq [A, \Sigma_\nu]$, [16].

2.6.6. Эффект «предварительной обработки» данных измерения. Рассмотрим линейное преобразование

$$\xi \rightarrow B\xi = BA\xi + B\nu, \quad (6.34)$$

иногда называемое «предобработкой» данных измерения ξ , и соответствующее преобразование (редукцию) ИП $[A, \Sigma_\nu] \rightarrow [BA, B\Sigma_\nu B^*]$ ($= [\tilde{A}, \tilde{\Sigma}_\nu]$). Нельзя ли таким путем повысить качество ИП? Ответ, разумеется, отрицательный: *любая линейная «предобработка» (6.34) не может повысить качество ИП, а именно, в итоге любого преобразования $B : \mathcal{R}^n \rightarrow \mathcal{R}^s$ $[A, \Sigma_\nu] \preceq [BA, B\Sigma_\nu B^*]$. Действительно,*

$$\begin{aligned} \inf \{M \|RB\xi - Uf\|^2 | R : \mathcal{R}^s \rightarrow \mathcal{R}^k, RBA = U\} &= \\ = \inf \{M \|\tilde{R}\xi - Uf\|^2 | \tilde{R} : \mathcal{R}^n \rightarrow \mathcal{R}^k, \tilde{R} = RB, \tilde{R}A = U\} &\geq \\ \geq \inf \{M \|\tilde{R}\xi - Uf\|^2 | \tilde{R} : \mathcal{R}^n \rightarrow \mathcal{R}^k, \tilde{R}A = U\}, \end{aligned}$$

где неравенство есть следствие того, что в его правой части $\inf$ вычисляется на множестве операторов $\{\tilde{R} : \mathcal{R}^n \rightarrow \mathcal{R}^k, \tilde{R}A = U\}$, содержащем множество $\{\tilde{R} : \mathcal{R}^n \rightarrow \mathcal{R}^k, \tilde{R}A = U, \tilde{R} = RB\}$ операторов $\tilde{R}$, представимых в виде $\tilde{R} = RB$ и, следовательно, удовлетворяющих условию $\tilde{R} = RB \Rightarrow \forall x \in \mathcal{R}^n Bx = 0 \Rightarrow \tilde{R}x = 0$. Более того, в то время как в рассматриваемом в теореме 2.6.1 случае уравнение $RA = U$ разрешимо при любом U , уравнение $\tilde{R}BA = U$ при некоторых U может не иметь решения. Разумеется, $[A, \Sigma_\nu] \sim [BA, B\Sigma_\nu B^*]$, если и только если оператор $A^* B^* B A > 0$ (невыврожденный). В частности, при $B = (A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1}$, см. Теорему 2.6.1, $[BA, B\Sigma_\nu B^*] = [I, (A^* \Sigma_\nu^{-1} A)^{-1}] \sim [A, \Sigma_\nu]$.

2.6.7. Комбинирование данных независимых измерений. Речь пойдет о проблеме оптимального использования информации, полученной в независимых измерительных экспериментах.

Пусть при исследовании объекта, см. рис. 2.4, выполнены два независимых¹⁾ измерения согласно схемам

$$\xi_i = A_i g_i + \nu_i, \quad i = 1, 2, \quad (6.35)$$

где ИП $[A_i, \Sigma_i]$, $i = 1, 2$, в системе I на рис. 2.3, и соответствующие идеальные ИП $[U_i, 0]$, $i = 1, 2$, измеряют одни и те же характеристики $t \in \mathcal{R}^k$ исследуемого объекта на рис. 2.5, т. е.

$$U_1 g_1 = U_2 g_2 = t, \quad (6.36)$$

где $U_i : \mathcal{R}^{m_i} \rightarrow \mathcal{R}^k$, $i = 1, 2$, а в (6.35) $A_i : \mathcal{R}^{m_i} \rightarrow \mathcal{R}^{n_i}$, $m_i \leq n_i$, g_i — воздействие на ИП $[A_i, \Sigma_i]$ при измерении в системе I, и $k \leq \min(m_1, m_2)$.

Рассмотрим задачу синтеза ИВП, оптимально комбинирующего данные измерений (6.35) на ИП $[A_i, \Sigma_i]$, $i = 1, 2$, при их редукции к виду, свойственному измерению на идеальном ИП параметров $t \in \mathcal{R}^k$ исследуемого объекта.

Покажем, что в рассматриваемой задаче комбинирования данных измерений (6.35) можно использовать результаты редукции ИП $[A_i, \Sigma_i]$, $i = 1, 2$, к соответствующим идеальным ИП $[U_i, 0]$, $i = 1, 2$.

Равенства (6.35) и (6.36) с учетом независимости $g_i \in \mathcal{R}^{m_i}$, $i = 1, 2$, и i , $i = 1, 2$, позволяют редуцировать схемы измерений (6.35) к схеме измерения характеристик $t \in \mathcal{R}^k$ исследуемого объекта

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \rightarrow \begin{pmatrix} R_1 \xi_1 \\ R_2 \xi_2 \end{pmatrix} = \begin{pmatrix} R_1 A_1 g_1 \\ R_2 A_2 g_2 \end{pmatrix} + \begin{pmatrix} R_1 \nu_1 \\ R_2 \nu_2 \end{pmatrix} = \begin{pmatrix} U_1 g_1 \\ U_2 g_2 \end{pmatrix} + \begin{pmatrix} R_1 \nu_1 \\ R_2 \nu_2 \end{pmatrix} = \begin{pmatrix} t \\ t \end{pmatrix} + \begin{pmatrix} R_1 \nu_1 \\ R_2 \nu_2 \end{pmatrix} \equiv \begin{pmatrix} I \\ I \end{pmatrix} t + \begin{pmatrix} R_1 \nu_1 \\ R_2 \nu_2 \end{pmatrix} \quad (6.37)$$

на ИП $\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right]$, где $I : \mathcal{R}^k \rightarrow \mathcal{R}^k$ — единичный оператор, $\tilde{\Sigma}_i = R_i \Sigma_i R_i^*$, $i = 1, 2$, и вследствие независимости $g_i \in \mathcal{R}^{m_i}$, $i = 1, 2$, $R_i A_i = U_i$, $i = 1, 2$. Поэтому

$$R_i = U_i (A_i^* \Sigma_i^{-1} A_i)^{-1} A_i^* \Sigma_i^{-1} + Z_i Z_i^*, \quad Z_i A_i = 0, \quad i = 1, 2, \quad (6.38)$$

¹⁾ Независимость в данном случае означает независимость $g_i \in \mathcal{R}^{m_i}$, $i = 1, 2$, и статистическую независимость ν_i , $i = 1, 2$.

Данные в (6.35) можно считать полученными на ИП $\left[\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \begin{pmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{pmatrix} \right]$.

и при любых $Z_i : \mathcal{R}^{n_i} \rightarrow \mathcal{R}^k$, $i = 1, 2$, качество ИП $\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right]$ при редукции $\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right] \sim > [I, 0]$ характеризуется ковариационным оператором с. к. погрешности редукции

$$\begin{aligned} & \left(\begin{pmatrix} I \\ I \end{pmatrix}^* \begin{pmatrix} \Sigma_{U_1} + Z_1 Z_1^* & 0 \\ 0 & \Sigma_{U_2} + Z_2 Z_2^* \end{pmatrix}^{-1} \begin{pmatrix} I \\ I \end{pmatrix} \right)^{-1} = \\ & = ((\Sigma_{U_1} + Z_1 Z_1^*)^{-1} + (\Sigma_{U_2} + Z_2 Z_2^*)^{-1})^{-1} \geq (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1}, \end{aligned} \quad (6.39)$$

определяющим, в частности, с. к. ошибку редукции

$\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right] \sim > [I, 0]$. Чтобы избежать не относящихся к существу проблемы сложностей, в (6.39) считается, что ковариационные операторы

$$\Sigma_{U_i} \stackrel{\text{def}}{=} U_i(A_i^* \Sigma_i^{-1} A_i)^{-1} U_i^*, \quad i = 1, 2, \quad (6.40)$$

невыврожденные ¹⁾.

Согласно неравенству в (6.39) для достижения наивысшего качества ИП $\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right]$ при редукции $\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \right] \sim > [I, 0]$ преобразование схем (6.35) к схеме (6.37) следует выполнить при ²⁾

$$R_i = R_{*i} = U_i(A_i^* \Sigma_i^{-1} A_i)^{-1} A_i^* \Sigma_i^{-1}, \quad i = 1, 2, \quad (6.41)$$

а измерение характеристик $t \in \mathcal{R}^k$ следует выполнить на ИП

$\left[\begin{pmatrix} I \\ I \end{pmatrix}, \begin{pmatrix} \Sigma_{U_1} & 0 \\ 0 & \Sigma_{U_2} \end{pmatrix} \right]$ по схеме

$$\begin{pmatrix} R_{*1} \xi_1 \\ R_{*2} \xi_2 \end{pmatrix} = \begin{pmatrix} I \\ I \end{pmatrix} t + \begin{pmatrix} R_{*1} \nu_1 \\ R_{*2} \nu_2 \end{pmatrix}, \quad (6.42)$$

¹⁾ В [16] показано, как задача редукции с вырожденным ковариационным оператором может быть сведена к задаче с невырожденным оператором.

²⁾ То есть так, как следовало бы преобразовать каждое измерение в (6.35) при редукции $[A_i, \Sigma_i] \sim > [U_i, 0]$, $i = 1, 2$.

поэтому редукция схемы измерения (6.42) к виду, свойственному измерению $t \in \mathcal{R}^k$ на идеальном ИП $[I, 0]$ определится равенством

$$\begin{aligned} R_{**} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} &\equiv R_* \begin{pmatrix} R_{*1}\xi_1 \\ R_{*2}\xi_2 \end{pmatrix} = \\ &= \left(\begin{pmatrix} I \\ I \end{pmatrix}^* \begin{pmatrix} \Sigma_{U_1}^{-1} & 0 \\ 0 & \Sigma_{U_2}^{-1} \end{pmatrix} \begin{pmatrix} I \\ I \end{pmatrix} \right)^{-1} \begin{pmatrix} I \\ I \end{pmatrix}^* \begin{pmatrix} \Sigma_{U_1}^{-1} & 0 \\ 0 & \Sigma_{U_2}^{-1} \end{pmatrix} \begin{pmatrix} R_{*1}\xi_1 \\ R_{*2}\xi_2 \end{pmatrix} = \\ &= (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1} (\Sigma_{U_1}^{-1} R_{*1}\xi_1 + \Sigma_{U_2}^{-1} R_{*2}\xi_2) \equiv \\ &\equiv (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1} (\Sigma_{U_1}^{-1} R_{*1}; \Sigma_{U_2}^{-1} R_{*2}) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = t + R_{**} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \end{aligned} \quad (6.43)$$

где

$$R_* = (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1} (\Sigma_{U_1}^{-1}; \Sigma_{U_2}^{-1}) : \mathcal{R}^k \oplus \mathcal{R}^k \rightarrow \mathcal{R}^k, \quad (6.44)$$

а оператор, определяющий оптимальное правило комбинирования данных (6.35), дается равенством

$$R_{**} = (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1} (\Sigma_{U_1}^{-1} R_{*1}; \Sigma_{U_2}^{-1} R_{*2}) : \mathcal{R}^{n_1} \oplus \mathcal{R}^{n_2} \rightarrow \mathcal{R}^k. \quad (6.45)$$

Ковариационный оператор с.к. погрешности редукции (6.43) есть $(\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1}$, а с.к. погрешность редукции

$$\text{tr} (\Sigma_{U_1}^{-1} + \Sigma_{U_2}^{-1})^{-1} < \text{tr} \Sigma_{U_i}, \quad i = 1, 2.$$

Понятно, что для любой системы данных $\xi_i = A_i g_i + \nu_i$, $i = 1, \dots, s$, независимых измерений оператор оптимального их комбинирования

$$R_{**} = (\Sigma_{U_1}^{-1} + \dots + \Sigma_{U_s}^{-1})^{-1} (\Sigma_{U_1}^{-1} R_{*1}; \dots; \Sigma_{U_s}^{-1} R_{*s}) : \mathcal{R}^{n_1} \oplus \dots \oplus \mathcal{R}^{n_s} \rightarrow \mathcal{R}^k$$

а неравенства $\text{tr} (\Sigma_{U_1}^{-1} + \dots + \Sigma_{U_s}^{-1})^{-1} < \text{tr} \Sigma_{U_i}$, $i = 1, \dots, s$, определяют выигрыш в точности при комбинировании данных.

2.6.8. Собственные базисы ИП $[A, \Sigma_\nu]$ и ИВП $[A, \Sigma_\nu, U]$.

Свойства ИП $[A, \Sigma_\nu]$ удобно охарактеризовать, определив понятия его *собственных базисов*. С этой целью рассмотрим задачу на собственные значения для оператора $A^* \Sigma_\nu^{-1} A : \mathcal{R}^m \rightarrow \mathcal{R}^m$

$$A^* \Sigma_\nu^{-1} A e_j = \delta_j^2 e_j, \quad j = 1, \dots, m, \quad \delta_1^2 \geq \dots \geq \delta_m^2 > 0. \quad (6.46)$$

Поскольку $A^* \Sigma_\nu^{-1} A$ — самосопряженный оператор, его собственные векторы $e_j \in \mathcal{R}^m$, $j = 1, \dots, m$, можно выбрать так, чтобы они образовали ортонормированный базис $\mathcal{R}^m$, а так как оператор $A^* \Sigma_\nu^{-1} A > 0$, его собственные значения > 0 ; условимся считать их упорядоченными, как в (6.46).

Замечание 2.6.3. Если выполнено условие в (6.28*), означающее, что $[A, \Sigma_\nu] \preccurlyeq [\tilde{A}, \tilde{\Sigma}_\nu]$, то $\delta_j^2 \geq \tilde{\delta}_j^2$, $j = 1, \dots, m$, где, как в (6.46), $\tilde{A}^* \tilde{\Sigma}_\nu^{-1} \tilde{A} \tilde{e}_j = \tilde{\delta}_j^2 \tilde{e}_j$, $j = 1, \dots, m$, $\tilde{\delta}_1^2 \geq \dots \geq \tilde{\delta}_m^2 > 0$, иными словами, неравенства $\delta_j^2 \geq \tilde{\delta}_j^2$, $j = 1, \dots, m$, суть необходимое условие $[A, \Sigma] \preccurlyeq [\tilde{A}, \tilde{\Sigma}_\nu]$.

Так как $A^* \Sigma_\nu^{-1} A = (\Sigma_\nu^{-1/2} A)^* \Sigma_\nu^{-1/2} A$, то векторы

$$s_i \stackrel{\text{def}}{=} \delta_i^{-1} \Sigma_\nu^{-1/2} A e_i, \quad i = 1, \dots, m, \quad (6.47)$$

образуют ортонормированную систему в $\mathcal{R}^n$, поскольку согласно (6.46), (6.47)

$$\begin{aligned} (s_i, s_k)_n &= \delta_i^{-1} \delta_k^{-1} (\Sigma_\nu^{-1/2} A e_i, \Sigma_\nu^{-1/2} A e_k) = \delta_i^{-1} \delta_k^{-1} (e_i, A^* \Sigma_\nu^{-1} A e_k) = \\ &= \delta_i^{-1} \delta_k (e_i, e_k) = \begin{cases} 0, & \text{если } i \neq k, \\ 1, & \text{если } i = k, \end{cases} \quad i, k = 1, \dots, m. \end{aligned} \quad (6.48)$$

Заметим, что согласно (6.46), (6.47) $A^* \Sigma_\nu^{-1/2} s_i = \delta_i^{-1} A^* \Sigma_\nu^{-1} A e_i = \delta_i e_i$, поэтому

$$e_i = \delta_i^{-1} A^* \Sigma_\nu^{-1/2} s_i, \quad i = 1, \dots, m, \quad (6.49)$$

и согласно (6.47), (6.49) $s_1, \dots, s_m$ — ортонормированная система собственных векторов оператора $\Sigma_\nu^{-1/2} A A^* \Sigma_\nu^{-1/2} : \mathcal{R}^n \rightarrow \mathcal{R}^n$

$$\Sigma_\nu^{-1/2} A A^* \Sigma_\nu^{-1/2} s_i = \delta_i \Sigma_\nu^{-1/2} A e_i = \delta_i^2 s_i, \quad i = 1, \dots, m. \quad (6.50)$$

Так как ранги операторов $A^* \Sigma_\nu^{-1} A = (\Sigma_\nu^{-1/2} A)^* \Sigma_\nu^{-1/2} A$ и $\Sigma_\nu^{-1/2} A A^* \Sigma_\nu^{-1/2}$ совпадают и равны m , ортонормированную систему $s_1, \dots, s_m$ можно дополнить до ортонормированного базиса $s_1, \dots, s_m, s_{m+1}, \dots, s_n$ в $\mathcal{R}^n$, добавив ортонормированную систему $s_{m+1}, \dots, s_n$ векторов, каждый из которых ортогонален $s_1, \dots, s_m$ и является собственным вектором оператора $\Sigma_\nu^{-1/2} A A^* \Sigma_\nu^{-1/2}$, отвечающим его нулевому собственному значению,

$$\Sigma_\nu^{-1/2} A A^* \Sigma_\nu^{-1/2} s_i = 0, \quad i = m+1, \dots, n. \quad (6.51)$$

Определение 2.6.3. Ортонормированный базис $e_1, \dots, e_m$, определенный в (6.46), называется *собственным базисом* ИП $[A, \Sigma_\nu]$ в $\mathcal{R}^m$, [15].

Ортонормированный базис $s_1, \dots, s_n$, определенный в (6.50), (6.51), назовем *собственным базисом* ИП $[A, \Sigma_\nu]$ в $\mathcal{R}^n$.

Определение 2.6.4. Ортонормированные базисы $\{u_1, \dots, u_k\} \subset \mathcal{R}^k$ и $\{v_1, \dots, v_n\} \subset \mathcal{R}^n$ называются *собственными базисами* ИВП $[A, \Sigma_\nu, U]$ в $\mathcal{R}^k$ и соответственно в $\mathcal{R}^n$, если

$$\begin{aligned} U S^{-1} U^* u_i &= \omega_i^2 u_i, \quad i = 1, \dots, k, \quad \omega_1^2 \geq \dots \geq \omega_l^2 > \omega_{l+1}^2 = \dots = \omega_k^2 = 0, \\ \Sigma_\nu^{-1/2} A S^{-1} U^* U S^{-1} A^* \Sigma_\nu^{-1/2} v_j &= \omega_j^2 v_j, \quad j = 1, \dots, n, \\ \omega_1^2 \geq \dots \geq \omega_l^2 > \omega_{l+1}^2 = \dots = \omega_n^2 = 0, \quad l &= \text{rank } U, \\ u_i &= \omega_i^{-1} U S^{-1} A^* \Sigma_\nu^{-1/2} v_i, \quad v_i = \omega_i^{-1} \Sigma_\nu^{-1/2} A S^{-1} U^* u_i, \quad i = 1, \dots, l, \\ \text{Er}(A, \Sigma_\nu, U) &= \text{tr}(U S^{-1} U^*) = \sum_{i=1}^l \omega_i^2, \quad S \stackrel{\text{def}}{=} A^* \Sigma_\nu^{-1} A. \end{aligned} \quad (6.52)$$

Собственные базисы ИП $[A, \Sigma_\nu]$ и ИВП $[A, \Sigma_\nu, U]$ позволяют полностью охарактеризовать свойства решений задач редукции $[A, \Sigma_\nu] \sim \sim [I, 0]$ и $[A, \Sigma_\nu] \sim \sim [U, 0]$ соответственно.

2.6.8.1 Редукция $[A, \Sigma_\nu] \sim \sim [I, 0]$ в собственных базисах ИП $[A, \Sigma_\nu]$.

Перепишем в координатном виде равенство (6.8), предварительно заменив его на эквивалентное $\Sigma_\nu^{-1/2}\xi = \Sigma_\nu^{-1/2}Ag + \Sigma_\nu^{-1/2}\nu$, т.е. в виде равенств

$$\xi_i = \delta_i g_i + \nu_i, \quad i = 1, \dots, m, \quad \xi_i = \nu_i, \quad i = m+1, \dots, n, \quad (6.53)$$

где согласно (6.46)–(6.51)

$$\begin{aligned} \xi_i &= (\Sigma_\nu^{-1/2}\xi, s_i)_n, \quad \nu_i = (\Sigma_\nu^{-1/2}\nu, s_i)_n, \quad i = 1, \dots, n, \\ \delta_i g_i &= (\Sigma_\nu^{-1/2}Ag, s_i)_n = \delta_i (g, e_i)_m, \\ i &= 1, \dots, m, \quad \delta_i g_i = 0, \quad i = m+1, \dots, n, \end{aligned} \quad (6.54)$$

и

$$\begin{aligned} \text{Er}(A, \Sigma_\nu, I) &= M\|\hat{g} - g\|_m^2 = \text{tr}(A^* \Sigma_\nu^{-1} A)^{-1} = \\ &= \sum_{j=1}^m (e_j, (A^* \Sigma_\nu^{-1} A)^{-1} e_j)_m = \sum_{j=1}^m \delta_j^{-2}. \end{aligned}$$

Заметим, что в таком представлении

$$\begin{aligned} M\nu_i \nu_j &= M(s_i, \Sigma_\nu^{-1/2}\nu)_n (s_j, \Sigma_\nu^{-1/2}\nu)_n = \\ &= (s_i, s_j) = \begin{cases} 0, & \text{если } i \neq j, \\ 1, & \text{если } i = j, \end{cases} \quad i, j = 1, \dots, n, \end{aligned}$$

т.е. в (6.53)–(6.55) "энергии координат шума" $M\nu_i^2 = 1, i = 1, \dots, n$.

При этом согласно теореме 2.6.1 и (6.54)

$$\hat{g} = \sum_{j=1}^m (\hat{g}, e_j)_m e_j = \sum_{i=1}^m (g_i + \delta_i^{-1} \nu_i) e_i, \quad (6.55)$$

где в силу (6.47), (6.49)

$$(\hat{g}, e_j) = \delta_j^{-1} (\Sigma_\nu^{-1/2}\xi, s_j)_n = (g, e_j)_m + \delta_j^{-1} (\Sigma_\nu^{-1/2}\nu, s_j)_n, \quad j = 1, \dots, m. \quad (6.56)$$

Поэтому оценка $\hat{g}$ определяется ортогональной проекцией данных измерения $\Sigma_\nu^{-1/2}\xi$ на $\mathcal{L}(s_1, \dots, s_m)$, а их ортогональная проекция $\sum_{i=m+1}^n (\Sigma_\nu^{-1/2}\xi, s_i) s_i$ на $\mathcal{L}(s_{m+1}, \dots, s_n)$ не зависит от g и позволяет проверить состоятельность модели ИП, см. замечание 2.6.2.

Выражения (6.55) и (6.54) позволяют объяснить на первый взгляд парадоксальный феномен, а именно, наилучшая в с.к. оценка $\hat{g}$ нередко полностью разрушена шумом и лишь отдаленно напоминает g , в то время как $\xi = Ag + \nu \approx g + \nu$ «воспроизводит g » существенно лучше, чем $\hat{g}$. Покажем, в каком случае и как проявляется разруша-

ющее действие шума на редукцию $[A, \Sigma_\nu] \sim> [I, 0]$, т. е. на оценку $\hat{g}$ сигнала g (6.55).

Пусть, например, $A = A^*$, $\Sigma_\nu = I$, $Ae_j = \alpha_j e_j$, где $|\alpha_j| \approx 1$, $i = 1, \dots, m-1$, $|\alpha_m| \ll 1$. В таком случае $\delta_j^2 = \alpha_j^2 \approx 1$, $j = 1, \dots, m-1$, $\delta_m^2 = \alpha_m^2 \ll 1$ и $s_j = e_j$, $j = 1, \dots, m$. Поэтому, если $g \approx \sum_{j=1}^{m-1} g_j e_j$, то $\xi = Ag + \nu \approx g + \nu$, $\hat{g} \approx g + \nu + (\alpha_m^{-1} - 1)\nu_m e_m$, и если сигнал g существенно превышает шум ν , $\|g\|_m \gg \|\nu\|_{n=m}$, то $\xi \approx g$, но так как $|\alpha_m| \ll 1$, то $\|\hat{g} - g\|_m \sim |\alpha_m^{-1}|$, что говорит о существенном отличии $\hat{g}$ от g .

Наиболее эффективный способ противостоять разрушающему действию шума на редукцию $[A, \Sigma_\nu] \sim> [I, 0]$ основан на следующем факте, характеризующем роль собственных базисов ИП $[A, \Sigma_\nu]$: для любого $t = 1, \dots, m$ ортогональная проекция $\hat{g}$ (6.55) $\hat{g}_{(t)} = \sum_{i=1}^t (g_i + \delta_i^{-1} \nu_i) e_i$

на $\mathcal{L}(e_1, \dots, e_t)$ поражена шумом $\nu_{(t)} = \sum_{i=1}^t \delta_i^{-1} \nu_i e_i$, интенсивность

которого $M\|\nu_{(t)}\|_m^2 = \sum_{i=1}^t \delta_i^{-2}$ не больше, чем интенсивность шума, искажающего любую другую t -мерную ортогональную составляющую $\hat{g}$, см. теорему 2.6.2.

2.6.8.2 Редукция $[A, \Sigma_\nu] \sim> [U, 0]$ в собственных базисах ИВП $[A, \Sigma_\nu, U]$.

Согласно соотношениям (6.52) и теореме 2.6.1 выходной сигнал ИВП $[A, \Sigma_\nu, U]$

$$R_* \xi \stackrel{\text{def}}{=} \widehat{Ug} = Ug = Ug + US^{-1} A^* \Sigma_\nu^{-1} \nu \sum_{i=1}^k u_i ((Ug, u_i)_k + \omega_i (\Sigma_\nu^{-1/2} \nu, v_i)_n), \quad (6.57)$$

где $M(\Sigma_\nu^{-1/2} \nu, v_i)_n = 0$, $M(\Sigma_\nu^{-1/2} \nu, v_i)^2 = 1$, $i = 1, \dots, n$, и, следовательно, ω_i^2 — уровень шума, искажающего составляющую $(Ug, u_i)_k u_i$ выходного сигнала Ug идеального ИП $[U, 0]$, $i = 1, \dots, l$, $((Ug, u_i) = 0$, $i = l+1, \dots, k$).

Согласно (6.52), (6.57) t -мерная ортогональная проекция $\widehat{Ug}$

$$(\widehat{Ug})_{(t)} = \sum_{i=l-t+1}^l ((Ug, u_i)_k + \omega_i (\Sigma_\nu^{-1/2} \nu, v_i)_n) u_i \quad (6.58)$$

на $\mathcal{L}(u_l, \dots, u_{l-t+1})$ поражена шумом $\mu_{(t)} = \sum_{i=l-t+1}^l \omega_i (\Sigma_\nu^{-1/2} \nu, v_i)_n u_i$,

интенсивность которого $M\|\mu_{(t)}\|_k^2 = \omega_l^2 + \dots + \omega_{l-t+1}^2$ не больше, чем интенсивность шума, искажающего любую другую t -мерную ортогональную составляющую $\widehat{Ug}$, $t = 1, \dots, l$. Этот факт характеризует

роль собственных базисов ИВП $[A, \Sigma_\nu, U]$ при анализе его качества и предельных возможностей как средства измерений.

2.6.9. Эффективные ранги ИП $[A, \Sigma_\nu]$ и ИВП $[A, \Sigma_\nu, U]$. Если $\tilde{\Pi}_t : \mathcal{R}^m \rightarrow \mathcal{R}^m$ — ортогональный проектор на произвольное линейное подпространство $\mathcal{R}^t$, то в задаче редукции $[A, \Sigma_\nu] \rightarrow [\tilde{\Pi}_t, \Sigma_{\tilde{\Pi}_t}]$

$$\begin{aligned} \text{Er}(A, \Sigma_\nu, \tilde{\Pi}_t) &= M \|\tilde{\Pi}_t(\hat{g} - g)\| = \text{tr } \tilde{\Pi}_t(A^* \Sigma_\nu^{-1} A)^{-1} = \\ &= \sum_{i=1}^m \delta_i^{-2} (e_i, \tilde{\Pi}_t e_i)_m. \end{aligned} \quad (6.59)$$

Согласно (6.59) составляющая $\tilde{\Pi}_t g$ любого вектора $g \in \mathcal{R}^m$ оценивается $\tilde{\Pi}_t \hat{g}$ с с.к. погрешностью $\text{tr}(\tilde{\Pi}_t(A^* \Sigma_\nu^{-1} A)^{-1})$, минимальной в классе всех линейных минимаксных оценок $\hat{\Pi}_t g$. Иными словами, линейное подпространство $\mathcal{R}^t$ сигналов, измеренных ИП $[A, \Sigma_\nu]$ по схеме (6.8), поражено шумом $\tilde{\Pi}_t(A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \nu$, интенсивность которого определена в (6.59). В этой связи представляет интерес задача отыскания t -мерного линейного подпространства $\mathcal{R}^m$, минимально пораженного шумом, т. е. следующая задача на минимум:

$$\text{Er}(A, \Sigma_\nu, \Pi) \sim \min_{\Pi, \text{rank } \Pi=t}, \quad (6.60)$$

в которой $\Pi : \mathcal{R}^m \rightarrow \mathcal{R}^m$ — ортогональный проектор. Решение этой задачи хорошо известно, см., например, [15].

Теорема 2.6.2. Пусть $e_1, \dots, e_m$ — собственный базис ИП $[A, \Sigma_\nu]$ в $\mathcal{R}^m$, определенный в (6.46). Тогда в задаче (6.60) минимум достигается на ортогональном проекторе Π_t на $\mathcal{L}(e_1, \dots, e_t)$ и равен $\sum_{j=1}^t \delta_j^{-2}$, $t = 1, \dots, m$.

В данном случае основной результат состоит в следующем.

Теорема 2.6.3. Собственные базисы ИП $[A, \Sigma_\nu]$ порождают последовательности расширяющихся линейных подпространств $\mathcal{L}(e_1) \subset \mathcal{L}(e_1, e_2) \subset \dots \subset \mathcal{L}(e_1, \dots, e_m) = \mathcal{R}^m$ и $\mathcal{L}(s_1) \subset \mathcal{L}(s_1, s_2) \subset \dots \subset \mathcal{L}(s_1, \dots, s_m) = \mathcal{R}^n$ такие, что для каждого $t = 1, \dots, m$

- ортогональная проекция $\Pi_t g$ на $\mathcal{L}(e_1, \dots, e_t)$ сигнала $g \in \mathcal{R}^m$, измеренного посредством ИП $[A, \Sigma_\nu]$ по схеме (6.8), оценивается $\Pi_t \hat{g}$ не менее точно, чем любая другая t -мерная ортогональная составляющая $\tilde{\Pi}_t g$ оценивается $\tilde{\Pi}_t \hat{g}$;
- оценка $\Pi_t \hat{g}$ определяется ортогональной проекцией $\sum_{i=1}^t (\Sigma_\nu^{-1/2} \xi, s_i)_n s_i$ результата измерения $\sigma_\nu^{-1/2} \xi$ на $\mathcal{L}(s_1, \dots, s_t)$.

Доказательство следует из (6.46), (6.50), (6.51), (6.55) и теоремы 2.6.2. Образно говоря, линейное подпространство $\mathcal{L}(e_1, \dots, e_t)$ поражено шумом, интенсивности $\text{Er}(A, \Sigma_\nu, \Pi_t) = \sum_{j=1}^t \delta_j^{-2}$, ибо ортогональная

составляющая $\Pi_t g \in \mathcal{L}(e_1, \dots, e_t)$ любого вектора $g \in \mathcal{R}^m$ оценивается с с.к. погрешностью $\sum_{j=1}^t \delta_j^{-2}$, не превосходящей с.к. погрешности (6.59) оценивания его ортогональной проекции $\tilde{\Pi}_t g$ на любое подпространство $\mathcal{R}^m$ той же размерности t . Последнее утверждение является следствием того, что

$$\sum_{j=1}^t \delta_j^{-2} = \inf \{ \text{Er}(A, \Sigma_\nu, \Pi) | \Pi : \mathcal{R}^m \rightarrow \mathcal{R}^m, \text{rank } \Pi \geq t \}, \quad (6.61)$$

где Π — ортогональный проектор, $\text{Er}(A, \Sigma_\nu, \Pi) = \text{tr}(\Pi(A^* \Sigma_\nu^{-1} A)^{-1})$. Равенство (6.61) следует из теоремы Пуанкаре, см. §3 гл. 2 в [15].

Эти факты позволяют ввести *понятие эффективного ранга* ИП $[A, \Sigma_\nu]$ как одной из основных характеристик его качества. Как известно, ранг линейного оператора $B : \mathcal{R}^m \rightarrow \mathcal{R}^n$ определяется как размерность пространства $\mathcal{R}(B) = \{Bg, g \in \mathcal{R}^m\}$ его значений и равен максимальной размерности *составляющей вектора* $g \in \mathcal{R}^m$, *которая определяется значением* $Bg \in \mathcal{R}(B) \subset \mathcal{R}^n$. Но если значение Bg , $g \in \mathcal{R}^m$, точно не известно, а наблюдаемы лишь значения случайного вектора $\xi = Bg + \nu$, то «максимальная размерность составляющей $g \in \mathcal{R}^m$ », «определенной значением ξ », будет зависеть от того, с какой точностью эта составляющая g должна быть определена [16].

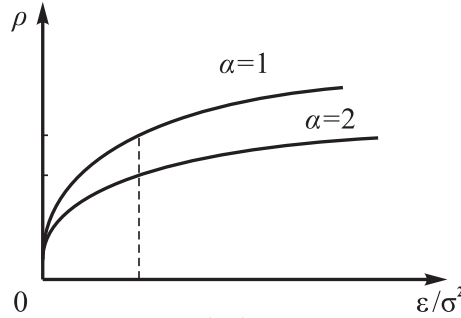


Рис. 2.9. Эффективный ранг ИП $[A^{(\alpha, \beta)}, \sigma^2 I]$, см. пример 2.6.2. Для каждого $\varepsilon > 0$ $\rho_{[A^{(\alpha, \beta)}, \sigma^2 I]}$ — размерность *оптимальной конечномерной аппроксимации* ИП $[A^{(\alpha, \beta)}, \sigma^2 I]$.

Определение 2.6.5. Эффективным рангом ИП $[A, \Sigma_\nu]$ называется функция $\rho_{[A, \Sigma_\nu]}(\cdot) : [0, \infty] \rightarrow \{0, 1, \dots, m\}$, определенная равенством

$$\rho_{[A, \Sigma_\nu]}(\varepsilon) = \begin{cases} \max\{k \in \{1, \dots, m\} | \sum_{j=1}^k \delta_j^{-2} \leq \varepsilon\}, & \text{если } \delta_1^{-2} \leq \varepsilon, \\ 0, & \text{если } \delta_1^{-2} > \varepsilon, \varepsilon \in [0, \infty), \end{cases}$$

где $\delta_1^2 \geq \delta_2^2 \geq \dots \delta_m^2 > 0$, см. (6.46).

Ее значение $\rho_{[A, \Sigma_\nu]}(\varepsilon)$ равно максимальной размерности ортогональной составляющей $g \in \mathcal{R}^m$, которая может быть оценена с с.к. погрешностью, не превосходящей $\varepsilon \geq 0$, на основе данных измерения, выполненного по схеме $\xi = Ag + \nu$ на ИП $[A, \Sigma_\nu]$. Согласно теореме 2.6.3 эта составляющая равна ортогональной проекции g на $\mathcal{L}(e_1, \dots, e_{\rho_{[A, \Sigma_\nu]}(\varepsilon)})$.

Отметим следующие свойства эффективного ранга ИП $[A, \Sigma_\nu]$.

1. Функция $\rho_{[A, \Sigma_\nu]}(\cdot) : [0, \infty) \rightarrow \{0, 1, \dots, m\}$ монотонно не убывает, причем $\lim_{\varepsilon \rightarrow \infty} \rho_{[A, \Sigma_\nu]}(\varepsilon) = \text{rank } A$.

2. Если $[A, \Sigma_\nu] \preceq [\tilde{A}, \tilde{\Sigma}_\nu]$, то $\rho_{[A, \Sigma_\nu]}(\varepsilon) \geq \rho_{[\tilde{A}, \tilde{\Sigma}_\nu]}(\varepsilon)$, $\varepsilon \in [0, \infty)$, т.е. чем выше качество ИП $[A, \Sigma_\nu]$, тем больше его эффективный ранг, см. гл. 10 [16].

Определение 2.6.6. Эффективным рангом ИВП $[A, \Sigma_\nu, U]$ называется функция

$$\rho_{[A, \Sigma_\nu, U]}(\varepsilon) = \begin{cases} \max\{t \in \{1, \dots, l\} \mid \sum_{i=l-t+1}^l \omega_i^2 \leq \varepsilon\}, & \text{если } \omega_l^2 \leq \varepsilon, \\ 0, & \text{если } \omega_l^2 > \varepsilon, \end{cases}$$

$0 \leq \varepsilon \leq \infty$.

2.6.10. Методы инструментального синтеза линейного ИП на ИВП. На практике стремление к точному синтезу выходного сигнала идеального ИП $[\tilde{U}, 0]$ редукцией $[A, \Sigma_\nu] \sim [\tilde{U}, 0]$ за редкими исключениями не может быть признано вполне оправданным. Прежде всего, в силу возможной неточности модели $[A, \Sigma_\nu]$ условие точного синтеза $RA = \tilde{U}$, фактически, не может гарантировать, что $\sup_g M \|R\xi - \tilde{U}g\|^2 < \infty$. Кроме того, точный синтез выходного сигнала ИП $[\tilde{U}, 0]$ на ИВП $[A, \Sigma_\nu, U]$ может сопровождаться неприемлемо большим шумом или, наконец, может быть принципиально невозможен [15].

Эти соображения определяют принципиально другой взгляд на ИВП как на средство измерения, согласно которому выходной сигнал ИВП определяется не как наиболее точная версия выходного сигнала идеального ИП $[\tilde{U}, 0]$, а как выходной сигнал ИП $[U, 0]$, ближайшего к ИП $[\tilde{U}, 0]$, при тех или иных требованиях к его качеству и качеству его выходного сигнала.

В этом разделе речь пойдет о семействе задач

$$\inf\{\|RA - \tilde{U}\|_2^2 \mid R : \mathcal{R}^n \rightarrow \mathcal{R}^k, M\|R\nu\|^2 \leq \varepsilon\} = \|R_\varepsilon A - \tilde{U}\|_2^2 \stackrel{\text{def}}{=} \rho_\varepsilon, \quad \varepsilon \geq 0, \quad (6.62)$$

решение R_ε каждой из которых определит сигнал $R_\varepsilon \xi$ как искаженный шумом $R_\varepsilon \nu$ выходной сигнал ИП $[R_\varepsilon A, 0]$, самого близкого к ИП $[\tilde{U}, 0]$ при условии $M\|R_\varepsilon \nu\|^2 \leq \varepsilon$, $\varepsilon \geq 0$. В монографии [15] показано, что для

любого $\overset{\circ}{U} : \mathcal{R}^m \rightarrow \mathcal{R}^k$ и $0 < \varepsilon < \text{tr}(\overset{\circ}{U}(A^*\Sigma_\nu^{-1}A)^{-1}\overset{\circ}{U}^*)$ $R_\varepsilon = R(\omega)|_{\omega=\omega_\varepsilon}$, где

$$R(\omega) = \overset{\circ}{U}A^*(AA^* + \omega\Sigma_\nu)^{-1} = \overset{\circ}{U}(A^*\Sigma_\nu^{-1}A + \omega I)^{-1}A^*\Sigma_\nu^{-1} \quad (6.63)$$

и ω_ε — корень уравнения $\text{tr}(R(\omega)\Sigma_\nu R^*(\omega)) = \varepsilon$. Функции $h(\omega) = \text{tr}(R(\omega)\Sigma_\nu R^*(\omega))$, $q(\omega) = \|R(\omega)A - \overset{\circ}{U}\|_2^2 \stackrel{\text{def}}{=} \text{tr}((R(\omega)A - \overset{\circ}{U})(R(\omega)A - \overset{\circ}{U})^*)$, $0 \leq \omega < \infty$, определяют *оперативную характеристику семейства задач* (6.62), для любого $\omega > 0$ связывающую уровень шума $h(\omega)$ и невязку $q(\omega)$ синтеза $\overset{\circ}{U}$.

Семейство операторов $R(\omega)$, $0 \leq \omega < \infty$, (6.63) определяет класс так называемых гребневых оценок [30].

Замечание 2.6.4. Хотя при $\omega > 0$ $\sup M\|R(\omega)\xi - \overset{\circ}{U}g\|^2 = \infty$, тем не менее для любого сигнала $g \in \mathcal{R}^m$ можно указать $\omega > 0$, при котором $M\|R(\omega)\xi - \overset{\circ}{U}g\|^2 < \text{tr}(\overset{\circ}{U}(A^*\Sigma^{-1}A)^{-1}\overset{\circ}{U}^*)$, [16].

Замечание 2.6.5. Оператор $R(\omega)$ (6.63) совпадает с $\tilde{R}_*$ в (6.32) редукции $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu] \sim > [\overset{\circ}{U}, 0]$, если $\tilde{\Sigma}_\gamma = \omega^{-1}I$. Это означает, что $R_\varepsilon\xi$ можно считать наиболее точным в с. к. приближением $\overset{\circ}{U}g$, если g определен как случайный вектор с корреляционным оператором $\tilde{\Sigma}_\gamma = \omega_\varepsilon^{-1}I$.

Замечание 2.6.6. В подразделе 2.6.8.1 отмечена «неустойчивость редукции» $[A, \Sigma_\nu] \sim > [I, 0]$, проявляющаяся в «разрушающем» действии шума. В данном случае этот эффект может быть существенно ослаблен, поскольку в базисах e_i , $i = 1, \dots, m$, (6.46) и s_j , $j = 1, \dots, n$, (6.50), (6.51), в которых записано равенство $R_*\xi = (A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1}\xi = \hat{g} = \sum_{i=1}^m (g_i + \delta_i^{-1}\nu_i)e_i$ (6.55), теперь для $R(\omega)$

(6.63) (в этих же базисах) $R(\omega)\xi = \sum_{i=1}^m (\delta_i^2 g_i + \delta_i \nu_i)(\delta_i^2 + \omega)^{-1}e_i$. По этому в то время как при $\delta_i \rightarrow 0$ в (6.55) $M(g_i + \delta_i^{-1}\nu_i)^2 \rightarrow \infty$, $M((\delta_i^2 g_i + \delta_i \nu_i)/(\delta_i^2 + \omega))^2 \rightarrow 0$ при любом $\omega > 0$.

В тех случаях, когда, с одной стороны, при приемлемом уровне шума не больше ε невязка синтеза $\|R_\varepsilon A - \overset{\circ}{U}\|_2^2$ в (6.62) оказывается слишком большой, а с другой — можно априори указать класс $\mathcal{U}$ операторов $U : \mathcal{R}^m \rightarrow \mathcal{R}^k$, определяющих ИП $[U, 0]$ "приемлемого" качества (близких по качеству к $[\overset{\circ}{U}, 0]$), задачу синтеза следует поставить как семейство задач, в которых при каждом ε определяется ИП $[U_\varepsilon, 0]$, $U_\varepsilon \in \mathcal{U}$, и ближайший к нему ИП $[R_\varepsilon A, R_\varepsilon \Sigma_\nu R_\varepsilon^*]$, такие что

$$\|U_\varepsilon - R_\varepsilon A\|_2^2 = \inf\{\|U - RA\|_2^2 | U \in \mathcal{U}, R : \mathcal{R}^n \rightarrow \mathcal{R}^k, M\|R\nu\|^2 \leq \varepsilon\}, \varepsilon \geq 0. \quad (6.64)$$

В каждой из задач (6.62) или (6.64) синтезируется не выходной сигнал идеального ИП, а ИП, ближайший к идеальному, удовлетворяющий

тем или иным требованиям к его качеству как средства измерения. Задачи (6.64) и примеры их практических постановок рассмотрены в [15, 16].

2.6.11. Гаусовский ИП $[A, \Sigma_\nu]$. Рассмотрим модель ИП $[A, \Sigma_\nu]$, в которой кроме того, что $M\nu = 0$ и известен ковариационный оператор Σ_ν , известно также, что $\nu \sim N(0, \Sigma_\nu)$. Посмотрим, что это дает.

2.6.11.1 Доверительный (оценивающий) эллипсоид в $\mathcal{R}^m$. Согласно схеме измерения

$$\xi = Ag + \nu \quad (6.65)$$

для наилучшей в с.к. оценки $\hat{g}$ сигнала g в теореме 2.6.1 получено выражение

$$\hat{g} = (A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \xi = g + (A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \nu \quad (6.66)$$

и найден ковариационный оператор $\hat{g}$

$$\Sigma_{\hat{g}} = M(\hat{g} - g)(\hat{g} - g)^* = (A^* \Sigma_\nu^{-1} A)^{-1}. \quad (6.67)$$

В рассматриваемом случае гауссовского ИП это означает, что ¹⁾ $\hat{g} - g = (A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \nu \sim N(0, (A^* \Sigma_\nu^{-1} A)^{-1})$ и, следовательно

$$(A^* \Sigma_\nu^{-1} A)^{1/2} (\hat{g} - g) \sim N(0, I), \\ \|(A^* \Sigma_\nu^{-1} A)^{1/2} (\hat{g} - g)\|_m^2 \equiv ((A^* \Sigma_\nu^{-1} A)(\hat{g} - g), (\hat{g} - g))_m \sim \chi_m^2. \quad (6.68)$$

Выберем $\alpha \in (0, 1)$, например, $\alpha = 0.01$ и определим $\varepsilon > 0$ так, чтобы $P(\chi_m^2 \leq \varepsilon) = 1 - \alpha = 0.99$. Тогда случайный эллипсоид

$$\mathcal{E}_{\hat{g}, 1-\alpha} = \{g \in \mathcal{R}^m, \|(A^* \Sigma_\nu^{-1} A)^{1/2} (\hat{g} - g)\|_m^2 \leq \varepsilon\} \quad (6.69)$$

с центром $\hat{g} \in \mathcal{R}^m$ покрывает истинное значение $g \in \mathcal{R}^m$ с вероятностью $1 - \alpha = 0.99$ и называется доверительным с уровнем доверия $1 - \alpha$.

Заметим, что в базисе модели $\mathcal{E}_{\hat{g}, 1-\alpha} = \{g \in \mathcal{R}^m, \sum_{i=1}^m \delta_i^2 (\hat{g}_i - g_i)^2 \leq \varepsilon\}$, где $g_i = (g, e_i)_m$, $\hat{g}_i = (\hat{g}, e_i)_m$, $i = 1, \dots, m$.

2.6.11.2 Интервальные оценки для координат g . Согласно равенству (6.66) в некотором ортонормированном базисе $\mathcal{R}^m$ i -ая координата

$$\hat{g}_i = ((A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} (Ag + \nu))_i = g_i + ((A^* \Sigma_\nu^{-1} A)^{-1} A^* \Sigma_\nu^{-1} \nu)_i,$$

и в силу (6.67), (6.68) $\hat{g}_i - g_i \sim N(0, (A^* \Sigma_\nu^{-1} A)^{-1}_{ii})$, $i = 1, \dots, m$, где $(A^* \Sigma_\nu^{-1} A)^{-1}_{ii}$ — i, i -матричный элемент матрицы оператора ²⁾ $(A^* \Sigma_\nu^{-1} A)^{-1}$. Согласно (6.68) $(\hat{g}_i - g_i) / \sqrt{(A^* \Sigma_\nu^{-1} A)^{-1}_{ii}} \sim N(0, 1)$.

¹⁾ $A^* \Sigma_\nu^{-1} A > 0$ и, следовательно, $(A^* \Sigma_\nu^{-1} A)^{-1} > 0$

²⁾ Так как $(A^* \Sigma_\nu^{-1} A)^{-1} > 0$, то в любом ортонормированном базисе $(A^* \Sigma_\nu^{-1} A)^{-1}_{ii} > 0$, $i = 1, \dots, m$.

Выберем как в §2.6.11.1 $\alpha \in (0, 1)$, скажем $\alpha = 0.01$, и определим $\varepsilon > 0$ так, чтобы $P(|\hat{g}_i - g_i|/\sqrt{(A^*\Sigma_\nu^{-1}A)_{ii}^{-1}} \leq \varepsilon) = 1 - \alpha = 0.99$. Тогда случайный интервал

$$\left[\hat{g}_i - \varepsilon \sqrt{(A^*\Sigma_\nu^{-1}A)_{ii}^{-1}}, \hat{g}_i + \varepsilon \sqrt{(A^*\Sigma_\nu^{-1}A)_{ii}^{-1}} \right] \quad (6.70)$$

с центром в $\hat{g}_i$ покрывает истинное значение координаты g_i , $i = 1, \dots, m$, с вероятностью $1 - \alpha = 0.99$ и называется доверительным для g_i с уровнем доверия $1 - \alpha$, $i = 1, \dots, m$. Заметим, что в собственном базисе ИП $[A, \Sigma_\nu]$ $(A^*\Sigma_\nu^{-1}A)_{ii}^{-1} = \delta_i^{-1}$, $i = 1, \dots, m$, и доверительный интервал (6.70) выглядит так: $[\hat{g}_i - \varepsilon/\delta_i, \hat{g}_i + \varepsilon/\delta_i]$, $i = 1, \dots, m$.

2.6.11.3 Проверка непротиворечивости гауссовской модели ИП $[A, \Sigma_\nu]$. Запишем схему измерения (6.65) в виде

$$\xi' = A'g + \nu', \quad (6.71)$$

где $\xi' = \Sigma_\nu^{-1/2}\xi$, $A' = \Sigma_\nu^{-1/2}A$, $\nu' = \Sigma_\nu^{-1/2}\nu \sim N(0, I)$, в котором шум ν' — «белый». Как было отмечено в §3, оператор

$$\Pi' = A'(A^*A')^{-1}A'^* : \mathcal{R}^n \rightarrow \mathcal{R}^n$$

является ортогональным проектором, проецирующим на $\mathcal{R}(A')$, а, следовательно, $I - \Pi'$ ортогонально проецируется на $\mathcal{R}^\perp(A')$. Так как размерность линейного подпространства $\mathcal{R}^\perp(A)$ равна $n - m$, то $\|(I - \Pi')\xi'\|_n^2 \equiv \|(I - \Sigma_\nu^{-1/2}A(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1/2})\xi'\|_n^2$ есть χ_{n-m}^2 -статистика¹⁾. Зададим $\alpha \in (0, 1)$ например, $\alpha = 0.001$ и определим $\delta > 0$ так, чтобы $P(\chi_{n-m}^2 \geq \delta) = \alpha = 0.001$. Если гауссовская модель ИП $[A, \Sigma_\nu]$ верна, то $\|(I - \Pi')\xi'\|_n^2 \geq \delta$ с весьма малой вероятностью $\alpha = 0.001$, поэтому, отвергая гауссовскую модель ИП $[A, \Sigma_\nu]$, как противоречащую данным измерения ξ , когда $\|(I - \Pi')\xi'\|_n^2 \geq \delta$, мы будем неправы с вероятностью 0.001, т.е. примерно в 1/1000 доле случаев принятия такого решения.

Заметим, что такое правило, отвергающее модель ИП, принято потому, что ошибка в модели приводит в среднем к увеличению статистики $\|(I - \Pi')\xi'\|_n^2$. Если, например, на самом деле $\xi'' = B'g + \nu'$, то истинные значения статистики определяются не выражением $\|(I - \Pi')\xi'\|_n^2 \equiv \|(I - \Pi')\nu'\|_n^2$, а выражением $\|(I - \Pi')\xi''\|_n^2 = \|(I - \Pi')B'g + (I - \Pi')\nu'\|_n^2$, что приводит, в среднем к бóльшим ее значениям²⁾.

2.6.12. Гауссовский ИП $[A, \Sigma_\nu^2 I]$, Σ_ν^2 неизвестно. Рассмотрим важный для практики случай неполностью определенного гауссов-

¹⁾ Проверьте, что выражение для χ_{n-m}^2 не зависит от g и равно $\|(I - \Sigma_\nu^{-1/2}A(A^*\Sigma_\nu^{-1}A)^{-1}A^*\Sigma_\nu^{-1/2})\Sigma_\nu^{-1/2}\nu\|_n^2$.

²⁾ Проверьте, что $M\|(I - \Pi')\xi''\|_n^2 \leq M\|(I - \Pi')\xi'\|_n^2$.

ского ИП $[A, \Sigma_\nu]$ с неизвестной дисперсией координат шума ν в (6.65). В этом случае

$$\|\xi - Ag\|^2 = \underbrace{\|\xi - \Pi\xi\|_n^2}_{\in \mathcal{R}^\perp(A)} + \underbrace{\|\Pi\xi - Ag\|_n^2}_{\in \mathcal{R}(A)} = \|\xi - \Pi\xi\|_n^2 + \|\Pi\xi - Ag\|_n^2, \quad (6.72)$$

где $\Pi = A(A^*A)^{-1}A^*$ — ортогональный проектор в $\mathcal{R}^n$ на $\mathcal{R}(A)$. В (6.72)

$$\|\xi - \Pi\xi\|_n^2 = \|(I - \Pi)\nu\|_n^2 = \Sigma_\nu^2 \chi_{n-m}^2, \quad \|\Pi\xi - Ag\|_n^2 = \|\Pi\nu\|_n^2 = \Sigma_\nu^2 \chi_m^2 \quad (6.73)$$

причем статистики χ_{n-m}^2 и χ_m^2 независимы, поскольку случайные векторы $\Pi\nu$ и $(I - \Pi)\nu$ ортогональны.

2.6.12.1 *Оценка Σ_ν^2* . Несмещенной оценкой Σ_ν^2 является статистика

$$\begin{aligned} \widehat{\Sigma}_\nu^2 &= \|\xi - \Pi\xi\|_{n-m}^2 / (n - m) = \Sigma_\nu^2 \chi_{n-m}^2 / (n - m), \\ M\widehat{\Sigma}_\nu^2 &= \Sigma_\nu^2, \quad D\widehat{\Sigma}_\nu^2 = M(\widehat{\Sigma}_\nu^2 - \Sigma_\nu^2)^2 = 2\Sigma_\nu^4 / (n - m). \end{aligned} \quad (6.74)$$

2.6.12.2 *Доверительный эллипсоид Хоттелинга*. Как было отмечено, в силу ортогональности $(I - \Pi)\nu$ и $\Pi\nu$ случайные векторы $(I - \Pi)\nu$ и $\Pi\nu$ независимы, следовательно независимы и χ^2 — статистики в (6.73), поэтому отношение

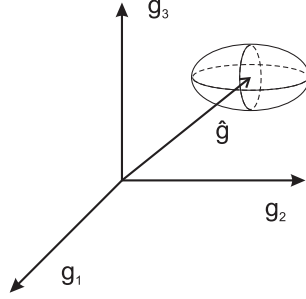
$$\frac{\|\Pi\xi - Ag\|_n^2 / m}{\|\xi - \Pi\xi\|_n^2 / (n - m)} = \varphi_{m, n-m} = \frac{\|A(\widehat{g} - g)\|_n^2 / m}{\|\xi - \Pi\xi\|_n^2 / (n - m)} \quad (6.75)$$

есть статистика Снедекора–Фишера с m и $n - m$ степенями свободы, где использовано равенство $\Pi\xi - Ag = A(A^*A)^{-1}A^*\xi - Ag = A(\widehat{g} - g)$. Выберем $\alpha \in (0, 1)$, например $\alpha = 0.001$, и определим $\varepsilon > 0$ так, чтобы $P(|\varphi_{m, n-m}| \leq \varepsilon) = 1 - \alpha = 0.999$. Тогда $P(\|A(\widehat{g} - g)\|_n^2 \leq \varepsilon \frac{m\|\xi - \Pi\xi\|_n^2}{n - m}) = 1 - \alpha = 0.999$, т.е. доверительный эллипсоид Хоттелинга

$$\mathfrak{D}_{1-\alpha} = \left\{ g \in \mathcal{R}^m, \quad \|A(\widehat{g} - g)\|_n^2 \leq \varepsilon \frac{m\|\xi - \Pi\xi\|_n^2}{n - m} \right\}, \quad (6.76)$$

см. рис. (2.10), покрывает истинное значение g с вероятностью $1 - \alpha$. В отличие от случайного эллипсоида (6.69) случайный эллипсоид Хоттелинга (6.76) имеет не только случайный центр $\widehat{g}$, но и случайные полуоси, пропорциональные статистике $\frac{\|\xi - \Pi\xi\|_n}{\sqrt{n - m}} = \widehat{\Sigma}_\nu$ в (6.74).

2.6.12.3 *Интервальные оценки для координат $g \in \mathcal{R}^m$* . Так как в рассматриваемом случае $\widehat{g} - g = (A^*A)^{-1}A^*\nu = (A^*A)^{-1}A^*\Pi\nu$, а $(I - \Pi)\xi = (I - \Pi)\nu$, то статистики $\widehat{g} - g$ и $(I - \Pi)\xi$, как зависящие от ортогональных $\Pi\nu$ и от $(I - \Pi)\nu$, независимы, причем

Рис. 2.10. Эллипсоид Хоттеллинга, оценка $\hat{g}$ — его центр

$\hat{g} - g \sim N(0, \Sigma_\nu^2 (A^* A)^{-1})$ и в любом ортонормированном базисе $\hat{g}_i - g_i \sim N(0, \Sigma_\nu^2 (A^* A)_{ii}^{-1})$. Поскольку

$$(\hat{g}_i - g_i) / (\Sigma_\nu \sqrt{(A^* A)_{ii}^{-1}}) \sim N(0, 1). \quad (6.77)$$

и $\|(I - \Pi)\|_n^2 = \Sigma_\nu^2 \chi_{n-m}^2$ не зависит от $\hat{g}_i - g_i$, то

$$\frac{(\hat{g}_i - g_i)}{\Sigma_\nu \sqrt{(A^* A)_{ii}^{-1}}} \cdot \frac{\Sigma_\nu \sqrt{n-m}}{\|(I - \Pi)\xi\|_n} = \tau_{n-m} \quad (6.78)$$

— статистика Стьюдента с $n - m$ степенями свободы.

Зададим $\alpha \in (0, 1)$, например, $\alpha = 0.01$ и определим $\varepsilon > 0$ из условия

$$P(\tau_{n-m} \leq \varepsilon) = 1 - \alpha = 0.99. \text{ Тогда } P(|\hat{g}_i - g_i| \leq \varepsilon \frac{\sqrt{(n-m)(A^* A)_{ii}^{-1}}}{\|(I - \Pi)\xi\|_n}) = 1 - \alpha \text{ и, следовательно случайный интервал}$$

$$\left\{ g_i \in \mathcal{R}^1, |g_i - \hat{g}_i| \leq \varepsilon \frac{\sqrt{(A^* A)_{ii}^{-1}} \|(I - \Pi)\xi\|_n}{\sqrt{n-m}} \right\}$$

со случайным центром $\hat{g}_i$ и случайной длиной $2\varepsilon \frac{\sqrt{(n-m)(A^* A)_{ii}^{-1}}}{\|(I - \Pi)\xi\|_n}$ покрывает истинное значение g_i с вероятностью $1 - \alpha$, $i = 1, \dots, m$.

2.7. Задачи проверки статистических гипотез

В этом параграфе речь пойдет о простейших задачах проверки гипотез о параметре распределения случайной величины. Как правило, гипотеза H и альтернатива K формулируются о возможных значениях неизвестного параметра $\vartheta \in \Theta$, а решение в пользу H или K должно быть вынесено на основании наблюдения выборки $\xi_1, \dots, \xi_n$ из распределения $F(\cdot, \vartheta)$. Понятно, что в таком случае H и K можно просто отождествить с соответствующими подмножествами Θ . Например, если параметр $\vartheta = \mu$ нормального распределения $N(\mu, \sigma^2)$ неизвестен, то

можно рассматривать гипотезу H о том, что $\mu = \mu_0$, при альтернативе $K : \mu \neq \mu_0$. В этом случае $\Theta = \mathcal{R}^1$, H можно рассматривать как подмножество прямой $\mathcal{R}^1$, состоящее из одной точки $\mu_0 : H = \{\mu_0\}$. Соответственно $K = K(\mu_0) = \mathcal{R}^1 \setminus \{\mu_0\} = \{\mu, -\infty < \mu < \infty, \mu \neq \mu_0\}$.

Пусть D — подмножество выборочного пространства $\mathcal{R}^n$, не зависящее от ϑ . Множество D называется критическим в задаче проверки гипотезы H при альтернативе K , если гипотеза H отвергается всякий раз, когда $(\xi_1, \dots, \xi_n) \in D$, и принимается в противном случае. Решение принять или отвергнуть гипотезу принимается на основе выборки $\xi_1, \dots, \xi_n$ из распределения $F(x, \vartheta)$, $x \in \mathcal{R}^1$, $\vartheta \in \Theta$.

Такой решающей процедуре свойственны ошибки двух типов. Говорят об *ошибке 1-го рода*, когда на основании выборки отклоняется на самом деле истинная гипотеза, и об *ошибке 2-го рода*, когда принимается на самом деле неверная гипотеза. При заданном критическом множестве D вероятности ошибок 1-го и 2-го рода равны соответственно

$$P(\{(\xi_1, \dots, \xi_n) \in D\}, \vartheta), \quad \vartheta \in H, \text{ и} \\ 1 - P(\{(\xi_1, \dots, \xi_n) \in D\}, \vartheta) = P(\{(\xi_1, \dots, \xi_n) \in \mathcal{R}^n \setminus D\}, \vartheta), \quad \vartheta \in K,$$

и зависят не только от D , H и K , но и от конкретного значения параметра $\vartheta \in \Theta$. Чем меньше вероятности ошибок первого и второго рода, тем лучше решающая процедура.

Вероятность отвергнуть гипотезу, когда она и на самом деле не верна, равна значению функции

$$P(\{(\xi_1, \dots, \xi_n) \in D\}, \vartheta), \quad \vartheta \in K,$$

называемой *мощностью решающей процедуры (или критерия)*. Значение мощности критерия определяется истинным значением $\vartheta \in K$.

Для иллюстрации введенных понятий вернемся к примеру проверки гипотезы о параметре нормального распределения. Дисперсию σ^2 будем считать известной. Если $\mu = \mu_0$ истинное значение математического ожидания и $(\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\mu_0, \sigma^2)$, то согласно (2.3) μ_0 покрывается случайным доверительным интервалом (2.4) с вероятностью $1 - \alpha$. Следовательно, выбирая для заданного α критическое множество в виде

$$D = \left\{ (x_1, \dots, x_n) \in \mathcal{R}^n, \frac{1}{n} \sum_{j=1}^n x_j \geq \mu_0 + \frac{\varepsilon \sigma}{\sqrt{n}} \right\} \cup \\ \cup \left\{ (x_1, \dots, x_n) \in \mathcal{R}^n, \frac{1}{n} \sum_{j=1}^n x_j \leq \mu_0 - \frac{\varepsilon \sigma}{\sqrt{n}} \right\}, \quad (7.1)$$

найдем следующее значение для вероятности ошибки 1-го рода:

$$1 - P\left(\left\{-\frac{\varepsilon\sigma}{\sqrt{n}} < \frac{1}{n} \sum_{j=1}^n x_j - \mu_0 < \frac{\varepsilon\sigma}{\sqrt{n}}\right\}\right) = \alpha.$$

Величина α ошибки 1-го рода называется *уровнем критерия*.

В данном случае гипотеза H содержит единственное значение параметра $\mu = \mu_0$. Альтернатива K содержит все остальные значения μ . Поэтому ошибка второго рода

$$1 - P(\{(\xi_1, \dots, \xi_n) \in D\}, \mu), \quad \mu \neq \mu_0$$

зависит от неизвестного параметра μ и не может быть вычислена априори. С этим обстоятельством связаны характерные трудности, возникающие при сравнении различных решающих процедур. Действительно, если $\tilde{D}$ — другая критическая область в рассматриваемой задаче проверки гипотезы, для которой ошибка первого рода также равна α : $P(\{(\xi_1, \dots, \xi_n) \in \tilde{D}\}, \mu_0) = \alpha$, то судить о том, какая из областей предпочтительнее, можно, лишь сравнивая вероятности ошибок второго рода для D и $\tilde{D}$ или соответствующие мощности. Но последние зависят от неизвестного параметра $\mu \neq \mu_0$, и может так случиться, что для некоторых $\mu \in K$ предпочтительнее область D , а для других предпочтительнее $\tilde{D}$. В таком случае говорят, что решающие процедуры *не сравнимы*. Если оказывается, что при одном и том же уровне α мощность критерия, отвечающего множеству D , не меньше, чем мощность, отвечающая $\tilde{D}$, для всех $\mu \neq \mu_0$, причем для некоторого $\mu \neq \mu_0$ неравенство строгое, то говорят, что критерий, отвечающий D , *равномерно более мощный*, чем отвечающий $\tilde{D}$. Разумеется, идеальным был бы равномерно наиболее мощный критерий, но, к сожалению, за редкими исключениями, такой критерий не существует, если K содержит более одного значения параметра.

2.7.1. Локально наиболее мощные критерии. Если в случае $H = \{\vartheta_0\}$ равномерно наиболее мощный критерий не существует, т. е. не существует критическое множество, наилучшее для каждого $\vartheta \in K$, то можно попытаться найти критическое множество, наилучшее для значений параметра $\vartheta \in K$, в известном смысле близких к ϑ_0 . Такой выбор представляет интерес, поскольку именно для близких к ϑ_0 значений ϑ велика вероятность ошибочного решения.

Рассмотрим задачу проверки гипотезы $H = \{\vartheta_0\} \subset \Theta$ при альтернативе $K = \{\vartheta \in \Theta, \vartheta > \vartheta_0\}$, где Θ — некоторое открытое подмножество действительной прямой $\mathcal{R}^1$. Если $(\xi_1, \dots, \xi_n)$ — выборка из распределения с плотностью $f(x, \vartheta)$, $x \in \mathcal{R}^1$, $\vartheta \in \Theta$, и D — критическое множество в выборочном пространстве $\mathcal{R}^n$, то вероятность ошибки первого рода

задается равенством

$$P(\{(\xi, \dots, \xi_n) \in D\}, \vartheta_0) = \int_D L(x, \vartheta_0) dx = \alpha, \quad (7.2)$$

где $L(x, \vartheta) = f(x_1, \vartheta) \dots f(x_n, \vartheta)$, $x = (x_1, \dots, x_n) \in \mathcal{R}^n$, $\vartheta \in \Theta$, — функция правдоподобия. При этом мощность критерия для каждого $\vartheta \in K$ равна

$$\beta(\vartheta) = \int_D L(x, \vartheta) dx. \quad (7.3)$$

Предположим, что для функции $\beta(\cdot)$ имеет место формула Тейлора

$$\beta(\vartheta) = \alpha + (\vartheta - \vartheta_0)\beta'(\vartheta_0) + (\vartheta - \vartheta_0)^2\beta''(\vartheta_0)/2 + o(\vartheta - \vartheta_0)^2, \quad \alpha = \beta(\vartheta_0). \quad (7.4)$$

Для этого достаточно, чтобы функция $\beta(\cdot)$ была дважды дифференцируема в некоторой окрестности ϑ_0 и вторая производная $\beta''(\vartheta)$ была непрерывна при $\vartheta = \vartheta_0$. Если $K = \{\vartheta \in \Theta, \vartheta > \vartheta_0\}$, то для получения локально наиболее мощного критерия *следует максимизировать $\beta'(\vartheta_0)$ как функцию области D* , или

$$\beta'(\vartheta_0) = \int_D (\partial L(x, \vartheta)/\partial \vartheta)|_{\vartheta=\vartheta_0} dx \sim \max_D, \quad (7.5)$$

если предположить возможность дифференцирования под знаком интеграла в (7.3).

Если $K = \{\vartheta \in \Theta, \vartheta < \vartheta_0\}$, то $\beta'(\vartheta_0)$ следует минимизировать, а в случае $K = \{\vartheta \in \Theta, \vartheta \neq \vartheta_0\}$ разумно наложить *условие локальной несмещенности $\beta'(\vartheta_0) = 0$ и максимизировать $\beta''(\vartheta_0)$ как функцию области D* . Такой критерий естественно назвать *локально наиболее мощным несмещенным*.

Во всех перечисленных задачах при *отыскании оптимального критического множества* может быть использована

Лемма 2.7.1 (Нейман-Пирсон). Пусть $f_0(x), f_1(x), \dots, f_m(x)$, $x \in \mathcal{R}^n$, — интегрируемые функции и D — подмножество $\mathcal{R}^n$, для которого

$$\int_D f_j(x) dx = c_j, \quad j = 1, \dots, m, \quad (7.6)$$

где c_j , $j = 1, \dots, m$, заданные числа. Если существуют постоянные $k_1, \dots, k_m$, такие, что для подмножества $D_0 \subset \mathcal{R}^n$, в точках которого

$$f(x_0) \geq k_1 f_1(x) + \dots + k_m f_m(x)$$

и вне которого

$$f(x_0) \leq k_1 f_1(x) + \dots + k_m f_m(x),$$

выполнены равенства (7.6), то

$$\int_{D_0} f_0(x) dx \geq \int_D f_0(x) dx.$$

Доказательство. Согласно условиям леммы

$$\begin{aligned} \int_{D_0} f_0(x) dx - \int_D f_0(x) dx &= \int_{D_0 \setminus (D_0 \cap D)} f_0(x) dx - \int_{D \setminus (D_0 \cap D)} f_0(x) dx \geq \\ &\geq \int_{D_0 \setminus (D_0 \cap D)} \sum_{j=1}^m k_j f_j(x) dx - \int_{D \setminus (D_0 \cap D)} \sum_{j=1}^m k_j f_j(x) dx = 0. \end{aligned}$$

Равенство нулю следует из равенств

$$\int_{D_0 \setminus (D_0 \cap D)} f_j(x) dx = \int_{D \setminus (D_0 \cap D)} f_j(x) dx, \quad j = 1, \dots, m,$$

которые, в свою очередь, следуют из равенств

$$\int_D f_j(x) dx = \int_{D_0} f_j(x) dx = c_j, \quad j = 1, \dots, m. \quad \blacksquare$$

Возвращаясь к задаче построения критического множества в локально наиболее мощном критерии, при оговоренных выше предположениях получаем следующий результат.

Теорема 2.7.1. 1. Если $K = \{\vartheta \in \Theta, \vartheta > \vartheta_0\}$, $H = \{\vartheta_0\}$, то критическое множество D локально наиболее мощного критерия уровня α имеет вид

$$D = \left\{ x \in \mathcal{R}^n, \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0} \geq k_1 L(x, \vartheta_0) \right\}, \quad (7.7)$$

где k_1 определяется условием

$$\beta(\vartheta_0) = \int_D L(x, \vartheta_0) dx = \alpha. \quad (7.8)$$

Аналогично, если $K = \{\vartheta \in \Theta, \vartheta < \vartheta_0\}$, $H = \{\vartheta_0\}$, то

$$D = \left\{ x \in \mathcal{R}^n, \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0} \leq k_1 L(x, \vartheta_0) \right\},$$

и k_1 определяется условием (7.8).

2. Если $K = \{\vartheta \in \Theta, \vartheta \neq \vartheta_0\}$, $H = \{\vartheta_0\}$, то критическое множество локально наиболее мощного несмещенного критерия уровня α дается равенством

$$D = \left\{ x \in \mathcal{R}^n, \frac{\partial^2 L(x, \vartheta_0)}{\partial \vartheta_0^2} \geq k_1 \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0} + k_2 L(x, \vartheta_0) \right\},$$

где k_1 и k_2 определяются условиями

$$\beta(\vartheta_0) = \int_D L(x, \vartheta_0) dx = \alpha, \quad \beta'(\vartheta_0) = \int_D \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0} dx = 0. \quad (7.9)$$

В обоих случаях предполагается, что постоянные k_1 и k_2 , удовлетворяющие соответственно условиям (7.8) и (7.9), существуют.

Доказательство.

1. Следует доказать, что критическое множество (7.7) доставляет максимум производной (7.5). Действительно, пусть $\tilde{D}$ — любое другое критическое множество уровня α :

$$\int_{\tilde{D}} L(x, \vartheta_0) dx = \alpha. \quad (7.10)$$

Выберем в лемме 2.7.1 $f_1(x) = L(x, \vartheta_0)$, $f_0(x) = \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0}$. Поскольку согласно (7.7) для $x \in D$ $f_0(x) \geq k_1 f_1(x)$, то на основании леммы 2.7.1: $\int_D f_0(x) dx \geq \int_{\tilde{D}} f_0(x) dx$. Следовательно, при указанных в теореме

условиях производная $\beta'(\vartheta_0) = \int_D \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0} dx$ максимальна.

Случай альтернативы $K = \{\vartheta \in \Theta, \vartheta \in \vartheta_0\}$ вполне аналогичен рассмотренному.

2. На этот раз выберем в лемме 2.7.1

$$f_0(x) = \frac{\partial^2 L(x, \vartheta_0)}{\partial \vartheta_0^2}, \quad f_1(x) = \frac{\partial L(x, \vartheta_0)}{\partial \vartheta_0}, \quad f_2(x) = L(x, \vartheta_0).$$

Тогда для любого множества $\tilde{D}$, удовлетворяющего условиям (7.10), согласно лемме

$$\int_D \frac{\partial^2 L(x, \vartheta_0)}{\partial \vartheta_0^2} dx \geq \int_{\tilde{D}} \frac{\partial^2 L(x, \vartheta_0)}{\partial \vartheta_0^2} dx,$$

т. е. D определяет несмещенный локально наиболее мощный критерий уровня α . ■

Пример 2.7.1. Рассмотрим задачу проверки гипотезы о параметре μ нормального распределения $N(\mu, \sigma^2)$ при известной дисперсии σ^2 . Пусть $H = \{\mu_0\}$, $K = \{\mu \in (-\infty, \infty), \mu \neq \mu_0\}$. Так как

$$L(x, \mu) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2 \right\}, \quad (7.11)$$

$$\frac{\partial \ln L(x, \mu)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu), \quad \frac{\partial^2 \ln L(x, \mu)}{\partial \mu^2} = -\frac{n}{\sigma^2}, \quad x \in \mathcal{R}^n,$$

то

$$\begin{aligned}\frac{\partial L(x, \mu)}{\partial \mu} &= L(x, \mu) \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu), \\ \frac{\partial^2 L(x, \mu)}{\partial \mu^2} &= L(x, \mu) \left\{ \left[\frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu)^2 \right]^2 - \frac{n}{\sigma^2} \right\}, \quad x \in \mathcal{R}^n.\end{aligned}\quad (7.12)$$

Следовательно, критическое множество имеет вид

$$D = \left\{ x \in \mathcal{R}^n, -\frac{n}{\sigma^2} + \left[\frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu_0) \right]^2 \geq \frac{k_1}{\sigma^2} \sum_{j=1}^n (x_j - \mu_0) + k_2 \right\},$$

где k_1 и k_2 определяются условиями

$$\int_D L(x, \mu_0) dx = \alpha, \quad \int_D L(x, \mu_0) \sum_{j=1}^n (x_j - \mu_0) dx = 0. \quad (7.13)$$

Второе уравнение (7.13) будет удовлетворено для любого k_2 , если $k_1 = 0$. Первое уравнение (7.13) можно записать в виде

$$P \left(\left\{ \left[\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n (\xi_j - \mu_0) \right]^2 \geq 1 + k_2 \frac{\sigma^2}{n} \right\}, \mu_0 \right) = \alpha,$$

где $(\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\mu_0, \sigma^2)$. Так как статистика $\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n (\xi_j - \mu_0)$ имеет распределение $N(0, 1)$, то

$\left[\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n (\xi_j - \mu_0) \right]^2 = \chi_1^2$, и k_2 может быть найдено по таблице распределения χ^2 из условия

$$P \left(\left\{ \chi_1^2 \geq 1 + k_2 \frac{\sigma^2}{n} \right\} \right) = \alpha.$$

Заметим, что найденное критическое множество можно записать в виде

$$\begin{aligned}D &= \left\{ x \in \mathcal{R}^n, \frac{1}{n} \sum_{j=1}^n (x_j - \mu_0) \geq \frac{\sigma(k_2 \sigma^2 + n)^{1/2}}{n} \right\} \cup \\ &\cup \left\{ x \in \mathcal{R}^n, \frac{1}{n} \sum_{j=1}^n (x_j - \mu_0) \leq -\frac{\sigma(k_2 \sigma^2 + n)^{1/2}}{n} \right\},\end{aligned}$$

аналогичном с (7.1).

Пример 2.7.2. Пусть теперь $H = \mu_0$ и $K = \{\mu_0 < \mu < \infty\}$. Согласно теореме 2.7.1 критическое множество локально наиболее мощного критерия следует искать в виде

$$D = \left\{ x \in \mathcal{R}^n, \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu_0) \geq k_1 \right\} \quad (7.14)$$

при условии, что

$$\int_D L(x, \mu_0) dx = \alpha. \quad (7.15)$$

Так как статистика $\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n (\xi_j - \mu_0) = \eta_n$ имеет распределение $N(0, 1)$, а условие (7.15) согласно (7.14) можно записать в виде $P\left(\left\{\eta_n \geq k_1 \frac{\sigma}{\sqrt{n}}\right\}\right) = \alpha$, то k_1 может быть найдено по таблице нормального распределения.

2.7.2. Случай простой гипотезы и простой альтернативы. Наиболее мощный критерий. Если гипотеза H (альтернатива K) содержит лишь одно значение $\vartheta = \vartheta_0$ ($\vartheta = \vartheta_1$), то гипотеза (альтернатива) называется *простой*. Выше рассматривались задачи проверки простой гипотезы при непростых альтернативах. Посмотрим, какие возможны упрощения в том случае, когда как гипотеза, так и альтернатива простые: $H = \{\vartheta_0\}$, $K = \{\vartheta_1\}$, $\vartheta_1 \neq \vartheta_0$. Ограничимся случаем распределения, имеющего плотность $f(x, \vartheta)$, $x \in \mathcal{R}^1$, $\vartheta = \vartheta_0, \vartheta_1$. Рассмотрим задачу отыскания наиболее мощного критерия, отвечающего заданному уровню α . Речь идет о задаче отыскания критического множества $D \subset \mathcal{R}^n$, для которого $\int_D L(x, \vartheta_1) dx \sim \max$ при условии $\int_D L(x, \vartheta_0) dx = \alpha$, где $L(x, \vartheta) = f(x_1, \vartheta)f(x_2, \vartheta) \dots f(x_n, \vartheta)$, $\vartheta = \vartheta_0, \vartheta_1$, $x = (x_1, \dots, x_n) \in \mathcal{R}^n$, — функция правдоподобия.

Выберем в лемме 2.7.1 $f_0(x) = L(x, \vartheta_1)$, $f_1(x) = L(x, \vartheta_0)$ и определим область D условием

$$D = \{x \in \mathcal{R}^n, f_0(x) \geq k_1 f_1(x)\}, \quad (7.16)$$

где k_1 находится из уравнения

$$\int_D f_1(x) dx = \alpha. \quad (7.17)$$

Если существует k_1 , удовлетворяющее уравнению (7.17), то для любого критического множества $\tilde{D}$ уровня α , т. е. такого, что $\int_{\tilde{D}} f_1(x) dx = \alpha$, согласно лемме

$$\int_D f_0(x) dx \geq \int_{\tilde{D}} f_0(x) dx.$$

Это и означает, что критерий, отвечающий D (7.16), наиболее мощный среди всех критериев уровня α . Сформулируем полученный результат.

Теорема 2.7.2. В задаче проверки гипотезы $H = \{\vartheta_0\}$ при альтернативе $K = \{\vartheta_1\}$ критическая область

$$D = \{x \in \mathcal{R}^n, L(x, \vartheta_1) \geq k_1 L(x, \vartheta_0)\},$$

определяет наиболее мощный критерий уровня α , если k_1 удовлетворяет условию

$$\int_D L(x, \vartheta_0) dx = \alpha.$$

Пример 2.7.3. Вернемся еще раз к задаче проверки гипотезы о параметре μ нормального распределения $N(\mu, \sigma^2)$ при известной дисперсии σ^2 . Пусть $H = \{\mu_0\}$, $K = \{\mu_1\}$, $\mu_1 > \mu_0$. Согласно теореме 2.7.2 критическую область D , отвечающую наиболее мощному критерию, можно искать в виде

$$D = \{x \in \mathcal{R}^n, L(x, \mu_1) \geq k L(x, \mu_0)\}, \quad 0 \leq k < \infty, \quad (7.18)$$

где функция правдоподобия имеет вид (7.11). Семейство областей (7.18) можно задать иначе, воспользовавшись явным видом функции правдоподобия,

$$\begin{aligned} D &= \left\{ x \in \mathcal{R}^n, \sum_{j=1}^n (x_j - \mu_1)^2 \leq \sum_{j=1}^n (x_j - \mu_0)^2 + c \right\} = \\ &= \left\{ x \in \mathcal{R}^n, (\mu_1 - \mu_0) \sum_{j=1}^n x_j \geq c_1 \right\} = \\ &= \left\{ x \in \mathcal{R}^n, \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (x_j - \mu_0) \geq c_2 \right\}, \end{aligned} \quad (7.19)$$

где c, c_1, c_2 — параметры, определяющие семейство (7.18). Значение c_2 в (7.19) следует определить из уравнения

$$\alpha = \int_D L(x, \mu_0) dx = P \left(\left\{ \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (\xi_j - \mu_0) \geq c_2 \right\} \right). \quad (7.20)$$

В (7.19) и (7.20) использовано условие $\mu_1 - \mu_0 > 0$, а также тот факт, что статистика $\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (\xi_j - \mu_0)$ имеет распределение $N(0, 1)$, если $(\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\mu_0, \sigma^2)$. Поскольку уравнение (7.20) разрешимо относительно c_2 при любом $\alpha \in [0, 1]$, то равенство (7.19) при c_2 , удовлетворяющем (7.20), определяет критическую область, отвечающую наиболее мощному критерию уровня α , $0 \leq \alpha \leq 1$.

В данном случае характерно, что критическая область не зависит от альтернативы μ_1 и, следовательно, определяет наиболее

мощный критерий при любом $\mu_1 > \mu_0$. Поэтому найденный ранее локально наиболее мощный критерий $\mu_0 < \mu < \infty$ (7.14) в задаче проверки гипотезы $H = \{\mu_0\}$ при альтернативе $K = \{\infty > \mu > \mu_0\}$ на самом деле является равномерно наиболее мощным.

2.7.3. Доверительные множества и задачи проверки гипотез.

Напомним, что $100\gamma\%$ -ным доверительным множеством для параметра ϑ распределения $F(\cdot, \vartheta)$, или, иначе, доверительным множеством уровня γ , называется случайное подмножество $A(\xi)$ множества Θ , такое, что

$$P(\{\vartheta \in A(\xi)\}, \vartheta) = \gamma,$$

где $\xi = (\xi_1, \dots, \xi_n)$ выборка из распределения $F(\cdot, \vartheta)$. Можно сказать, что случайное множество $A(\xi)$ покрывает истинную параметрическую точку $\vartheta \in \Theta$ с вероятностью γ .

Связь между задачей построения доверительных множеств и задачей проверки статистических гипотез устанавливает

Теорема 2.7.3. Рассмотрим для каждого $\vartheta \in \Theta$ какой-либо критерий уровня α для проверки гипотезы $H = \{\vartheta\}$. Обозначим $S\{\vartheta\} = \mathcal{R}^n \setminus D$ область выборочного пространства $\mathcal{R}^n$ принятия гипотезы H (D — критическое множество). Для каждого $x \in \mathcal{R}^n$ определим множество $A(x)$ в пространстве параметров Θ , полагая

$$A(x) = \{\vartheta \in \Theta, x \in S(\vartheta)\}.$$

Тогда $A(x)$, $x \in \mathcal{R}^n$, — семейство доверительных множеств для ϑ уровня $1 - \alpha$, т. е.

$$P(\{\vartheta \in A(\xi)\}, \vartheta) = 1 - \alpha.$$

Если $S(\vartheta_0)$ определяет равномерно наиболее мощный критерий уровня α для проверки гипотезы $H = \{\vartheta_0\}$ при альтернативе $K(\vartheta_0)$, то ¹⁾ $A(\cdot)$ минимизирует вероятность

$$P(\{\vartheta' \in A(\xi)\}, \vartheta)$$

для всех $\vartheta \in K(\vartheta')$ в классе всех доверительных множеств уровня $1 - \alpha$.

Доказательство. По определению множества $A(x)$ $\vartheta \in A(x)$ тогда и только тогда, когда $x \in S(\vartheta)$. Следовательно,

$$P(\{\vartheta \in A(\xi)\}, \vartheta) = P(\{\xi \in S(\vartheta)\}, \vartheta) = 1 - \alpha, \quad \vartheta \in \Theta.$$

Это равенство описывает структуру доверительных множеств $A(x)$, $x \in \mathcal{R}^n$, а именно, $A(x)$ есть совокупность тех значений ϑ , для которых принимается гипотеза $H = \{\vartheta\}$, когда наблюдается $\xi = x \in \mathcal{R}^n$.

¹⁾ $A(\cdot)$ — многозначное отображение, определенное на $\mathcal{R}^n$, со значениями в классе $\mathcal{P}(\Theta)$ всех подмножеств множества Θ .

Пусть $\tilde{A}(x)$ — любое семейство доверительных множеств уровня $1 - \alpha$ и

$$\tilde{S}'(\vartheta) = \{x \in \mathcal{R}^n, \vartheta \in \tilde{A}(x)\}.$$

Тогда по определению

$$P(\{\xi \in \tilde{S}(\vartheta)\}, \vartheta) = (P\{\vartheta \in \tilde{A}(\xi)\}, \vartheta) = 1 - \alpha,$$

так что $\tilde{S}(\vartheta_0)$ является областью $\mathcal{R}^n$ принятия гипотезы $H = \{\vartheta_0\}$ уровня α . Так как $S(\vartheta_0)$ — область принятия гипотезы, соответствующая равномерно наиболее мощному критерию, то при любом $\vartheta \in K(\vartheta_0)$

$$P(\{\xi \in \tilde{S}(\vartheta_0)\}, \vartheta) \geq P(\{\xi \in S(\vartheta_0)\}, \vartheta)$$

и, следовательно,

$$P(\{\vartheta_0 \in \tilde{A}(\xi)\}, \vartheta) \geq P(\{\vartheta_0 \in A(\xi)\}, \vartheta). \blacksquare$$

2.8. Элементы теории статистических решений

Рассмотрим типичную ситуацию, в которой возникает задача принятия решения. Предположим, что нам известны возможные «состояния природы» $\vartheta \in \Theta$, например, $\vartheta_1, \vartheta_2, \dots, \vartheta_k$, и определены возможные «действия» $d \in D$, например, $d_1, \dots, d_N$, которые связаны с состояниями природы таким образом, что действие d_i , выполненное при состоянии природы ϑ_j , сопряжено с риском (потерь или других неприятностей), оцениваемым числом $l(\vartheta_j, d_i)$, причем значения риска, сопутствующего каждой комбинации ϑ_j, d_i , известны, или, иначе говоря, известен риск потерь $l(\vartheta, d)$, $\vartheta \in \Theta$, $d \in D$. Разумеется, на практике множества Θ и D не обязательно конечны.

Если состояние природы ϑ известно, то вопрос о действии d естественно решается следующим образом: в каждом состоянии природы $\vartheta \in \Theta$ следует выполнять то или те действия $d \in D$, при которых риск $l(\vartheta, d)$ минимален. В данном случае *правило действия* состоит в наблюдении за состояниями природы и принятии определенного *решения о действии* $d = d_i$, если минимум $l(\vartheta, d)$ как функции $d \in D$ достигается на одном действии d_i . Если минимум $l(\vartheta, d) = l(\vartheta)$, соответствующий состоянию природы ϑ , достигается на нескольких $d \in D$, скажем, на $d_{i_1}, \dots, d_{i_m}$, $m \leq N$, то можно предпринять любое из них. Но можно также воспользоваться экспериментом с *т случайными исходами* $\alpha_1, \alpha_2, \dots, \alpha_m$ с вероятностями $p_1, p_2, \dots, p_m$, $p_1 + \dots + p_m = 1$. В этом случае, прежде чем принять решение о действии в состоянии природы ϑ можно разыграть случайный эксперимент и принять решение о действии d_{i_p} , если исходом эксперимента окажется α_p . Такое правило называется *рандомизированным*, в отличие от правил $d_{i_1}, \dots, d_{i_m}$, которые называются *чистыми*. В случае рандомизирован-

ного правила риск $l(\vartheta, d)$ при фиксированном ϑ является случайной величиной, причем ожидаемый (средний) риск

$$Ml(\vartheta, d) = \sum_{t=1}^m l(\vartheta, d_{i_t}) p_t = l(\vartheta, d_i) = l(\vartheta), \quad i = i_1, \dots, i_m.$$

На самом деле, конечно, состояние природы в момент принятия решения обычно неизвестно. Если, однако, о состояниях природы неизвестно ничего ¹⁾, то нет и задачи принятия решения: можно принять любое решение, так как в терминах риска невозможно привести аргументов в пользу какого-либо одного из них. Если же известно множество Θ *возможных* состояний природы, то оптимальное правило решения можно определить, например, как решение $d^* \in D$ задачи

$$c^* = \max_{\vartheta \in \Theta} l(\vartheta, d^*) = \min_{d \in D} \max_{\vartheta \in \Theta} l(\vartheta, d), \quad (8.1)$$

минимизирующее в (8.1) максимальный риск $\max_{\vartheta \in \Theta} l(\vartheta, d) = l(\vartheta(d), d)$, $d \in D$, отвечающий «наиболее неблагоприятному» состоянию природы $\vartheta = \vartheta(d^*) \in \Theta$.

Примечательно, что если в этой ситуации решение должно приниматься неоднократно, то правило d^* , найденное в (8.1), может быть улучшено в среднем путем его рандомизации, согласно которой, как уже было сказано, решения $d_1, \dots, d_N$ каждый раз принимаются, случайно с некоторыми фиксированными вероятностями $p_1, \dots, p_N$. Точнее *рандомизированное решение*, (или рандомизированное действие) — это случайная величина δ со значениями в D , распределенная согласно условию $P(\delta = d_i) = p_i$, $i = 1, \dots, N$.

Теперь, чтобы определить оптимальное рандомизированное правило действия δ^* , в отличие от задачи (8.1), требуется найти распределение $p_1^*, \dots, p_N^*$, минимизирующее максимальное значение математического ожидания риска, или короче — *ожидаемый риск*, который как функция δ , является случайной функцией $\lambda(\vartheta) = l(\vartheta, \delta)$, $\vartheta \in \Theta$. Иначе говоря, оптимальное рандомизированное действие δ^* определяется как решение задачи

$$\max_{\vartheta \in \Theta} Ml(\vartheta, \delta^*) = \min_{\delta} \max_{\vartheta \in \Theta} Ml(\vartheta, \delta), \quad (8.2)$$

в которой $M(l, (\vartheta, \delta)) = \sum_{i=1}^N p_i l(\vartheta, d_i)$ и $\min_{\delta}$ вычисляется на множестве $\mathcal{P} \triangleq \{p \triangleq (p_1, \dots, p_N), p_i \geq 0, i = 1, \dots, N, p_1 + \dots + p_N = 1\}$ всех распределений δ .

¹⁾ В том числе и множество Θ возможных состояний природы.

Для $\Theta = \{\vartheta_1, \dots, \vartheta_k\}$ определим *ожидаемый (маргинальный) риск*, отвечающий состоянию природы ϑ_t ,

$$l_t(p) \triangleq Ml(\vartheta_t, \delta) = \sum_{i=1}^N p_i l(\vartheta_t, d_i), \quad t = 1, \dots, k, \quad (8.3)$$

вектор $l(p) = (l_1(p), \dots, l_k(p)) \in \mathcal{R}^k$, $p \in \mathcal{P}$, и его значения $l^{(i)} = l(p^{(i)})$ при $p = p^{(i)} \triangleq (0, \dots, 0, p_i = 1, 0, \dots, 0)$, $i = 1, \dots, N$.

В задаче (8.1) требуется найти точку $l(p^*) = (l_1(p^*), \dots, l_k(p^*)) \triangleq (Ml(\vartheta_1, \delta^*), \dots, Ml(\vartheta_k, \delta^*)) \triangleq \left(\sum_{i=1}^N p_i^* l(\vartheta_1, d_i), \dots, \sum_{i=1}^N p_i^* l(\vartheta_k, d_i) \right) \in \mathcal{L} \triangleq \{l(p), p \in \mathcal{P}\} = \text{co}\{l^{(1)}, \dots, l^{(N)}\}$, максимальная координата которой минимальна¹⁾, $\max_{1 \leq t \leq k} l_t(p^*) = \min_{p \in \mathcal{P}} \max_{1 \leq t \leq k} l_t(p) = \min_{l \in \mathcal{L}} \max_{1 \leq t \leq k} l_t$, $l = (l_1, \dots, l_k)$. Поскольку $\mathcal{L}$ — ограниченное выпуклое и замкнутое множество в $\mathcal{R}^k$, а $\max(l_1, \dots, l_k)$, $l = (l_1, \dots, l_k) \in \mathcal{R}^k$, — непрерывная функция на $\mathcal{R}^k$, то задача на минимум (8.2), записанная в виде

$$\begin{aligned} c_r^* &= \max(l_1^*, \dots, l_k^*) = \min\{\max\{l_1, \dots, l_k\} | l \in \mathcal{L}\} = \\ &= \min\{\max\{l_1(p), \dots, l_k(p)\} | p \in \mathcal{P}\}, \end{aligned} \quad (8.4)$$

всегда имеет решение.

На рис. 2.11 представлены графические иллюстрации решений задач (8.1) и (8.2), в постановке (8.3), (8.4) в случае $k = 2$, $N = 8$

Рассмотрим теперь задачу принятия решения, в которой возможны наблюдения за природой $x \in X = \{x_1, \dots, x_q\}$ с возможными значениями $x_1, \dots, x_q$. Наблюдения должны содержать некоторую информацию о состоянии природы. Предположим, что *эта информация задается распределением переходных вероятностей $p(x|\vartheta)$ наблюдений $x \in X$ для каждого состояния природы $\vartheta \in \Theta$* . Теперь решение о действии следует принимать с учетом результата наблюдения.

Определим *правило решения s* как отображение множества наблюдений X на множество действий D , $s(\cdot) : X \rightarrow D$. Если

$$s(x) = d, \quad (8.5)$$

то правило s при наблюдении x предписывает действие d . В данном случае всего N^q отображений $s_i(\cdot) : \{x_1, \dots, x_q\} \rightarrow \{d_1, \dots, d_N\}$, $i = 1, \dots, N$, множество всех таких отображений (чистых правил) обозначим S .

При этом с каждым правилом s связано разбиение (которое также обозначим s) множества наблюдений X на подмножества $D_1, \dots, D_N$:

¹⁾ $\text{co}(l^{(1)}, \dots, l^{(N)}) = \{l^{(1)}p_1 + \dots + l^{(N)}p_N, p = (p_1, \dots, p_N) \in \mathcal{P}\}$ — выпуклая оболочка векторов $l^{(i)} = (l(\vartheta_1, d_i), \dots, l(\vartheta_k, d_i)) \in \mathcal{R}^k$, $i = 1, \dots, N$, см. рис. 2.11.

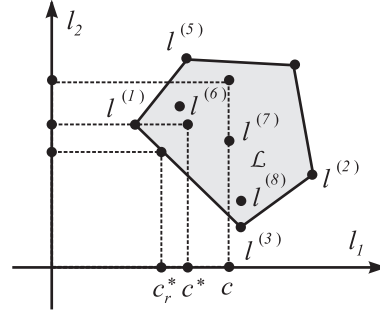


Рис. 2.11. Множество $\mathcal{L} = \text{co}(l^{(1)}, \dots, l^{(8)})$; множество $\{(l_1, l_2) \in \mathcal{R}^2, \max(l_1, l_2) = c\}$, множество $\{(l_1, l_2) \in \mathcal{R}^2, \max(l_1, l_2) = \max(l_1^*, l_2^*) = c^*\}$, где $l_1^* = l(\vartheta_1, d^*)$, $l_2^* = l(\vartheta_2, d^*)$, d^* — решение задачи (8.1); c_r^* — значение минимума в задаче (8.4), определяющее распределение $p_1^*, \dots, p_8^*$ оптимального рандомизированного действия δ^* , $p_2^* = p_4^* = p_5^* = p_6^* = p_7^* = p_8^* = 0$, $p_1^* l_1^{(1)} + p_3^* l_1^{(3)} = p_1^* l_2^{(1)} + p_3^* l_2^{(3)} = c_r^*$, $p_1^* + p_3^* = 1$; $\max_{s=1,2} Ml(\vartheta_s, \delta^*) = c_r^* \leq \max_{s=1,2} Ml(\vartheta_s, d^*) = c^*$, т.е. ожидаемый риск c_r^* , сопутствующий рандомизированному решению δ^* в (8.4), меньше риска c^* , сопутствующего решению d^* в (8.1).

$X = D_1 + \dots + D_N$, где

$$D_j = \{x \in X, s(x) = d_j\}, \quad j = 1, \dots, N.$$

Каждое *правило действия* сопряжено с риском, и, естественно, лучшим является то, которому сопутствует меньший риск. Задача сводится к выбору лучшего правила.

2.8.1. Средний (ожидаемый) риск. Рандомизация решения.

Пусть s — некоторое правило действия. Тогда распределение $p(x|\vartheta)$, $x \in X$, можно пересчитать в распределение $p_s(d|\vartheta)$, $d \in D$, по формуле

$$p_s(d|\vartheta) = \sum_{x:s(x)=d} p(x|\vartheta), \quad d \in D, \quad \vartheta \in \Theta, \quad s(\cdot) \in S, \quad (8.6)$$

и вычислить ожидаемый риск потерь, сопутствующий применению правила s в состоянии природы $\vartheta_i \in \Theta$,

$$L_i(s) = \sum_{t=1}^N l(\vartheta_i, d_t) p_s(d_t|\vartheta_i), \quad i = 1, \dots, k. \quad (8.7)$$

Поскольку то или иное правило интересует нас лишь с точки зрения сопутствующего риска, то исчерпывающей характеристикой s является точка $l(s)$ в $\mathcal{R}^k$ с координатами $L_1(s), \dots, L_k(s)$, $l(s) = (L_1(s), \dots, L_k(s))$, $s \in S$. На рис. 2.12 как и на рис. 2.11 пред-

ставлен случай двух состояний природы ϑ_1 и ϑ_2 и изображены шесть точек в $\mathcal{R}^2$, представляющих потери шести правил решения ¹⁾.

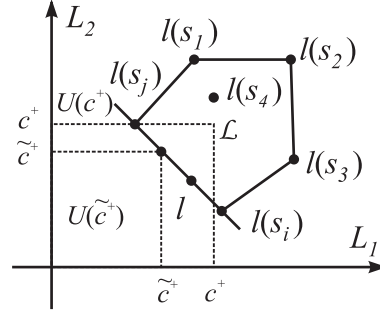


Рис. 2.12. Шесть точек представляют потери, отвечающие шести отмеченным чистым правилам $s_1, \dots, s_4, s_i, s_j$, множество $U(\tilde{c}^+)$ определяет минимаксный риск, равный $\tilde{c}^+$, c^+ — минимальный риск, полученный как решение задачи (8.9)

Если для правил s_1 и s_2 $l(s_1) = (L_1(s_1), L_2(s_1)) \leq (L_1(s_2), L_2(s_2)) = l(s_2)$, что означает $L_1(s_1) \leq L_1(s_2)$, $L_2(s_1) \leq L_2(s_2)$, то говорят, что правило s_1 *доминирует* над s_2 , см. рис. 2.12. При этом правилу s_2 в любом случае будут сопутствовать потери не меньшие, чем правилу s_1 , и, следовательно, s_2 можно исключить из рассмотрения. Точки, отвечающие правилам $s_1, s_2, \dots, s_i, s_j$, и их выпуклая оболочка $\mathcal{L} \triangleq \text{co}\{l(s_1), l(s_2), \dots, l(s_i), l(s_j)\}$, представлены на рис. 2.12. Из сказанного следует, что нас могут интересовать лишь правила s_i и s_j , которым на рис. 2.12 соответствуют точки $l(s_i)$ и $l(s_j)$, для которых нет других, доминирующих над ними, расположенных левее их и ниже.

Если p_i и p_j — вероятности, с которыми будут применяться правила s_i и s_j , $p_i + p_j = 1$, то точка $l = p_i l(s_i) + p_j l(s_j)$, представляющая рандомизированное правило $\tilde{s}$, согласно которому с вероятностью p_i применяется правило s_i и с вероятностью p_j — правило s_j , лежит на прямой, соединяющей $l(s_i)$ и $l(s_j)$, причем — между точками $l(s_i)$ и $l(s_j)$. Ожидаемые маргинальные риски, сопутствующие правилу $\tilde{s}$, даются равенствами

$$\begin{aligned} ML_1(\tilde{s}) &= p_i L_1(s_i) + p_j L_1(s_j), \text{ при } \vartheta = \vartheta_1, \\ ML_2(\tilde{s}) &= p_i L_2(s_i) + p_j L_2(s_j), \text{ при } \vartheta = \vartheta_2, \end{aligned} \quad (8.8)$$

а вероятности p_i, p_j определяются как решение линейных уравнений $ML_1(\tilde{s}) = ML_2(\tilde{s})$, $p_1 + p_2 = 1$, при условии $p_1 \geq 0$, $p_2 \geq 0$.

¹⁾ Остальные из $N^q - 6$, как и $l(s_4)$, лежат в пятиугольнике $\mathcal{L} = \text{co}\{l(s_1), \dots, l(s_{N^q})\}$ и не показаны, поскольку не влияют на выбор оптимального правила.

Если точки, представляющие доминирующие правила, не могут быть выделены априори (как $l(s_i)$ и $l(s_j)$ на рис. 2.12), то следует рассматривать рандомизированные правила, в которых учитываются все N^q чистых правил. Понятно, что, например, множество точек на рис. 2.12, соответствующих *всем рандомизированным правилам, является выпуклой оболочкой* $\mathcal{L} = \text{co}\{l(s_1), \dots, l(s_{N^q})\} \triangleq \{\lambda_1 l(s_1) + \dots + \lambda_{N^q} l(s_{N^q}), \lambda_1 \geq 0, \dots, \lambda_{N^q} \geq 0, \lambda_1 + \dots + \lambda_{N^q} = 1\}$, *натянутой на* $l(s_1), \dots, l(s_{N^q})$. При этом важно отметить, что множество $\mathcal{L}$ точек, представляющих все рандомизированные правила, выпукло и замкнуто в $\mathcal{R}^2$.

2.8.2. Минимаксное правило решения. Рассмотрим *минимаксное правило действия* s^+ , минимизирующее максимальный риск среди $L_i(s)$, $i = 1, \dots, k$ в (8.7) на множестве S всех известных правил

$$\max_{1 \leq i \leq k} L_i(s^+) = \min_{s \in S} \max_{1 \leq i \leq k} L_i(s). \quad (8.9)$$

Если семейство множеств $U(c) \triangleq \{l = (L_1, \dots, L_k) \in \mathcal{R}^k, \max_{1 \leq i \leq k} L_i \leq c\}$, $c \in \mathcal{R}^1$, см. рис. 2.12, то правило $s^+ \in S$ определится как соответствующее точке $l(s^+) \in U(c^+)$, где c^+ — минимальное значение $c \in \mathcal{R}^1$, при котором $l(s^+) \in U(c)$, $c^+ = \min\{c \in \mathcal{R}^1, \{l(s), s \in S\} \cap U(c) \neq \emptyset\}$. Минимаксный риск $\max_{1 \leq i \leq k} L_i(s^+) = c^+$.

Пусть рандомизированное правило $\tilde{s}$ состоит из чистых правил $s_1, s_2, \dots, s_t$, применяемых с вероятностями $p_1, p_2, \dots, p_t$. Тогда ожидаемый маргинальный риск, связанный с применением $\tilde{s}$ в состоянии ϑ_i , равен

$$ML_i(\tilde{s}) = \sum_{j=1}^t L_i(s_j) p_j = \sum_{j=1}^t \sum_{m=1}^N l(\vartheta_i, d_m) p_{s_j}(d_m | \vartheta_i) p_j, \quad i = 1, \dots, k. \quad (8.10)$$

Минимаксное рандомизированное правило $\tilde{s}^*$, минимизирующее максимальный ожидаемый маргинальный риск среды $ML_i(\tilde{s})$, $i = 1, \dots, k$, в (8.10), определим условием

$$\max_{1 \leq i \leq k} ML_i(\tilde{s}^+) = \min_{\tilde{s} \in \tilde{S}} \max_{1 \leq i \leq k} ML_i(\tilde{s}), \quad (8.11)$$

в котором $\tilde{S}$ — класс всех рандомизированных правил. Минимаксное рандомизированное правило $\tilde{s}^+$ определится как соответствующее точке $l(\tilde{s}^+) \in U(\tilde{c}^+)$, где $\tilde{c}^+$ минимальное значение $c \in \mathcal{R}^1$, при котором $\mathcal{L} \cap U(c) \neq \emptyset$, $\tilde{c}^+ = \min\{c \in \mathcal{R}^1, \mathcal{L} \cap U(c) \neq \emptyset\}$; минимаксный ожидаемый риск

$\max_{1 \leq i \leq k} ML_i(\tilde{s}^+) = \tilde{c}^+$. Так как $\{l(s), s \in S\} = \{l(s_1), \dots, l(s_{N^q})\} \subset \mathcal{L} = \text{co}\{l(s_1), \dots, l(s_{N^q})\}$, то в любом случае $\tilde{c}^+ \leq c^+$ см. рис. 2.13, поэтому *минимаксным называется рандомизированное правило* $\tilde{s}^+$,

а задача (8.9) и правило s^+ , доминируемое $\tilde{s}^+$, обычно, не рассматриваются.

Согласно равенствам (8.8) вероятности p_i^* , p_j^* , определяющие искоемое рандомизированное правило $\tilde{s}^+$ на рис. 2.12, удовлетворяют системе линейных уравнений $\tilde{c}^+ = p_i^* L_1(s_i) + p_j^* L_1(s_j) = p_i^* L_2(s_i) + p_j^* L_2(s_j)$, $p_i^* + p_j^* = 1$ (при условии $p_i^* \geq 0$, $p_j^* \geq 0$).

На рис. 2.13 приведены примеры ситуаций в которых $c_1^+ = \tilde{c}_1^+$, $c_2^+ > \tilde{c}_2^+$ и $c_3^+ = \tilde{c}_3^+$.

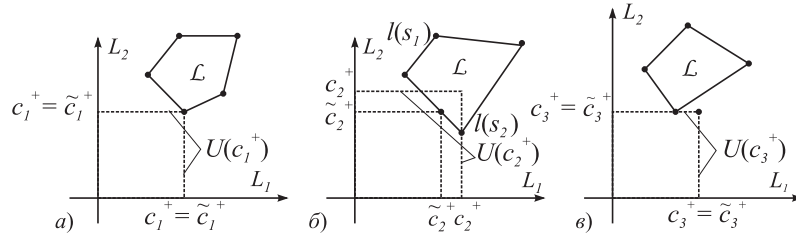


Рис. 2.13. Минимаксные риски c_1^+ , c_2^+ , c_3^+ . Риск $\tilde{c}_2^+$, соответствующий рандомизированному правилу, основанному на чистых правилах s_1 , s_2 , $\tilde{c}_2^+ < c_2^+$.

2.8.3. Байесовское правило решения. Байесовское правило применяется в случае, когда состояние природы ϑ случайно, и известны априорные вероятности состояний природы $p(\vartheta_1), \dots, p(\vartheta_k)$. *Байесовским называется правило решения s^* (называемое также байесовской стратегией), минимизирующее ожидаемый байесовский риск*

$$L(s^*) = \sum_{i=1}^k L_i(s^*)p(\vartheta_i) = \sum_{t=1}^N \sum_{i=1}^k l(\vartheta_i, d_t) p_s(d_t|\vartheta_i) p(\vartheta_i). \quad (8.12)$$

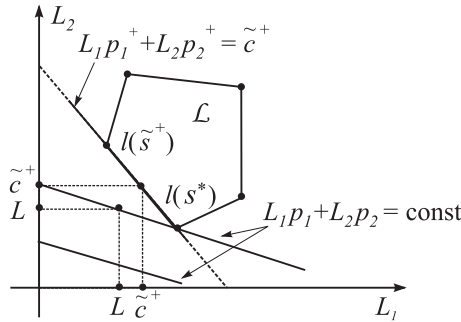


Рис. 2.14. Байесовское правило s^* отвечает точке $l(s^*)$, через которую проходит прямая $L_1p_1 + L_2p_2 = \text{const}$ при значении const , равном ожидаемому байесовскому риску L . Прямая $L_1p_1^+ + L_2p_2^+ = \tilde{c}^+$ отвечает априорным вероятностям состояний природы p_1^+ , p_2^+ , при которых байесовский риск максимален и совпадает с минимаксным $\tilde{c}^+$.

Рассмотрим графическое решение задачи (8.12) в случае $\Theta = \{\vartheta_1, \vartheta_2\}$. Определим на плоскости $\{(L_1, L_2)\}$ семейство прямых $p(\vartheta_1)L_1 + p(\vartheta_2)L_2 = \text{const}$, $p(\vartheta_1) + p(\vartheta_2) = 1$. Очевидно, байесовское правило s^* соответствует первой точке пересечения прямой при ее движении от начала координат, и выпуклого множества $\mathcal{L}$, как это показано на рис. 2.14, где $p_1 = p(\vartheta_1)$, $p_2 = p(\vartheta_2)$, отмечена точка $l(s^*)$, соответствующая байесовскому правилу $s^* \in S$, и значение L байесовского риска, равного значению const для прямой семейства, проходящей через точку $l(s^*)$, где s^* — байесовское правило, при котором $L(s^*)$ в (8.11) равно минимальному значению $\text{const} = L = p(\vartheta_1)L_1 + p(\vartheta_2)L_2$ при $L_1 = L_2 = L$, $l(s^*) = (L, L)$.

Таким образом сформулировано определение байесовского правила и показано, как его найти. Однако существует другой способ отыскания байесовского правила, не требующий рассмотрения всех правил из S . Мы получим этот способ несколько позже, а сейчас заметим, что *минимаксное (рандомизированное) правило может быть получено как частный случай байесовского, если подобрать априорное распределение вероятностей состояний природы так, чтобы соответствующий ожидаемый риск (8.12) оказался максимальным и равным минимаксному*, ¹⁾ см. рис. 2.14, задачу (8.11), рис. 2.12.

2.8.4. Байесовское правило в случае невозможности наблюдений над природой. Если задано априорное распределение вероятностей состояний природы $p(\vartheta_1), \dots, p(\vartheta_k)$, но наблюдения над природой невозможны, то *ожидаемый риск, связанный с действием d_j , равен*

$$L(d_j) = \sum_{i=1}^k l(\vartheta_i, d_j)p(\vartheta_i), \quad j = 1, \dots, N, \quad (8.13)$$

и байесово действие определится из условия

$$L(d_j) \sim \min_j. \quad (8.14)$$

Рассмотрим плоскость $\{(p, L)\}$ и зададим распределения состояний природы в виде $p(\vartheta_1) = p$, $p(\vartheta_2) = 1 - p$, где p — параметр, $0 \leq p \leq 1$. Тогда в зависимости от значения p байесово действие будет d_1 , d_2 или d_3 , как это показано на рис. 2.15, где представлены три прямых $\{(p, L), l(\vartheta_1, d_j)p + l(\vartheta_2, d_j)(1 - p) = L\}$, $j = 1, 2, 3$, определяющих зависимости ожидаемых рисков $L(d_1, p)$, $L(d_2, p)$ и $L(d_3, p)$, обусловленных действиями d_1 , d_2 и d_3 , от параметра $p \in [0, 1]$. Для каждого значения p байесовским будет действие $d_{i(p)}$, для которого $L(d_{i(p)}, p) = \min_j L(d_j, p)$, см. условие (8.14) и рис. 2.15. В точках пересечения прямых возможна рандомизация.

¹⁾ Так подобранное распределение вероятностей состояний природы называется наименее благоприятным.

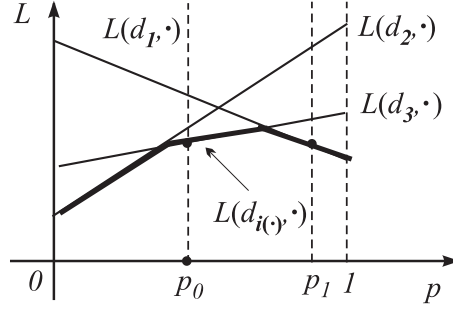


Рис. 2.15. При $p = p_0$ байесово действие есть d_3 , при p_1 есть d_1 , $L(d_{i(p)}, p)$, $p \in [0, 1]$, — выделенная кусочно-линейная кривая.

Заметим, что в рассмотренной ситуации определяется не правило решения, а непосредственно байесово действие. Однако оказывается, что и в общем случае байесовского правила решения, когда производится наблюдение над природой, задача может быть сведена к только что рассмотренной, если *пересчитать априорное распределение состояний природы в условное апостериорное, учитывающее наблюдение за природой*.

2.8.5. Байесово действие. Рассмотрим вместе с правилом s соответствующее разбиение s множества наблюдений

$$X = D_1 + \dots + D_N, \quad D_j = \{x \in X, s(x) = d_j\}, \quad j = 1, \dots, N. \quad (8.15)$$

Согласно (8.15) вероятность действия d_m в состоянии природы ϑ_j равна

$$p_s(d_m|\vartheta_j) = P(\{x \in D_m|\vartheta_j\}) = \sum_{x \in D_m} p(x|\vartheta_j), \quad m = 1, \dots, N, \quad j = 1, \dots, k, \quad (8.16)$$

и выражение (8.12) для ожидаемого риска, свойственного правилу $s \in S$, теперь может быть переписано в виде

$$\begin{aligned} L(s) &= \sum_{i=1}^k \sum_{t=1}^N l(\vartheta_i, d_t) p_s(d_t|\vartheta_i) p(\vartheta_i) = \\ &= \sum_{t=1}^N \sum_{i=1}^k l(\vartheta_i, d_t) p(\vartheta_i) \sum_{x \in D_t} p(x|\vartheta_i). \end{aligned} \quad (8.17)$$

Как следует из (8.17), (8.16), для того чтобы риск (8.17) был минимальным, необходимо и достаточно, чтобы в разбиении (8.15)

$$\begin{aligned} D_t \subset \left\{ x \in X, \sum_{i=1}^k l(\vartheta_i, d_t) p(x|\vartheta_i) p(\vartheta_i) \leq \right. \\ \left. \leq \sum_{i=1}^k l(\vartheta_i, d_j) p(x|\vartheta_i) p(\vartheta_i), \quad j = 1, \dots, N \right\}, \quad t = 1, \dots, N, \end{aligned} \quad (8.18)$$

Соответственно, по наблюдению x следует принять решение d_t , если $x \in D_t$ (8.18), $t = 1, \dots, N$. Если с каждым наблюдением будет связано такое байесовское действие, то ожидаемый риск (8.17) и правило s , соответствующее так определенному разбиению $X = \sum_{j=1}^N D_j$, также будут байесовскими.

Покажем, что решение этой задачи сводится к решению предыдущей с помощью байесовского пересчета априорных вероятностей состояний природы в условные апостериорные, при условии, что при наблюдении над природой получено значение $x \in X$. Решение d_t при условии, что наблюденно x , приводит к условному ожидаемому риску

$$L(d_t|x) \triangleq \sum_{i=1}^k l(\vartheta_i, d_t) p(\vartheta_i|x) = \sum_{i=1}^k l(\vartheta_i, d_t) \left[\frac{p(\vartheta_i)p(x|\vartheta_i)}{\sum_{j=1}^k p(\vartheta_j)p(x|\vartheta_j)} \right]. \quad (8.19)$$

Сравнивая правую часть в (8.19) с левой в (8.18), нетрудно видеть, что условие (8.18), определяющее байесовское действие d_t , эквивалентно следующему условию: *при наблюдении $x \in D_t$ принимается решение d_t , для которого условный ожидаемый риск $L(d_t|x)$ (8.19) минимален. Но выражение (8.19) совпадет с (8.13), если в последнем априорную вероятность $p(\vartheta_i)$ заменить на апостериорную $p(\vartheta_i|x)$, $i = 1, \dots, k$.*

Одновременно, разумеется, определено и байесовское правило s . Однако s теперь определено в терминах байесовских действий: в связи с каждым наблюдением x принимается решение о байесовском действии d . Эти действия и определяют байесовское правило $s(\cdot) : X \rightarrow D$.

Рассмотрим произвольное правило s . Согласно (8.19) правилу $s(\cdot)$ при условии, что наблюденно x , сопутствует условный ожидаемый риск

$$L(s(x)|x) = \sum_{i=1}^k l(\vartheta_i, s(x)) p(\vartheta_i|x) = \sum_{i=1}^k l(\vartheta_i, s(x)) \left[\frac{p(\vartheta_i)p(x|\vartheta_i)}{\sum_{j=1}^k p(\vartheta_j)p(x|\vartheta_j)} \right]. \quad (8.20)$$

Покажем, что ожидаемый риск $L(s)$ (8.12), сопутствующий правилу s , может быть получен из (8.20) усреднением по всем (случайным) наблюдениям $\xi = x \in X$, т. е. что

$$L(s) = M L(s(\xi)|\xi). \quad (8.21)$$

Действительно, математическое ожидание M можно представить в виде $M = \sum_{i=1}^k p(\vartheta_i) M_i$, где M_i — оператор условного математиче-

ского ожидания при условии, что наблюдения распределены согласно $p(x|\vartheta_i)$, $x \in X$, или, иначе говоря, M_i — оператор математического ожидания при условии, что природа находится в состоянии ϑ_i . Пусть $D_j = \{x : s(x) = d_j\}$ и $\chi_{D_j}(\cdot)$ — индикаторная функция D_j , так что

$$P(\{s(x) = d_j|\vartheta_i\}) = \sum_{x \in D_j} p(x|\vartheta_i) = \sum_{x \in D} \chi_{D_j}(x)p(x|\vartheta_i), \quad (8.22)$$

$$i = 1, \dots, k, \quad j = 1, \dots, N.$$

Теперь доказательство следует из цепочки равенств:

$$\begin{aligned} ML(s(\xi)|\xi) &= \sum_{i=1}^k p(\vartheta_i) M_i \left(\sum_{j=1}^N \chi_{D_j} L(s(x)|x) \right) = \\ &= \sum_{i=1}^k \sum_{j=1}^N p(\vartheta_i) \sum_{x \in X} p(x|\vartheta_i) \chi_{D_j}(x) L(s(x)|x) = \\ &= \sum_{j=1}^N \sum_{x \in X} \chi_{D_j}(x) \sum_{i=1}^k l(\vartheta_i, s(x)) p(\vartheta_i) p(x|\vartheta_i) = \\ &= \sum_{j=1}^N \sum_{i=1}^k l(\vartheta_i, d_j) p(d_j|\vartheta_i) p(\vartheta_i) = L(s), \end{aligned}$$

где использованы равенства (8.22) и $l(\vartheta_i, s(x)) = l(\vartheta_i, d_j)$, $x \in D_j$, $j = 1, \dots, N$.

Приведем теперь формальное доказательство того, что последовательность байесовых действий действительно определяет байесовское правило решения (байесовскую стратегию).

Теорема 2.8.1. Пусть $\mathcal{K}$ — класс правил s , таких, что решение $s(x) = d_t$ принимается для $x \in X$, удовлетворяющих неравенствам

$$L(d_t|x) \leq L(d_j|x), \quad j = 1, \dots, N,$$

где $L(d_j|x)$ — условный ожидаемый риск, определенный в (8.19). Тогда для всякого правила $s \in \mathcal{K}$ выполняется неравенство $L(s) \leq L(s')$, где s' — произвольное правило. Иными словами всякое правило из $\mathcal{K}$ является байесовским.

Доказательство. Обозначим $l(x) = \min_t L(d_t|x)$. Тогда $ML(\xi) = ML_{s(\xi)}(\xi)$ при $s \in \mathcal{K}$. Действительно,

$$\begin{aligned} ML(\xi) &= M \sum_{j=1}^N \chi_{D_j} \min_t L(d_t|\xi) = M \sum_{j=1}^N \chi_{D_j} L(d_j|\xi) = \\ &= M \sum_{j=1}^N \chi_{D_j} L(s(\xi)|\xi) = ML(s(\xi)|\xi) = L(s), \end{aligned}$$

где использовано, что при $x \in D_t$

$$L(d_t|x) \leq L(d_j|x) \rightarrow l(x) = L(d_t|x), \quad s(x) = d_t.$$

Отсюда следует, что

$$L(s) = M \min_t L(d_t|\xi) \leq ML(s^*(\xi)|\xi) = L(s^*). \blacksquare$$

2.8.6. Байесовская классификация. Рассмотрим частный случай задачи статистического решения, в котором *действиями* $d \in D$ являются *решения о состоянии природы*. Множество действий в этом случае совпадает с множеством решений. Сохраним прежние обозначения: $\vartheta \in \Theta$ — состояние природы, $d \in D$ — решение о состоянии природы, принятое по наблюдениям $x \in X$ над природой.

Полученные ранее результаты, разумеется, справедливы и в рассматриваемом случае. Байесово действие теперь является байесовским решением.

2.8.7. Правило решения, минимизирующее ожидаемое число ошибок. В задаче решения риск чаще всего оцениваются посредством количества ошибок. Рассмотрим в этой связи некоторые характерные решающие правила. Зададим функцию, определяющую риск потерь, условием

$$l(\vartheta_i, d_j) = 1 - \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad i, j = 1, \dots, k.$$

Тогда в (8.19)

$$L(d_t|x) = \sum_{i=1}^k l(\vartheta_i, d_t) p(\vartheta_i|x) = 1 - p(\vartheta_t|x), \quad (8.23)$$

и соответственно ожидаемый риск дается равенством (8.12)

$$L(s) = \sum_{i,j=1}^k (1 - \delta_{ij}) p(d_j|\vartheta_i) p(\vartheta_i) = \sum_{i=1}^k p(\vartheta_i) (1 - p(d_i|\vartheta_i)). \quad (8.24)$$

Но последнее выражение совпадает с *математическим ожиданием доли ошибочных решений*. Следовательно, *байесовское правило решения в этом случае совпадает с правилом, минимизирующим ожидаемое число ошибок решения*.

По данному наблюдению $x \in X$ байесовское правило предписывает принять решение d_t , для которого условный ожидаемый риск (8.23) минимален. Точнее, правило решения, минимизирующее среднее число ошибок, предписывает принять решение d_t , если наблюдается

$$\begin{aligned} \xi &= x \in D_t \subset \{x \in X : s(x) = d_t\} = \\ &= \{x \in X : p(\vartheta_t|x) \geq p(\vartheta_j|x), \quad j = 1, \dots, k\}, \quad t = 1, \dots, k. \end{aligned} \quad (8.25)$$

Иными словами, речь идет о решении по максимуму апостериорной вероятности состояния природы ϑ_t : *наблюдение* $x \in X$ *относится к*

тому состоянию природы, которое наиболее вероятно при этом наблюдении.

Согласно (8.25) можно также записать

$$D_t = \{x \in X : p(x|\vartheta_t)p(\vartheta_t) \geq p(x|\vartheta_j)p(\vartheta_j), j = 1, \dots, k\}. \quad (8.26)$$

Заметим, что вероятность верно опознать состояние природы ϑ_t , равна

$$P\{s(\xi) = d_t|\vartheta_t\} = \sum_{x \in D_t} p(x|\vartheta_t) = p(d_t|\vartheta_t), \quad t = 1, \dots, k,$$

и математическое ожидание доли верных решений дается равенством

$$P = \sum_{t=1}^k p(d_t|\vartheta_t)p(\vartheta_t) = \sum_{t=1}^k \sum_{x \in D_t} p(x|\vartheta_t)p(\vartheta_t).$$

Если априорные вероятности состояния природы одинаковы, то в (8.26)

$$D_t \subset \{x \in X : p(x|\vartheta_t) \geq p(x|\vartheta_j), j = 1, \dots, k\}, \quad t = 1, \dots, k,$$

и речь идет о решении по принципу максимума правдоподобия. Решение по принципу максимума правдоподобия есть частный случай байесовского при функции риска потерь $l(\vartheta_i, d_j) = 1 - \delta_{ij}$, $i, j = 1, \dots, k$, и равных априорных вероятностях состояний природы.

2.9. Приложение

В разделе курса, посвященном математической статистике, постоянно используются χ_n^2 -распределение (распределение Пирсона), t_n -распределение (распределение Стьюдента) и $F_{k,m}$ -распределение (распределение Фишера). Вычислим плотности этих распределений и их числовые характеристики.

Определение 2.9.1. χ_n^2 -распределением с n степенями свободы называется распределение случайной величины $\chi_n^2 = \xi_1^2 + \dots + \xi_n^2$, где все $\xi_i \sim \mathcal{N}(0, 1)$ и независимы. Найдем плотность распределения χ_n^2 . Случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ имеет плотность $p(x_1, \dots, x_n) = 1/(2\pi)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}$, $(x_1, \dots, x_n) \in \mathcal{R}^n$, поэтому

$$F(x) = P(\{\chi_n^2 < x\}) = \int \dots \int_{\sum_{i=1}^n x_i^2 < x} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} dx_1 \dots dx_n. \quad (9.1)$$

Введем сферические координаты в $\mathcal{R}^n$: $x_1 = \rho \cos \varphi_1$, $x_2 = \rho \sin \varphi_1 \cos \varphi_2$, ..., $x_n = \rho \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{n-1}$, $-\pi/2 \leq \varphi_i \leq \pi/2$, $i = 1, \dots, n-2$, $-\pi \leq \varphi_{n-1} \leq \pi$. Тогда

$$\begin{aligned} F(x) &= \frac{1}{(2\pi)^{n/2}} \int_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho \int_{-\pi/2}^{\pi/2} \dots \int_{-\pi/2}^{\pi/2} J(\varphi_1, \dots, \varphi_{n-1}) d\varphi_1 \dots d\varphi_{n-1} = \\ &= \frac{\omega_{n-1}}{(2\pi)^{n/2}} \int_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho, \end{aligned} \quad (9.2)$$

где ω_{n-1} — площадь поверхности единичной $(n-1)$ -мерной сферы; $\omega_{n-1} = \int_{-\pi/2}^{\pi/2} \dots \int_{-\pi/2}^{\pi/2} J(\varphi_1, \dots, \varphi_{n-1}) d\varphi_1 \dots d\varphi_{n-1}$ легко вычислить исходя из формулы (9.2). Поскольку $F(+\infty) = 1$, то $1 = \frac{\omega_{n-1}}{(2\pi)^{n/2}} \int_0^{\infty} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho = \frac{\omega_{n-1}}{(2\pi)^{n/2}} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}-1}$. Таким образом, $\omega_{n-1} = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}$. Подставляя ω_{n-1} в (9.2), находим окончательно

$$F(x) = \frac{1}{2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)} \int_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho, \quad x > 0. \quad (9.3)$$

Поскольку, очевидно, $P(\{\chi_n^2 < x\}) = 0$ при $x \leq 0$, то $F(x) = 0$, $x \leq 0$. Отсюда и из (9.3)

$$p_{\chi_n^2} = F'(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0. \end{cases} \quad (9.4)$$

Пользуясь (9.4), найдем $M\chi_n^2$ и $D\chi_n^2$:

$$M\chi_n^2 = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} x^{n/2} e^{-x/2} dx = \frac{2^{(n/2)+1} \Gamma\left(\frac{n}{2} + 1\right)}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} = n, \quad (9.5)$$

поскольку $\Gamma(s+1) = s\Gamma(s)$; $D\chi_n^2 = M(\chi_n^2)^2 - (M\chi_n^2)^2 = M(\chi_n^2)^2 - n^2$, и, так как

$$M(\chi_n^2)^2 = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} x^{n/2+1} e^{-x/2} dx = \frac{2^{(n/2)+2} \Gamma\left(\frac{n}{2} + 2\right)}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} = n^2 + 2n, \quad (9.6)$$

то

$$D\chi_n^2 = 2n. \quad (9.7)$$

Определение 2.9.2. t_n -распределением (Стьюдента) с n степенями свободы называется распределение случайной величины $t_n = \xi/\eta$, где $\xi \in N(0, 1)$, $\eta = \sqrt{\chi_n^2/n}$, причем ξ и η независимы. В силу последнего условия случайный вектор (ξ, η) имеет плотность $p(x_1, x_2) = p_\xi(x_1)p_\eta(x_2)$, а плотность частного ξ/η определяется на основании формулы (8.66)

$$p_{t_n}(x) = F'_{t_n}(x) = \int_0^\infty zp_\xi(zx)p_\eta(z)dz - \int_{-\infty}^0 zp_\xi(zx)p_\eta(z)dz, \quad x \in \mathcal{R}^1. \quad (9.8)$$

Далее, $F_\eta(x) = P\left(\left\{\sqrt{\frac{\chi_n^2}{n}} < x\right\}\right) = P(\{\chi_n^2 < x^2 n\}) = F_{\chi_n^2}(x^2 n)$, поэтому согласно (9.5) плотность распределения η равна

$$p_\eta(x) = F'_\eta(x) = p_{\chi_n^2}(x^2 n) \cdot 2xn = \begin{cases} \frac{n^{n/2}}{2^{(n/2)-1}\Gamma(\frac{n}{2})} x^{n-1} e^{-x^2 n/2}, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (9.9)$$

Ввиду (9.9) второй интеграл в (9.8) исчезает, и находим

$$\begin{aligned} p_{t_n}(x) &= \frac{n^{n/2}}{2^{(n/2)-1}\Gamma(\frac{n}{2})\sqrt{2\pi}} \int_0^\infty z^n e^{-\frac{(zx)^2}{2} - \frac{nz^2}{2}} dz = \\ &= \frac{n^{n/2} 2^{(n/2)-1/2}}{2^{(n/2)-1}\Gamma(\frac{n}{2})\sqrt{2\pi}} \frac{1}{(x^2 + n)^{(n/2)+1/2}} \int_0^\infty t^{(n-1)/2} e^{-t} dt = \\ &= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}. \end{aligned} \quad (9.10)$$

С помощью (9.10) вычисляем Mt_n , $n \geq 2$,

$$Mt_n = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\sqrt{\pi n}} \int_{-\infty}^\infty x \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} dx = 0 \quad (9.11)$$

ввиду нечетности подынтегральной функции. Теперь

$$Dt_n = Mt_n^2 = M\left(\frac{\xi^2}{\chi_n^2/n}\right) = M\xi^2 M\left(\frac{n}{\chi_n^2}\right), \quad (9.12)$$

так как ξ^2 и χ_n^2/n независимы. Обозначим $\eta = n/\chi_n^2$, тогда при $x > 0$

$$F_\eta(x) = P\left(\left\{\frac{n}{\chi_n^2} < x\right\}\right) = P\left(\left\{\frac{1}{x} < \frac{\chi_n^2}{n}\right\}\right) = 1 - P\left(\left\{\chi_n^2 < \frac{n}{x}\right\}\right).$$

Отсюда, используя (9.4) с n/x вместо x , найдем

$$p_\eta(x) = \begin{cases} \left(\frac{n}{2}\right)^{n/2} \frac{1}{\Gamma\left(\frac{n}{2}\right)} e^{-n/(2x)} x^{-(n/2+1)}, & x > 0 \\ 0, & x \leq 0. \end{cases} \quad (9.13)$$

Поэтому

$$\begin{aligned} M\left(\frac{n}{\chi_n^2}\right) &= M\eta = \left(\frac{n}{2}\right)^{n/2} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty x^{-n/2} e^{-n/(2x)} dx = \\ &= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \frac{n}{2} \int_0^\infty t^{\frac{n}{2}-2} e^{-t} dt = \frac{n}{n-2}. \end{aligned} \quad (9.14)$$

Сопоставляя (9.12) и (9.14) и учитывая, что $M\xi^2 = D\xi = 1$, получим $Dt_n = \frac{n}{n-2}$.

Определение 2.9.3. $F_{k,m}$ -распределением с (k, m) степенями свободы называется распределение случайной величины $F_{k,m} = \frac{\chi_k^2/k}{\chi_m^2/m}$, где χ_k^2, χ_m^2 независимы. Положим для краткости $F_{k,m} = \eta$, $\chi_k^2/k = \xi_1$, $\chi_m^2/m = \xi_2$. Ввиду независимости ξ_1 и ξ_2 плотность случайного вектора (ξ_1, ξ_2) равна $p(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$, $x_1, x_2 > 0$, а плотность частного ξ_1/ξ_2 равна (см. (9.8))

$$p_\eta(x) = \int_0^\infty zp_{\xi_1}(zx)p_{\xi_2}(z) dz - \int_{-\infty}^0 zp_{\xi_1}(zx)p_{\xi_2}(z) dz, \quad x > 0.$$

Отсюда с учетом (9.5) находим при $x > 0$

$$\begin{aligned} p_\eta(x) &= \frac{k^{\frac{k}{2}-1} m^{\frac{m}{2}-1} x^{\frac{k}{2}-1}}{2^{\frac{k+m}{2}} \Gamma\left(\frac{k}{2}\right) \Gamma\left(\frac{m}{2}\right)} \int_0^\infty z^{(k+m)/2-1} e^{-z(kx+m)/2} dz = \\ &= \frac{k^{\frac{k}{2}-1} m^{\frac{m}{2}-1} x^{\frac{k}{2}-1} (kx+m)^{-\frac{k+m}{2}}}{\Gamma\left(\frac{k}{2}\right) \Gamma\left(\frac{m}{2}\right)} \Gamma\left(\frac{k+m}{2}\right), \end{aligned} \quad (9.15)$$

и $p_\eta(x) = 0$ при $x \leq 0$. Далее, в силу (9.6) и (9.14)

$$M\eta = M\left(\frac{\chi_k^2/k}{\chi_m^2/m}\right) = M\left(\frac{\chi_k^2}{k}\right) M\left(\frac{m}{\chi_m^2}\right) = \frac{k}{k} \cdot \frac{m}{m-2} = \frac{m}{m-2}. \quad (9.16)$$

Найдем $D\eta$:

$$M\eta = M\eta^2 - (M\eta)^2 = M\eta^2 - \left(\frac{m}{m-2}\right)^2. \quad (9.17)$$

Но

$$M\eta^2 = M\left(\frac{\chi_k^2}{k}\right)^2 M\left(\frac{m}{\chi_m^2}\right)^2, \quad (9.18)$$

где

$$M\left(\frac{\chi_k^2}{k}\right)^2 = \frac{k^2 + 2k}{k^2} \quad (9.19)$$

ввиду (9.6). Используя (9.13), находим

$$\begin{aligned} \left(\frac{m}{\chi_m^2}\right)^2 &= \left(\frac{m}{2}\right)^{\frac{m}{2}} \frac{1}{\Gamma\left(\frac{m}{2}\right)} \int_0^\infty x^{-\frac{m}{2}+1} e^{-\frac{m}{2}x} dx = \\ &= \left(\frac{m}{2}\right)^{\frac{m}{2}} \frac{1}{\Gamma\left(\frac{m}{2}\right)} \left(\frac{m}{2}\right)^{-\frac{m}{2}+2} \int_0^\infty e^{-z} z^{\frac{m}{2}-3} dz = \\ &= \left(\frac{m}{2}\right)^2 \frac{\Gamma\left(\frac{m}{2}-2\right)}{\Gamma\left(\frac{m}{2}\right)} = \frac{m^2}{(m-4)(m-2)}. \end{aligned} \quad (9.20)$$

Собирая выражения (9.17)–(9.20), получим при $m > 4$

$$D\eta = \frac{2m^2}{(m-2)^2(m-4)} \left(1 + \frac{m-2}{k}\right). \quad (9.21)$$

Глава 3

ЭЛЕМЕНТЫ ТЕОРИИ ВОЗМОЖНОСТЕЙ И ЕЁ ПРИМЕНЕНИЯ

3.0. Предисловие

Теория вероятностей как математическая модель случайности широко используется в теоретических и прикладных исследованиях благодаря двум ее *фундаментальным* аспектам:

- ▶ математическому, согласно которому теория вероятностей есть специальный раздел теории меры и интеграла, и
- ▶ эмпирическому, основанному на математически обоснованных процедурах, позволяющих, пользуясь данными событийно-частотных наблюдений, *при определенных условиях* получать сколь угодно точные аппроксимации вероятности, математического ожидания и т.п., с одной стороны, а с другой — как предсказывать, так и эмпирически проверять предсказанные значения математически аккуратно определенных характеристик случайных процессов и явлений.

Вместе с тем вероятностные методы де-факто оказались неэффективными при моделировании сложных физических, технических, социальных и экономических объектов, субъективных суждений и т.д. Этим объясняется повышенный интерес к невероятностным моделям случайности, нечеткости и неопределенности, характерный для второй половины прошлого столетия, см. [5, 7, 8, 26, 35, 37–39].

Неэффективность вероятностных методов обусловлена многими факторами. Во-первых, неточность и нечеткость, свойственная формулировкам моделей названных объектов, зачастую не может быть охарактеризована в вероятностных терминах ¹⁾, а во-вторых, если в модели и удастся выделить стохастическую компоненту ²⁾, то возникают серьезные проблемы с эмпирическим построением и верификацией ее вероятностной модели. Дело в том, что для эмпирического построения

¹⁾ На самом деле, эмпирический аспект далеко не так фундаментален, как математический, поскольку не содержит критерия вероятностной природы наблюдаемого феномена случайности и, как следствие, не содержит критерия упомянутых "определенных условий". Поэтому не исключено, что нечеткость и неточность, свойственная формулировкам моделей названных объектов, не может быть охарактеризована в терминах теории вероятностей, но теория вероятностей не позволяет ни подтвердить это, ни опровергнуть.

²⁾ То есть такую компоненту, математической моделью которой в каждый момент является некоторое вероятностное пространство.

вероятностной модели сложного объекта, равно как и для ее верификации, требуются большие объемы данных наблюдений, которые в конечном счете обычно оказываются неполными, неточными и противоречивыми, поскольку за время получения данных объект и его окружение заметно эволюционируют, его вероятностные характеристики, как правило, непредсказуемо изменяются, а их оценки, естественно, оказываются неадекватными. В таких случаях эмпирическое построение вероятностной модели объекта невозможно, если данные наблюдений не позволяют оценить *эволюцию его вероятностных свойств*.

Рассматриваемая в этой главе теория возможностей лучше, чем теория вероятностей, приспособлена для математического моделирования упомянутых выше, в том числе стохастических, объектов. Хотя формально возможность, вообще говоря, не связана с вероятностью, ее математический аспект и схема построения подобны математическому аспекту и схеме построения вероятности. Поэтому идеи, лежащие в основе теории возможностей, удобно изложить, представив возможность как альтернативную вероятности математическую модель случайности и сравнив основные математические конструкции теорий вероятностей и возможностей, их эмпирическое построение, содержательную интерпретацию и прикладную ориентацию.

1. Вероятность. Проблемы эмпирического построения и интерпретации. Как известно, эмпирическая интерпретация вероятности, называемая статистической (или событийно-частотной), основана на законах больших чисел (З.Б.Ч.). Пусть $\nu^{(N)}(A)$ — частота события $A \in \mathcal{A}$ в серии N взаимно независимых испытаний, модель которых обозначим ¹⁾ $(\Omega, \mathcal{A}, \text{Pr}) \times \dots \times (\Omega, \mathcal{A}, \text{Pr}) = (\Omega, \mathcal{A}, \text{Pr})^N$, тогда вероятность любого отклонения $\nu^{(N)}(A)$ от $\text{Pr}(A)$ стремится к нулю при $N \rightarrow \infty$. Точнее $\forall A \in \mathcal{A} \forall \varepsilon > 0$

$$\lim_{N \rightarrow \infty} \text{Pr}(|\nu^{(N)}(A) - \text{Pr}(A)| > \varepsilon) = 0$$

(слабый З.Б.Ч., $\nu^{(N)}(A) \xrightarrow[N \rightarrow \infty]{\text{Pr}} \text{Pr}(A)$). Более того $\forall \varepsilon > 0 \forall A \in \mathcal{A}$

$$\lim_{N \rightarrow \infty} \text{Pr}^\infty(\sup_{n \geq N} |\nu^{(n)}(A) - \text{Pr}(A)| > \varepsilon) = 0, \quad (0.1)$$

т. е. частота $\nu^{(n)}(A)$ с увеличением n приближается и остается близкой к $\text{Pr}(A)$, ибо согласно (0.1) $\forall \varepsilon > 0 \forall A \in \mathcal{A} |\nu^{(n)}(A) - \text{Pr}(A)| > \varepsilon$ лишь для Pr^∞ -почти на верное ²⁾ (п. н.) конечного числа n испытаний (усиленный З.Б.Ч.).

¹⁾ Чтобы отличать вероятность (probability) от возможности (possibility), первую далее обозначим Pr , вторую — P .

²⁾ Условие (0.1) определяет сходимость $\nu^{(N)}(A)$ к $\text{Pr}(A)$ Pr^∞ -почти на верное (п. н.). Pr^∞ — вероятность, определенная на борелевских множествах бесконечных последовательностей испытаний, см, например [13].

Закон больших чисел (0.1) определяет *эмпирическую интерпретацию вероятности*, согласно которой вероятность любого исхода стохастического эксперимента ¹⁾ (С.Э.) сколь угодно точно оценивает его частоту в достаточно длинной последовательности взаимно независимых испытаний, и наоборот, — при этих условиях частота любого исхода сколь угодно точно оценивает его вероятность, а, следовательно, — и модель С.Э.

Однако, если в процессе испытаний вероятностные свойства С.Э. произвольно изменяются, то частота каждого исхода испытаний, вообще говоря, не характеризует его вероятность, а результаты испытаний не позволяют восстановить вероятностную модель С.Э. В то же время, как будет показано далее, при известных ограничениях на характер эволюции вероятностных свойств С.Э., его возможностная модель может быть восстановлена эмпирически, причем — безошибочно, на основе п. н. конечного числа наблюдений, см. [13].

Установить этот факт, существенно расширяющий класс стохастических объектов ²⁾, математические модели которых могут быть построены эмпирически, позволит усиленный З.Б.Ч. для последовательности взаимно независимых испытаний, в модели которых $(\Omega, \mathcal{A}, \text{Pr}_{(1)}) \times (\Omega, \mathcal{A}, \text{Pr}_{(2)}) \times \dots \times (\Omega, \mathcal{A}, \text{Pr}_{(n)}) \times \dots$ вероятность произвольно изменяется от испытания к испытанию.

Согласно усиленному З.Б.Ч. в этом случае $\forall \varepsilon > 0$

$$\lim_{N \rightarrow \infty} \text{Pr}^\infty \left(\sup_{n \geq N} \left| \nu^{(n)}(A) - \frac{1}{n} \sum_{i=1}^n \text{Pr}_{(i)}(A) \right| > \varepsilon \right) = 0, \quad (0.2)$$

т. е. частота $\nu^{(n)}(A)$ с увеличением n всё более точно следует за вероятностью $\text{Pr}^{(n)}(A) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \text{Pr}_{(i)}(A)$, ибо согласно (0.2) $\forall \varepsilon > 0$ $\forall A \in \mathcal{A}$ $|\nu^{(n)}(A) - \text{Pr}^{(n)}(A)| > \varepsilon$ лишь для Pr^∞ -п. н. конечного числа n

¹⁾ Стохастическим назовем эксперимент, математической моделью которого в каждый момент является некоторое вероятностное пространство, его повторения назовем испытаниями, математической моделью каждого испытания (повторения) является некоторое вероятностное пространство. Напомним, что в теории вероятностей, рассмотренной в гл. 1, математической моделью стохастического эксперимента является вероятностное пространство, *фиксированное* некоторым комплексом условий S , при которых выполняется эксперимент. Математической моделью такого С.Э. и каждого испытания является вероятностное пространство $(\Omega, \mathcal{A}, \text{Pr})$, моделью взаимно независимых испытаний — вероятностное пространство $(\Omega, \mathcal{A}, \text{Pr}) \times (\Omega, \mathcal{A}, \text{Pr}) \times \dots$

²⁾ Стохастическим назовем объект, математической моделью которого в каждый момент является (некоторое) вероятностное пространство.

испытаний ¹⁾ (усиленный З. Б. Ч., $\nu^{(n)}(A) - \text{Pr}^{(n)}(A) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0$), см., например, [10]. Это означает, что наблюдения частот $\nu^{(n)}(A)$, $n = 1, 2, \dots$, $A \in \mathcal{A}$, не позволят эмпирически охарактеризовать стохастическую модель испытаний $(\Omega, \mathcal{A}, \text{Pr}_{(1)}) \times \dots \times (\Omega, \mathcal{A}, \text{Pr}_{(n)}) \times \dots$. Более детально вопрос эмпирического построения модели стохастического объекта рассмотрен в § 8 и в § 3.3.

2. Возможность как мера предопределенности исходов стохастического эксперимента. В монографии [13] представлена точка зрения на возможность, согласно которой последняя, как и вероятность, является мерой, определяемой моделируемыми свойствами объекта исследования, значения которой полностью *упорядочивают предопределенности, шансы проявлений* этих свойств.

Рассмотрим идеи, лежащие в основе построения возможности, еще раз подчеркнув, что значение $\text{Pr}(A)$ следует рассматривать как прогнозируемое значение *частоты* исхода A в серии независимых испытаний, но не как меру предопределенности или — *возможности* исхода A при каждом испытании ²⁾.

Что можно сказать о предопределенности исхода стохастического эксперимента (С.Э.), если его моделью является вероятностное пространство $(\Omega, \mathcal{P}(\Omega), \text{Pr})$, в котором $\Omega = \{\omega_1, \omega_2, \dots\}$? В частности, что можно сказать о *возможностях* $\{\omega_i\}$, $i = 1, 2, \dots$, в этом случае, — об их *шансах*? Ясно лишь, что при любом определении возможности p_i исхода $\{\omega_i\} \subset \Omega$ как значения меры (возможности $\text{Pr}(\cdot)$), *при каждом испытании оценивающей, обусловленный свойствами С.Э., шанс его исхода $\{\omega_i\}$ в сравнении с шансами всех других его элементарных исходов*, естественно считать, что $\text{Pr}(\{\omega_i\}) \stackrel{\text{def}}{=} p_i \geq p_j \stackrel{\text{def}}{=} \text{Pr}(\{\omega_j\})$, если $\text{Pr}(\{\omega_i\}) \stackrel{\text{def}}{=} p_i \geq p_j \stackrel{\text{def}}{=} \text{Pr}(\{\omega_j\})$. Индуктивная мотивация в данном случае такова: чем больше вероятность p_i исхода $\{\omega_i\}$, тем чаще $\{\omega_i\}$ встретится в длинной серии испытаний, и, следовательно, тем более возможен (ожидаем, предопределен) исход $\{\omega_i\}$ в каждом очередном испытании.

В данном случае принципиально то, что для такого заключения *не требуются значения вероятностей* $p_1, p_2, \dots$, *достаточно лишь знать, как они упорядочены*. Более того, такое заключение останется

¹⁾ Событие $S_\varepsilon = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} S_\varepsilon^{(n)}$ происходит, если и только если происходит бесконечно много событий среди $S_\varepsilon^{(1)}, S_\varepsilon^{(2)}, \dots$. Согласно (0.2) для $S_\varepsilon^{(n)} = \{|\nu^{(n)}(A) - \text{Pr}^{(n)}(A)| > \varepsilon\}$, $n = 1, 2, \dots$, $\forall \varepsilon > 0$ $\text{Pr}^\infty(S_\varepsilon) = \lim_{N \rightarrow \infty} \text{Pr}^\infty(\bigcup_{n \geq N} S_\varepsilon^{(n)}) = \lim_{N \rightarrow \infty} \text{Pr}^\infty(\sup_{n \geq N} |\nu^{(n)}(A) - \text{Pr}^{(n)}(A)| > \varepsilon) = 0$.

²⁾ На моделирование модальностей индуктивных суждений типа «вероятно», «возможно», «необходимо» и т.п. ориентированы математические теории интуитивных [28] и субъективных [3, 29, 33] вероятностей.

в силе, если допустить, что *вероятности* $pr_1, pr_2, \dots$ *изменяются от испытания к испытанию, оставаясь лишь одинаково упорядоченными, например, согласно условию*

$$1 \geq pr_1 \geq pr_2 \geq \dots > 0, \quad pr_1 + pr_2 + \dots = 1. \quad (0.3)$$

В связи с этими замечаниями рассмотрим $\widetilde{C.Э.}$, математической моделью которого является класс $\mathcal{Pr} \stackrel{\text{def}}{=} \{(\Omega, \mathcal{P}(\Omega), Pr), Pr \in \mathbb{Pr}\}$ *дискретных вероятностных пространств, где* $\mathbb{Pr}$ *— класс всех вероятностей* $Pr(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$, *удовлетворяющих условию (0.3).*

Знания одной лишь упорядоченности (0.3) вероятностей $pr_1, pr_2, \dots$, конечно, недостаточно, чтобы охарактеризовать $\widetilde{C.Э.}$ в терминах стандартного формализма теории вероятностей, а допустив, что в рамках условия (0.3) значения $pr_1, pr_2, \dots$ произвольно изменяются от испытания к испытанию¹⁾, утратим возможность эмпирически оценивать вероятностную модель $(\Omega, \mathcal{P}(\Omega), Pr_{(1)}) \times (\Omega, \mathcal{P}(\Omega), Pr_{(2)}) \times \dots, Pr_{(i)} \in \mathbb{Pr}, i = 1, 2, \dots$ испытаний.

3. Классы эквивалентных возможностей. В любой *возможностной модели* $\widetilde{C.Э.}$, модель которого определена как класс $\mathcal{Pr}$, возможности $p_i \stackrel{\text{def}}{=} P(\{\omega_i\}), i = 1, 2, \dots$, должны быть априори подчинены условиям

$$1 = p_1 \geq p_2 \geq \dots \geq 0, \quad (0.4)$$

содержащим лишь требование упорядоченности, согласованной с упорядоченностью в (0.3), и условие нормировки $p_1 = 1$.

Предположим, что распределение²⁾ $p_1, p_2, \dots$ определяет возможность как функцию $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ и соответствующее пространство с возможностью $(\Omega, \mathcal{P}(\Omega), P)$, а каждая конкретная упорядоченность в (0.4), в которой встречаются только равенства и строгие неравенства, определяет класс взаимно эквивалентных возможностей, распределения которых одинаково упорядочены, и соответствующий класс эквивалентных пространств с возможностью $(\Omega, \mathcal{P}(\Omega), P)$.

¹⁾ Вероятностной моделью каждого испытания является некоторое $(\Omega, \mathcal{P}(\Omega), Pr) \in \mathcal{Pr}$.

²⁾ Нестандартный в русскоязычной литературе термин «распределение вероятности» используется с целью расширения «терминологического единства» теорий вероятностей и возможностей, поскольку в этой главе вероятность и возможность рассматриваются как метафизические понятия, характеризующие объект исследования подобно метафизическим понятиям силы, тепла и т. п. в духе пропенсивной интерпретации вероятности К. Поппера [11]. В теории вероятностей «распределение» и «вероятность», как правило, — синонимы; условимся термин «распределение вероятностей (возможностей)» понимать как распределение вероятностей (возможностей) элементарных событий, значений случайного (нечеткого) элемента и т. п.

Обозначим $\mathbb{P}$ класс возможностей, распределенных согласно (0.4), и $\mathcal{P} \stackrel{\text{def}}{=} \{(\Omega, \mathcal{P}(\Omega), P), P \in \mathbb{P}\}$ — соответствующий класс пространств с возможностью, т. е. — *возможностную модель* $\widetilde{\text{С.Э.}}$.

Представим класс $\mathbb{P}$ в виде разбиения на попарно непересекающиеся подклассы *взаимно эквивалентных возможностей*, каждый из которых определит *неприводимую, не допускающую дальнейшего уточнения и детализации* (единственную с точностью до эквивалентности) возможность модель $(\Omega, \mathcal{P}(\Omega), P)$. С этой целью заметим, что *всякую конкретную упорядоченность* значений $p_1, p_2, \dots$ в (0.4) можно задать двоичным числом $e = 0.e_1e_2\dots \in (0, 1)$, в котором $e_i = 1$, если $p_i > p_{i+1}$, и $e_i = 0$, если $p_i = p_{i+1}$, $i = 1, 2, \dots$. Обозначим $\mathbb{P}_{(e)}$ *класс возможностей, упорядоченность распределений* $p_1, p_2, \dots$ *которых определена значением* $e \in (0, 1)$. Тогда $\mathbb{P}_{(e)} \cap \mathbb{P}_{(e')} = \emptyset$, если $e \neq e'$, и

$$\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}. \quad (0.5)$$

Подчеркнем, что для любого $e \in (0, 1)$ *все возможности* $P \in \mathbb{P}_{(e)}$ *попарно эквивалентны*, а каждый класс $\mathbb{P}_{(e)}$ определяет *неприводимую* возможность модель

$$\mathcal{P}_{(e)} = \{(\Omega, \mathcal{P}(\Omega), P), P \in \mathbb{P}_{(e)}\}, \quad e \in (0, 1), \quad (0.6)$$

в которой все пространства с возможностью $(\Omega, \mathcal{P}(\Omega), P)$, $P \in \mathbb{P}_{(e)}$, определяют *единственную с точностью до эквивалентности возможность модель*; при этом $\mathcal{P} = \bigcup_{e \in (0,1)} \mathcal{P}_{(e)}$ — *возможностная*

модель $\widetilde{\text{С.Э.}}$, где $\mathcal{P}_{(e)} \cap \mathcal{P}_{(e')} = \emptyset$, $e \neq e'$, $e, e' \in (0, 1)$.

4. Шкала значений возможности. Возможность события. Рассмотрим, как распределение $p_i \stackrel{\text{def}}{=} P(\{\omega_i\})$, $i = 1, 2, \dots$, определяет возможность $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$. Поскольку каждая *конкретная упорядоченность* в (0.4) определяет единственную (с точностью до эквивалентности) возможность модель, то следует так определить возможность P и шкалу $\mathcal{L}$ ее значений, в частности, — операции сложения и умножения, чтобы конкретная упорядоченность в (0.4) *конкретно упорядочивала значения возможностей всех подмножеств* Ω .

Определим *шкалу* $\mathcal{L}$ *значений возможности* как интервал $[0, 1]$ с естественной упорядоченностью $\leq$ и двумя бинарными операциями — сложением $+$: $[0, 1] \times [0, 1] \rightarrow [0, 1]$ и умножением $\bullet$: $[0, 1] \times [0, 1] \rightarrow [0, 1]$, т. е. определим четверку $\mathcal{L} = ([0, 1], \leq, +, \bullet)$, и *группу* Γ *изотонных (сохраняющих конкретную упорядоченность¹⁾) автоморфизмов* $\mathcal{L}$, порожденную группой строго монотонных непрерывных функций $\gamma(\cdot) : [0, 1] \rightarrow [0, 1]$, удовлетворяющих условиям $\gamma(0) =$

¹⁾ В (0.5) $\forall \gamma(\cdot) \in \Gamma \quad \gamma \circ \mathbb{P}_{(e)} = \mathbb{P}_{(e)}$, где $\gamma \circ \mathbb{P}_{(e)} \stackrel{\text{def}}{=} \{\gamma \circ P, P \in \mathbb{P}_{(e)}\}$, $\gamma \circ P(A) \stackrel{\text{def}}{=} \gamma(P(A))$, $A \in \mathcal{P}(\Omega)$, $e \in (0, 1)$.

$= 0$, $\gamma(1) = 1$, с групповой операцией $\circ$, определённой равенством $\gamma' \circ \gamma(a) \stackrel{\text{def}}{=} \gamma'(\gamma(a))$, $a \in [0, 1]$. Поскольку Γ — группа автоморфизмов $\mathcal{L}$, то для любых $a, b \in [0, 1]$ и $\gamma(\cdot) \in \Gamma$ должны быть выполнены соотношения

$$\begin{aligned} a * b &\iff \gamma(a) * \gamma(b), & \gamma(a + b) &= \gamma(a) + \gamma(b), \\ \gamma(a \bullet b) &= \gamma(a) \bullet \gamma(b), & \gamma(0) &= 0, \quad \gamma(1) = 1, \end{aligned} \quad (0.7)$$

где $*$ означает либо $<$, либо $>$, либо $=$.

Теорема. Если

- операции $+$ и $\bullet$ суть непрерывные отображения из $[0, 1] \times [0, 1]$ в $[0, 1]$,
- для любых $a, b \in [0, 1]$

$$\begin{aligned} a \bullet b &= b \bullet a, & 0 \bullet a &= 0, & 1 \bullet a &= a, \\ a + b &= b + a, & 0 + a &= a, & 1 + a &= 1, \end{aligned} \quad \text{и} \quad (0.8)$$

- для любой функции $\gamma(\cdot) \in \Gamma$ выполнены условия¹⁾ (0.7), то $a + b = \max(a, b)$, $a \bullet b = \min(a, b)$, $a, b \in [0, 1]$. См. [13].

Определённые в теореме операции $+$ и $\bullet$ позволяют сформулировать правило, определяющее возможность по ее распределению,

$$P(A) \stackrel{\text{def}}{=} \bigoplus_{i: \omega_i \in A} p_i \stackrel{\text{def}}{=} \sup_{i: \omega_i \in A} P(\{\omega_i\}), \quad (0.9)$$

аналогичное правилу для вероятности, согласно которому

$$\Pr(A) = \sum_{i: \omega_i \in A} \text{pr}_i = \sum_{i: \omega_i \in A} \Pr(\{\omega_i\}), \quad A \in \mathcal{A}. \quad (0.9^*)$$

Согласно (0.9) конкретная упорядоченность в (0.4) конкретно упорядочивает значения возможностей всех подмножеств Ω , ибо если $A = \{\omega_{i_1}, \omega_{i_2}, \dots\}$, $i_1 < i_2 < \dots$, и $B = \{\omega_{j_1}, \omega_{j_2}, \dots\}$, $j_1 < j_2 < \dots$, то $P(A) = p_{i_1} \leq p_{j_1} = P(B)$, если и только если $i_1 \geq j_1$, причем $P(A) < P(B)$, если $\max_{j_1 \leq k < i_1} e_k = 1$, $P(A) = P(B)$, если $\max_{j_1 \leq k < i_1} e_k = 0$

или $j_1 = i_1$. Кроме этого, согласно (0.9) $P(\Omega) = 1$, $P(\emptyset) \stackrel{\text{def}}{=} 0$, и, как нетрудно проверить, для любых $A, B \in \mathcal{P}(\Omega)$:

- $A \subset B \Rightarrow P(A) \leq P(B)$,
- $P(A \cup B) = \max(P(A), P(B)) \stackrel{\text{def}}{=} P(A) + P(B)$,
- $P(A \cap B) \leq \min(P(A), P(B)) \stackrel{\text{def}}{=} P(A) \bullet P(B)$, в случае

¹⁾ Согласно (0.7) и (0.8) 0 и 1 суть нейтральные элементы $\mathcal{L}$.

P -независимости A и B естественно считать, что ¹⁾ $P(A \cap B) = P(A) \bullet P(B)$.

5. Необходимость. Шкала значений необходимости. Поскольку возможности противоположных событий связывает лишь равенство $P(\Omega) = \max(P(A), P(\Omega \setminus A)) \stackrel{\text{def}}{=} P(A) + P(\Omega \setminus A) = 1$, $A \in \mathcal{P}(\Omega)$, не позволяющее выразить $P(\Omega \setminus A)$ через $P(A)$, каждое событие естественно охарактеризовать значениями двух, априори не связанных мер, — возможности $P(\cdot) : \mathcal{P}(\Omega) \rightarrow \mathcal{L}$ и необходимости $N(\cdot) : \mathcal{P}(\Omega) \rightarrow \tilde{\mathcal{L}}$, принимающей значения в дуально-изоморфной $\mathcal{L}$ шкале ²⁾ $\tilde{\mathcal{L}} = ([0, 1], \leq, +, \bullet)$, определив вместе с (0.4) упорядоченность

$$0 = n_1 \leq n_2 \leq \dots \quad (0.10)$$

значений $n_i \stackrel{\text{def}}{=} N(\Omega \setminus \{\omega_i\})$, $i = 1, 2, \dots$, и соответственно $N(A) \stackrel{\text{def}}{=} \bigoplus_{i:\omega_i \in \Omega \setminus A} n_i \stackrel{\text{def}}{=} \inf_{i:\omega_i \in \Omega \setminus A} N(\Omega \setminus \{\omega_i\})$ — как меру, характеризующую событие A как противоположное $\Omega \setminus A$, $A \in \mathcal{P}(\Omega)$. Класс всех необходимостей, распределенных согласно (0.10), обозначим $\mathbb{N}$.

Конкретную упорядоченность в (0.10) зададим двоичным числом $\tilde{e} = 0.\tilde{e}_1\tilde{e}_2\dots \in (0, 1)$ так, чтобы $n_i < n_{i+1}$, если $\tilde{e}_i = 1$, $n_i = n_{i+1}$, если $\tilde{e}_i = 0$, $i = 1, 2, \dots$, $\mathbb{N}_{(\tilde{e})}$ обозначим класс необходимостей, упорядоченность распределений которых определена числом $\tilde{e}$. При этом аналогично разбиению (0.5)

$$\mathbb{N} = \bigcup_{\tilde{e} \in (0,1)} \mathbb{N}_{(\tilde{e})}$$

Как и для возможности, конкретная упорядоченность в (0.10) упорядочивает значения необходимостей всех событий: $N(A) = n_{i_1} \leq n_{j_1} = N(B)$, если $\Omega \setminus A = \{\omega_{i_1}, \omega_{i_2}, \dots\}$, $i_1 \leq i_2 \leq \dots$, $\Omega \setminus B = \{\omega_{j_1}, \omega_{j_2}, \dots\}$, $j_1 \leq j_2 \leq \dots$, и $i_1 \leq j_1$,

¹⁾ Заметим, что шкала значений вероятности — четверка $\mathcal{L} = ([0, 1], \leq, +, \times)$, где сложение $+$: $[0, 1]^2 \rightarrow [0, 1]$ и умножение $\times$: $[0, 1]^2 \rightarrow [0, 1]$ суть $a + b \stackrel{\text{def}}{=} a + b - ab$, $a \times b \stackrel{\text{def}}{=} ab$, поэтому в случае независимости A и B как и для возможности $\Pr(A \cup B) = \Pr(A) + \Pr(B) \stackrel{\text{def}}{=} \Pr(A) + \Pr(B) - \Pr(A)\Pr(B)$, $\Pr(A \cap B) = \Pr(A) \times \Pr(B) \stackrel{\text{def}}{=} \Pr(A)\Pr(B)$. Иначе говоря, в случае независимости A и B как для возможности, так и для вероятности операциям $\bigcup$ и $\bigcap$ над событиями соответствуют операции сложения и умножения в соответствующих шкалах [13].

²⁾ В $\tilde{\mathcal{L}}$: $a \tilde{\leq} b \Leftrightarrow a \geq b$, нейтральные элементы $\tilde{0} = 1$, $\tilde{1} = 0$, $\tilde{0} \tilde{\leq} \tilde{1}$, $a + b \stackrel{\text{def}}{=} \min(a, b)$, $a \bullet b \stackrel{\text{def}}{=} \max(a, b)$, $a, b \in [0, 1]$. Если $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1]$ — непрерывная строго монотонно убывающая функция, $\vartheta(0) = 1$, $\vartheta(1) = 0$, то $\tilde{0} = \vartheta(0)$, $\tilde{1} = \vartheta(1)$, $\vartheta(a + b) = \vartheta(a) + \vartheta(b) = \vartheta(a) \bullet \vartheta(b)$, $\vartheta(a \bullet b) = \vartheta(a) \tilde{\bullet} \vartheta(b) = \vartheta(a) + \vartheta(b)$, $a, b \in [0, 1]$, подробнее см. § 3.1.4.

причем $N(A) < N(B)$, если $\max_{i_1 \leq k < j_1} \tilde{e}_k = 1$, $N(A) = N(B)$, если $\max_{i_1 \leq k < j_1} \tilde{e}_k = 0$ или $i_1 = j_1$. Кроме этого $N(\Omega) = \inf_{i: \omega_i \in \Omega} n_i \stackrel{\text{def}}{=} 1$, $N(\emptyset) = \inf_{i: \omega_i \in \Omega} n_i = 0$, и для любых $A, B \in \mathcal{P}(\Omega)$ $A \subset B \Rightarrow N(A) \leq N(B)$, $N(A \cap B) = N(A) \tilde{+} N(B) \stackrel{\text{def}}{=} \min(N(A), N(B))$, $N(A \cup B) \tilde{\geq} N(A) \tilde{\bullet} N(B) \stackrel{\text{def}}{=} \max(N(A), N(B))$, а если A и B N -независимы, то $N(A \cup B) = N(A) \tilde{\bullet} N(B)$.

Таким образом, в общем случае нечеткая модель объекта определяется как пространство $(\Omega, \mathcal{P}(\Omega), P, N)$ с двумя мерами, связью между которыми определяется свойствами объекта. Например, в нечеткой модели стохастического объекта меры P и N дуально согласованы, $P < \approx > N$, [13], т.е. $\exists \vartheta(\cdot) \in \Theta \forall A \in \mathcal{P}(\Omega)$

$$\begin{aligned} N(A) &= \vartheta(P(\Omega \setminus A)), \\ n_i &= N(\Omega \setminus \{\omega_i\}) = \vartheta(P(\{\omega_i\})) = \vartheta(p_i), \quad i = 1, 2, \dots, \end{aligned} \quad (0.11)$$

где Θ — класс функций $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1]$, строго монотонных и удовлетворяющих условиям $\vartheta(0) = 1$, $\vartheta(1) = 0$. Если преобразование $\vartheta(\cdot)$ понимать как нечеткое отрицание, то равенство $N(A) = \vartheta(P(\Omega \setminus A))$ означает, что «необходимость A » равна «невозможности $\Omega \setminus A$ », в частности, событие A необходимо, $N(A) = 1$, если и только если событие $\Omega \setminus A$ невозможно.

В [13] подобные соображения определили математические основы теории возможностей как теории мер и интегралов со значениями в шкалах $\mathcal{L}$ и $\tilde{\mathcal{L}}$.

6. Принцип относительности. Принципиальное отличие рассматриваемой теории возможностей от теории вероятностей ¹⁾ обусловлено группой Γ изотонных автоморфизмов шкал $\mathcal{L} = ([0, 1], \leq, +, \bullet)$ и $\tilde{\mathcal{L}} = ([0, 1], \tilde{\leq}, \tilde{+}, \tilde{\bullet})$ значений возможности и необходимости. В то время как вероятностные модели формулируются в единой шкале значений вероятности, нечеткие модели могут формулироваться в различных (изоморфных!) шкалах ²⁾, выбираемых исследователями сообразно их предпочтениям. При этом формулируемые в некоторых шкалах $\mathcal{L}'$, $\tilde{\mathcal{L}}'$ и $\mathcal{L}''$, $\tilde{\mathcal{L}}''$ модели, доводы, заключения и т.п. считаются эквивалентными, если существуют шкалы $\mathcal{L} = \gamma' \mathcal{L}' = \gamma'' \mathcal{L}''$ и $\tilde{\mathcal{L}} = \tilde{\gamma}' \tilde{\mathcal{L}}' = \tilde{\gamma}'' \tilde{\mathcal{L}}''$, $\gamma', \gamma'', \tilde{\gamma}', \tilde{\gamma}'' \in \Gamma$, в которых их формулировки совпадают, а содержательно истолкованы могут быть только те из них, формулировки которых не зависят от выбора шкал (т.е. одинаковы для всех исследователей). Например, возможности P_1 и P_2

¹⁾ Этим отличается рассматриваемая теория и от теории возможностей, предложенной Л. Заде [39] и другими авторами [5, 8], [26, 37].

²⁾ Преобразование $\gamma \in \Gamma$ ($\tilde{\gamma} \in \Gamma$) определяет переход $\mathcal{L} \rightarrow \gamma \mathcal{L}$ ($\tilde{\mathcal{L}} \rightarrow \tilde{\gamma} \tilde{\mathcal{L}}$) к шкале $\gamma \mathcal{L}$ ($\tilde{\gamma} \tilde{\mathcal{L}}$), изоморфной $\mathcal{L}$ ($\tilde{\mathcal{L}}$).

со значениями в шкалах $\mathcal{L}_1$ и соответственно в $\mathcal{L}_2$ эквивалентны, если можно указать $\gamma_1 \in \Gamma$, такое, что для любого $A \in \mathcal{A}$ $\gamma_1(P_1(A)) = P_2(A)$. В противном случае P_1 и P_2 неэквивалентны, как, например, в случае, когда для некоторого $A \in \mathcal{A}$ $P_1(A) = 1$, а $P_2(A) < 1$, или $P_1(A) > 0$, а $P_2(A) = 0$, ибо равенства возможности нулю или единице остаются таковыми в любой шкале. Точно также неравенства $P(A) \neq P(B)$ или $1 > P(A) > P(B) > 0$ в шкале $\mathcal{L}$ сохраняются в любой шкале $\gamma\mathcal{L}$, $\gamma \in \Gamma$.

В то же время ответы на такие, например, вопросы, как «чему равно (отличное от 0 и 1) значение возможности того или иного события», «во сколько раз или насколько возможность одного события больше, чем возможность другого» и т. п., не могут быть содержательно истолкованы.

Принцип относительности в теории возможностей, аналогичный принципу относительности в физике, определил содержательную интерпретацию возможности, математические методы и алгоритмы ее эмпирического построения, математический формализм теории и области её применений [13].

7. Максимальная согласованность возможности с вероятностью. Стохастическая модель возможности. Согласно упорядоченностям (0.3), (0.4) и формулам (0.9), (0.9*)

- $\forall A_1 \in \mathcal{P}(\Omega)$, если $\omega_1 \in A_1$, то $P(A_1) = p_1 = 1$, $\Pr(A_1) \in \Delta_1 \stackrel{\text{def}}{=} [pr_1, 1]$, где Δ_1 — минимальный (по включению) интервал, содержащий $\Pr(A_1)$;
- $\forall A_2 \in \mathcal{P}(\Omega)$, если $\omega_1 \notin A_2$, $\omega_2 \in A_2$, то $P(A_2) = p_2$, $\Pr(A_2) \in \Delta_2 = [pr_2, 1 - pr_1]$, где Δ_2 — минимальный интервал, содержащий $\Pr(A_2)$;
- ...
- $\forall A_i \in \mathcal{P}(\Omega)$, если $\omega_1 \notin A_i, \dots, \omega_{i-1} \notin A_i, \omega_i \in A_i$, то $P(A_i) = p_i$, $\Pr(A_i) \in \Delta_i = [pr_i, 1 - pr_1 - \dots - pr_{i-1}]$, где Δ_i — минимальный интервал, содержащий $\Pr(A_i)$;
- ...

Следовательно, если потребовать, чтобы $\forall A \in \mathcal{P}(\Omega)$ $P(A) = \tilde{\gamma}(\Pr(A))$, где $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$ — монотонно неубывающая функция, определяющая "согласованность" P с $\Pr$, то $\tilde{\gamma}(a) = p_i$, $a \in \Delta_i$, $i = 1, 2, \dots$, и в случае "максимальной согласованности" P с $\Pr$ естественно считать, что $p_i > p_{i+1}$, если $\Delta_i \cap \Delta_{i+1} = \emptyset$, что эквивалентно условию $f_i \stackrel{\text{def}}{=} pr_1 + \dots + pr_{i-1} + 2pr_i > 1$, и $p_i = p_{i+1}$, если $\Delta_i \cap \Delta_{i+1} \neq \emptyset$, что эквивалентно условию $f_i \leq 1$, $i = 1, 2, \dots$. Поэтому каждому классу $\mathbb{P}_{(e)}$ в (0.5) согласно условиям

$$\begin{aligned} e_i = 1 &\iff p_i > p_{i+1} \iff pr_1 + \dots + pr_{i-1} + 2pr_i > 1, \\ e_i = 0 &\iff p_i = p_{i+1} \iff pr_1 + \dots + pr_{i-1} + 2pr_i \leq 1, \end{aligned} \quad (0.12)$$

где $\iff$ означает «если и только если», взаимно однозначно сопоставлен класс вероятностей $\mathbb{P}_{(e)}$, $e \in (0, 1)$, а разбиению (0.5)

класса $\mathbb{P}$ — разбиение

$$\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}, \quad \mathbb{P}_{(e)} \cap \mathbb{P}_{(e')} = \emptyset, \quad e \neq e', \quad e, e' \in (0,1), \quad (0.13)$$

класса $\mathbb{P}$ всех вероятностей, распределенных согласно условиям (0.3), причем так, что для любых $P \in \mathbb{P}_{(e)}$ и $\text{Pr} \in \mathbb{P}_{(e)}$ можно указать монотонно неубывающую непрерывную на $(0,1]$ функцию $\tilde{\gamma}_e(\cdot) : [0,1] \rightarrow [0,1]$, $\tilde{\gamma}_e(\cdot) \in \tilde{\Gamma}(\text{Pr})$, такую, что для любого $A \subset \Omega$

$$P(A) = \sup_{i: \omega_i \in A} p_i = \tilde{\gamma}_e\left(\sum_{i: \omega_i \in A} p_i\right) = \tilde{\gamma}_e(\text{Pr}(A)), \quad e \in (0,1). \quad (0.14)$$

Класс $\tilde{\Gamma}(\text{Pr})$ всех таких функций $\tilde{\gamma}_e(\cdot)$ определяется вероятностью $\text{Pr} \in \mathbb{P}_{(e)}$, $e \in (0,1)$. Любая возможность $P \in \mathbb{P}_{(e)}$ называется *максимально согласованной с любой вероятностью* $\text{Pr} \in \mathbb{P}_{(e)}$, *факт максимальной согласованности* P с Pr выражает символ $\text{Pr} \approx_{>} P$, означающий, что среди неравенств $\tilde{\gamma}_e(p_i) \geq \tilde{\gamma}_e(p_{i+1})$, $i = 1, 2, \dots$, максимальное число строгих $>$, [13]. Если $\text{Pr} \approx_{>} P$, то каждое событие $A \in \mathcal{P}(\Omega)$ в $(\Omega, \mathcal{P}(\Omega), \text{Pr})$ можно интерпретировать как событие в $(\Omega, \mathcal{P}(\Omega), P)$, а его возможность будет определена его вероятностью равенством (0.14); в этом смысле класс $\mathbb{P}_{(e)}$ называется *стохастической моделью* класса $\mathbb{P}_{(e)}$ и любой возможности $P \in \mathbb{P}_{(e)}$, любая вероятность $\text{Pr} \in \mathbb{P}_{(e)}$ определяет *вероятностную модель* любой возможности $P \in \mathbb{P}_{(e)}$, а последняя называется *Pr-измеримой*, $e \in (0,1)$, подробнее см. § 3.2.

8. Эмпирическое построение и эмпирическая интерпретация возможности. Рассмотрим вначале эмпирическую интерпретацию возможности. Согласно равенствам (0.14) и усиленному 3.Б.Ч. (0.1) для любых событий $A, B \in \mathcal{P}(\Omega)$, таких, что $\text{Pr}(A) > \text{Pr}(B)$, для любой функции $\tilde{\gamma}_e(\cdot) \in \tilde{\Gamma}(\text{Pr})$ и любого $\varepsilon > 0$ можно указать число $N = N(\varepsilon, A, B, \tilde{\gamma}_e(\cdot))$, такое, что для всех $n > N$ $\tilde{\gamma}_e(\text{Pr}(A)) - \tilde{\gamma}_e(\text{Pr}(B)) - \varepsilon < \tilde{\gamma}_e(\nu^{(n)}(A)) - \tilde{\gamma}_e(\nu^{(n)}(B)) < \tilde{\gamma}_e(\text{Pr}(A)) - \tilde{\gamma}_e(\text{Pr}(B)) + \varepsilon$. Поэтому, если $P(A) = \tilde{\gamma}_e(\text{Pr}(A)) > \tilde{\gamma}_e(\text{Pr}(B)) = P(B)$, то, выбрав $\varepsilon \in (0, \tilde{\gamma}_e(\text{Pr}(A)) - \tilde{\gamma}_e(\text{Pr}(B)))$, найдем, что для всех $n > N$

$$P(A) > P(B) \Rightarrow \tilde{\gamma}_e(\nu^{(n)}(A)) \stackrel{\text{п.н.}}{>} \tilde{\gamma}_e(\nu^{(n)}(B)) \Rightarrow \nu^{(n)}(A) \stackrel{\text{п.н.}}{>} \nu^{(n)}(B).$$

Этот факт определяет эмпирическую событийно-частотную интерпретацию Pr-измеримой возможности, а именно, *в достаточно длинной последовательности взаимно независимых испытаний упорядоченность возможностей любых событий п.н. точно прогнозирует такую же упорядоченность значений их частот*.

Взаимно однозначное соответствие между классами $\mathbb{P}_{(e)}$ и $\mathbb{P}_{(e)}$, $e \in (0,1)$, в (0.5) и в (0.13) по существу решает *проблему эмпирического построения возможностной модели стохастического объекта* [13]. Дело в том, что задача эмпирического построения воз-

возможности, эквивалентная задаче выбора одного из классов *взаимно эквивалентных* возможностей $\mathbb{P}_{(e)}$, $e \in (0, 1)$, в (0.5), в силу соответствия между классами $\mathbb{P}_{(e)}$ и $\mathbb{P}_{(e)}$, $e \in (0, 1)$, сводится к задаче статистической идентификации, в которой в предположении, что результаты наблюдений за стохастическим объектом контролируются, вообще говоря, изменяющейся от испытания к испытанию вероятностью $\text{Pr} \in \mathbb{P}_{(e_0)}$, требуется на основе результатов испытаний принять решение о значении $e_0 \in (0, 1)$, *фиксированном условиями наблюдений за объектом*.

Подчеркнем, что в то время как при эмпирическом оценивании вероятности, контролирующей результаты наблюдений, необходимо, чтобы *последняя была зафиксирована условиями наблюдений, при восстановлении возможности условия наблюдений должны фиксировать один из классов $\mathbb{P}_{(e)}$ в (0.13), в пределах которого вероятность, контролирующая результаты испытаний, может произвольно изменяться от испытания к испытанию*.

Характерно, что при этом класс $\mathcal{P}_{(e_0)} = \{(\Omega, \mathcal{P}(\Omega), \text{Pr}), \text{Pr} \in \mathbb{P}_{(e_0)}\}$, как (эмпирически восстановленная) модель стохастического объекта, не позволяет охарактеризовать его стандартными средствами теории вероятностей.

Проиллюстрируем этот важный результат в случае $\Omega = \{\omega_1, \omega_2, \omega_3\}$, представив P и Pr точками $\mathbf{p} = (p_1, p_2, p_3)$ и $\mathbf{pr} = (pr_1, pr_2, pr_3)$ в $\mathcal{R}^3$. Соответственно представленные как множества в $\mathcal{R}^3$ классы $\mathbb{P}$ и $\mathbb{P}_{(e)}$ в (0.5) обозначим

$$\mathbf{p} = \{(p_1, p_2, p_3), 1 = p_1 \geq p_2 \geq p_3 \geq 0\} = \bigcup_{e \in (0, 1)} \mathbf{p}_{(e)},$$

$$\mathbf{p}_{(e)} = \{(p_1, p_2, p_3), \begin{matrix} e & = & 0 \\ 1 & = & p_1 \end{matrix} \begin{matrix} . & e_1 \\ \geq & p_2 \end{matrix} \begin{matrix} e_2 \\ \geq & p_3 \end{matrix} \begin{matrix} e_3 \\ \geq & 0 \end{matrix} \},$$

где $\overset{e_i}{\geq}$ означает $>$, если $e_i = 1$, и, соответственно, $=$, если $e_i = 0$, $i = 1, 2, 3$, см. рис. 3.1. Классы $\mathbb{P}$ и $\mathbb{P}_{(e)}$ в (0.13) представим множествами

$$\mathbf{pr} = \{(pr_1, pr_2, pr_3), 1 \geq pr_1 \geq pr_2 \geq pr_3 \geq 0, pr_1 + pr_2 + pr_3 = 1\} = \bigcup_{e \in (0, 1)} \mathbf{pr}_{(e)},$$

где согласно условиям в (0.12),

$$\mathbf{pr}_{(e)} = \{(pr_1, pr_2, pr_3), \begin{matrix} e_1 \\ 2pr_1 \end{matrix} \geq 1, \begin{matrix} e_2 \\ pr_1 + 2pr_2 \end{matrix} \geq 1, \begin{matrix} e_3 \\ pr_1 + pr_2 + 2pr_3 \end{matrix} \geq 1\} \cap \mathbf{pr},$$

где $\overset{e_i}{\geq} 1$ означает > 1 , если $e_i = 1$, и, соответственно, ≤ 1 , если $e_i = 0$, $i = 1, 2$, а $\overset{e_3}{\geq} 1$ означает > 1 , если $e_3 = 1$, и $= 1$, если $e_3 = 0$, см. рис. 3.2, 3.1, где множество $\mathbf{p}_* = \{(p_1, p_2, p_3), 0 \leq p_i \leq 1, i =$

$= 1, 2, 3, \max_{1 \leq i \leq 3} p_i = 1\} \subset \mathcal{R}^3$ представляет класс $\mathbb{P}_*$ всех возможностей $P(\cdot) : \mathcal{P}(\{\omega_1, \omega_2, \omega_3\}) \rightarrow [0, 1]$, и треугольник $\mathbf{pr}_* = \{(pr_1, pr_2, pr_3), 0 \leq pr_i \leq 1, i = 1, 2, 3, pr_1 + pr_2 + pr_3 = 1\}$ представляет класс $\mathbb{Pr}_*$ всех вероятностей $Pr(\cdot) : \mathcal{P}(\{\omega_1, \omega_2, \omega_3\}) \rightarrow [0, 1]$, см. рис. 3.2.

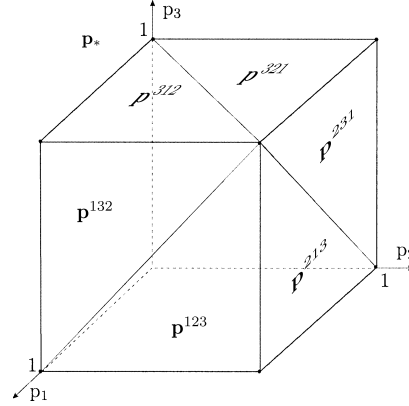


Рис. 3.1. «Треугольники распределений возможностей»

$\mathbf{p}^{\pi_s(1)\pi_s(2)\pi_s(3)} \stackrel{\text{def}}{=} \{(p_1, p_2, p_3), 1 = p_{\pi_s(1)} \geq p_{\pi_s(2)} \geq p_{\pi_s(3)} \geq 0\}, s = 1, \dots, 6;$ тройка ijk обозначает упорядоченность $1 = p_i \geq p_j \geq p_k$, определяющую «треугольник возможностей» $\mathbf{p}^{ijk}$, $123 = \pi_1(1)\pi_1(2)\pi_1(3)$, $213 = \pi_2(1)\pi_2(2)\pi_2(3)$, $231 = \pi_3(1)\pi_3(2)\pi_3(3)$, ..., $132 = \pi_6(1)\pi_6(2)\pi_6(3)$, где $\pi_s(\cdot) : \{1, 2, 3\} \rightarrow \{\pi_s(1), \pi_s(2), \pi_s(3)\}$ — перестановка, $s = 1, \dots, 6 = 3!$; $\mathbf{p}^{123} \stackrel{\text{def}}{=} \mathbf{p} \bigcup_{\{e\}} \mathbf{p}_{(e)}$, где $\{e\} = \{0.111, 0.110, 0.101, 0.100, 0.011, 0.010, 0.001\}$;

$$\mathbf{p}_* = \bigcup_{1 \leq s \leq 6} \mathbf{p}^{\pi_s(1)\pi_s(2)\pi_s(3)}.$$

Рассмотрим классы $\mathbf{p}_{(e)}$ и $\mathbf{pr}_{(e)}$, $e \in \{e\} = \{0.111, 0.110, 0.101, 0.100, 0.011, 0.010, 0.001\}$ на рис. 3.3 а, б, например, для $e = 0.111$. Каждая точка $\mathbf{pr} = (pr_1, pr_2, pr_3) \in \mathbf{pr}_{(0.111)}$ удовлетворяет условиям $2pr_1 > 1$, $pr_1 + 2pr_2 > 1$, $pr_1 + pr_2 + 2pr_3 > 1$, $pr_1 + pr_2 + pr_3 = 1$ в (0.12) и определяет класс $\Gamma(\mathbf{Pr})$ функций $\tilde{\gamma}_{0.111}(\cdot) : [0, 1] \rightarrow [0, 1]$, монотонно неубывающих и непрерывных на $(0, 1]$, таких, что для любого $A \subset \Omega = \{\omega_1, \omega_2, \omega_3\}$ выполнено равенство (0.14), в котором $e = 0.111$, $\mathbf{P}$ — возможность, определенная *любой точкой* $\mathbf{p} = (p_1, p_2, p_3) \in \mathbf{p}_{(0.111)} = \mathbf{p}^{123}$ и $\mathbf{Pr}$ — вероятность, определенная *выбранной точкой* $\mathbf{pr} = (pr_1, pr_2, pr_3) \in \mathbf{pr}_{(0.111)} = \mathbf{pr}^{123}$.

Пусть $(\Omega, \mathcal{P}(\Omega), \mathbf{Pr}_{(1)}) \times (\Omega, \mathcal{P}(\Omega), \mathbf{Pr}_{(2)}) \times \dots$ — последовательность взаимно независимых испытаний, в которой $\Omega = \{\omega_1, \omega_2, \omega_3\}$, вероятности $\mathbf{Pr}_{(1)}, \mathbf{Pr}_{(2)}, \dots$ неизвестны, но известно, что $\mathbf{Pr}_{(i)} \in \mathbb{Pr}_{(0.111)}$, $i = 1, 2, \dots$. Для простоты будем считать, что среди вероятностей $\mathbf{Pr}_{(1)}, \mathbf{Pr}_{(2)}, \dots$ конечное число различных, обозначим их $\mathbf{Pr}^1, \dots, \mathbf{Pr}^k$.

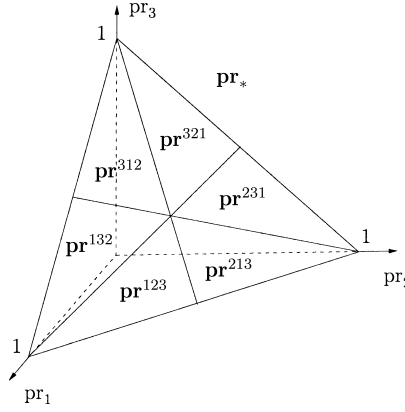


Рис. 3.2. «Треугольник распределений вероятностей» $\mathbf{pr}_* = \{(\mathbf{pr}_1, \mathbf{pr}_2, \mathbf{pr}_3), 0 \leq \mathbf{pr}_i \leq 1, i = 1, 2, 3, \mathbf{pr}_1 + \mathbf{pr}_2 + \mathbf{pr}_3 = 1\} = \bigcup_{1 \leq s \leq 6} \mathbf{pr}^{\pi_s(1)\pi_s(2)\pi_s(3)}$, где

$\mathbf{pr}^{\pi_s(1)\pi_s(2)\pi_s(3)} \stackrel{\text{def}}{=} \{(\mathbf{pr}_1, \mathbf{pr}_2, \mathbf{pr}_3), 1 \geq \mathbf{pr}_{\pi_s(1)} \geq \mathbf{pr}_{\pi_s(2)} \geq \mathbf{pr}_{\pi_s(3)} \geq 0, \mathbf{pr}_{\pi_s(1)} + \mathbf{pr}_{\pi_s(2)} + \mathbf{pr}_{\pi_s(3)} = 1\}$, $s = 1, \dots, 6$, тройка ijk обозначает упорядоченность $\mathbf{pr}_i \geq \mathbf{pr}_j \geq \mathbf{pr}_k$, определяющую «треугольник вероятностей» $\mathbf{pr}^{ijk}$, $123 = \pi_1(1)\pi_1(2)\pi_1(3)$, $213 = \pi_2(1)\pi_2(2)\pi_2(3)$, $231 = \pi_3(1)\pi_3(2)\pi_3(3)$, $\dots$, $132 = \pi_6(1)\pi_6(2)\pi_6(3)$, где $\pi_s(\cdot) : \{1, 2, 3\} \rightarrow \{\pi_s(1), \pi_s(2), \pi_s(3)\}$ — перестановка, $s = 1, \dots, 6 = 3!$; $\mathbf{pr}^{123} \stackrel{\text{def}}{=} \mathbf{pr} = \bigcup_{\{e\}} \mathbf{pr}_{(e)}$, где $\{e\} = \{0.111, 0.110, 0.101, 0.100, 0.011, 0.010, 0.001\}$.

Покажем, воспользовавшись рис. 3.3 а), б), что в этом случае возможностная модель $(\Omega, \mathcal{P}(\Omega), P)$, $P \in \mathbb{P}_{(0.111)}$, каждого испытания *может быть построена*, причем — *безошибочно и на основе п. н. конечного числа значений частот элементарных событий* $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$, в то время как построение вероятностной модели испытаний в таких условиях без дополнительных предположений невозможно.

Действительно, поскольку *любое множество* $\mathbf{pr}_{(e)} \subset \mathbf{pr}$, $e \in \{e\}$, см. рис. 3.3 б), — *выпуклое* и $\mathbf{pr}^s \in \mathbf{pr}_{(0.111)}$, $s = 1, \dots, k$, то при любом $n = 1, 2, \dots$

$$\mathbf{pr}^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n \mathbf{pr}_{(j)} = \sum_{s=1}^k \frac{n_s}{n} \mathbf{pr}^s \in \mathbf{pr}_{(0.111)}, \quad (0.15)$$

где $\mathbf{pr}_{(j)} = (\mathbf{pr}_{(j)1}, \mathbf{pr}_{(j)2}, \mathbf{pr}_{(j)3})$, $\mathbf{pr}_{(j)i} = \text{Pr}_{(j)}(\{\omega_i\})$, $i = 1, 2, 3$, $j = 1, \dots, n$, $\mathbf{pr}^s = (\mathbf{pr}_1^s, \mathbf{pr}_2^s, \mathbf{pr}_3^s)$, $\mathbf{pr}_i^s = \text{Pr}^s(\{\omega_i\})$, $i = 1, 2, 3$, и n^s/n — частота, с которой точка $\mathbf{pr}^s \in \mathbf{pr}_{(0.111)}$ встречается среди $\mathbf{pr}_{(1)}, \dots, \mathbf{pr}_{(n)}$, $s = 1, \dots, k$. А так как $\mathbf{pr}_{(0.111)}$ — *открытое в* $\mathbf{pr}$ *множество*, то выпуклая оболочка со $(\mathbf{pr}^1, \dots, \mathbf{pr}^k)$ точек $\mathbf{pr}^1, \dots, \mathbf{pr}^k$ содержится в $\mathbf{pr}_{(0.111)}$ вместе с некоторой ее окрестностью, и, следовательно, *каждая точка* $\mathbf{pr}^{(n)}$ в (0.15), $n = 1, 2, \dots$, *содержится в*

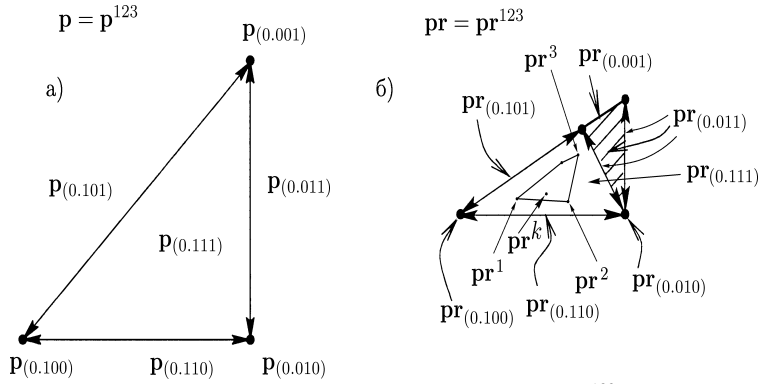


Рис. 3.3. а) «Треугольник распределений возможностей» $\mathbf{p}^{123} = \mathbf{p}$ и семь классов $\mathbf{p}_{(0.001)}, \dots, \mathbf{p}_{(0.111)}$ эквивалентных возможностей, образующих разбиение $\mathbf{p} = \mathbf{p}_{(0.001)} \cup \dots \cup \mathbf{p}_{(0.111)}$, см. (0.5); б) «Треугольник распределений вероятностей» $\mathbf{pr}^{123} = \mathbf{pr}$ и семь его подмножеств — прообразов $\mathbf{p}_{(0.001)}, \dots, \mathbf{p}_{(0.111)}$, образующих разбиение $\mathbf{pr} = \mathbf{pr}_{(0.001)} \cup \dots \cup \mathbf{pr}_{(0.111)}$, см. (0.13). Точки $\mathbf{pr}^1, \dots, \mathbf{pr}^k$ представляют вероятности $\text{Pr}^1, \dots, \text{Pr}^k$.

$\mathbf{pr}_{(0.111)}$ вместе с некоторой, не зависящей от n , окрестностью, ибо $\mathbf{pr}^{(n)} \in \text{co}(\mathbf{pr}^1, \dots, \mathbf{pr}^k)$, $n = 1, 2, \dots$. Наконец, поскольку согласно 3. Б.Ч. (0.2)

$$\nu^{(n)} - \mathbf{pr}^{(n)} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} 0, \quad (0.16)$$

где $\nu^{(n)} = (\nu_1^{(n)}, \nu_2^{(n)}, \nu_3^{(n)})$, $\nu_i^{(n)}$ — частота, с которой ω_i встречается в последовательности n испытаний, $i = 1, 2, 3$, то, начиная с некоторого $n = \bar{n}$, случайная точка $\nu^{(n)} \in \mathbf{pr}_{(0.111)}$ попадет и п. н. останется в упомянутой окрестности (произвольно) блуждающей в $\text{co}(\mathbf{pr}^1, \dots, \mathbf{pr}^k) \subset \mathbf{pr}_{(0.111)}$ точки $\mathbf{pr}^{(n)}$. Иными словами

$$\exists \bar{n} \quad \forall n \geq \bar{n} \quad \nu^{(n)} \overset{\text{п.н.}}{\in} \mathbf{pr}_{(0.111)},$$

и тем самым определен класс $\mathbb{P}_{(0.111)}$ вероятностей, контролирующих исходы испытаний, и п. н. безошибочно восстановлена вероятностная модель $(\Omega, \mathcal{P}(\Omega), \mathbf{P})$, $\mathbf{P} \in \mathbb{P}_{(0.111)}$, каждого испытания.

Заметим теперь, что в общем случае вероятность $\sum_{s=1}^k (n_s/n) \text{Pr}^s(A)$ с увеличением n произвольно «блуждает по отрезку $[\min_{1 \leq s \leq k} \text{Pr}^s(A), \dots]$ »

¹⁾ При каждом фиксированном $n = 1, 2, \dots$ $\text{Pr}^{(n)}(A) \stackrel{\text{def}}{=} \sum_{s=1}^k (n_s/n) \text{Pr}^s(A)$, $A \in \mathcal{P}(\Omega)$, — вероятность, $\sum_i \text{Pr}^{(n)}(\{\omega_i\}) = \sum_{s=1}^k (n_s/n) \sum_i \text{Pr}^s(\{\omega_i\}) = 1$.

$\max_{1 \leq s \leq k} \Pr^s(A)]\gg, A \in \mathcal{P}(\Omega)$, и за ней согласно З.Б.Ч. (0.2) все более точно следует частота $\nu^{(n)}(A)$. В этом случае знание вероятностей $\Pr^1(A), \dots, \Pr^k(A)$ не позволит оценить частоту $\nu^{(n)}(A)$, а наблюдение за последней не позволит восстановить вероятностную модель испытаний.

Вместе с тем, если существует $e \in (0, 1)$, такое, что вероятности $\Pr^1, \dots, \Pr^k$ содержатся в классе $\mathbb{P}_{(e)}$ (т.е. $\Pr^s \approx > P, s = 1, \dots, k$; возможность P называется $\Pr^1, \dots, \Pr^k$ -измеримой), то при условии регулярности $\Pr^1, \dots, \Pr^k$ [13], класс $\mathbb{P}_{(e)}$ эквивалентных возможностей можно восстановить безошибочно на основе п.н. конечного числа испытаний и любой возможности P из $\mathbb{P}_{(e)}$ дать событийно-частотную интерпретацию, согласно которой, как и в случае неменяющейся вероятности, если $P(A) > P(B)$, то существует число $N(A, B)$, такое, что $\nu^{(n)}(A) \xrightarrow{\text{п.н.}} \nu^{(n)}(B)$ для всех $n > N(A, B)$.

Рассмотрим, например, событийно-частотную интерпретацию $\Pr^1, \dots, \Pr^k$ -измеримой возможности. Заметим, что

$$\nu^{(n)}(A) = \sum_{s=1}^k (n_s/n) \nu^{(n_s)}(A), \quad (0.17)$$

где $\nu^{(n_s)}(A)$ — частота события A в подпоследовательности последовательности n испытаний, в которой исход испытания контролируется вероятностью $\Pr^s$, причем согласно усиленному З.Б.Ч. (0.1) при $n \rightarrow \infty$

$$\nu^{(n_s)}(A) \xrightarrow{\text{п.н.}} \Pr^s(A), \quad s = 1, \dots, k. \quad (0.18)$$

Условия $\Pr^s \approx > P, s = 1, \dots, k$, означают, что как в (0.14) существуют функции $\tilde{\gamma}_e^s(\cdot) \in \tilde{\Gamma}(\Pr^s)$, $s = 1, \dots, k$, такие, что для любого $A \in \mathcal{P}(\Omega)$ $P(A) = \tilde{\gamma}_e^1(\Pr^1(A)) = \dots = \tilde{\gamma}_e^k(\Pr^k(A))$. Поэтому согласно сходимостям в (0.18) при $n \rightarrow \infty$ $\tilde{\gamma}_e^s(\nu^{(n_s)}(A)) \xrightarrow{\text{п.н.}} \tilde{\gamma}_e^s(\Pr^s(A)) = P(A)$, $s = 1, \dots, k$, и, следовательно, если $P(A) > P(B)$, то $\Pr^s(A) > \Pr^s(B)$, $s = 1, \dots, k$, а это означает, что существует число $N(A, B)$, такое, что $\nu^{(n_s)}(A) \xrightarrow{\text{п.н.}} \nu^{(n_s)}(B)$ для всех $n > N(A, B)$ и $s = 1, \dots, k$. Отсюда согласно равенству (0.17) следует, что и в рассматриваемом случае изменяющейся вероятности неравенство $P(A) > P(B)$ п.н. влечет неравенство $\nu^{(n)}(A) > \nu^{(n)}(B)$ для всех $n > N(A, B)$.

Заметим однако, что если вероятности $\Pr^1, \dots, \Pr^k$ встречаются в последовательности независимых испытаний с определенной закономерностью, например, — с (неизвестными) вероятностями $\text{pr}(1), \dots, \text{pr}(k)$, то при $n \rightarrow \infty$ в (0.15) $n_s/n \xrightarrow{\text{п.н.}} \text{pr}(s)$, $s = 1, \dots, k$, и, следовательно, $\Pr^{(n)}(A) \stackrel{\text{def}}{=} \sum_{s=1}^k (n_s/n) \Pr^s(A) \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \sum_{s=1}^k \Pr^s(A) \text{pr}(s) \stackrel{\text{def}}{=} \Pr(A)$, $A \in \mathcal{P}(\Omega)$. В этом случае $\nu^{(n)}(A) \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \Pr(A)$, $A \in \mathcal{P}(\Omega)$, т.е. может быть восстановлена эмпирически вероятностная модель $(\Omega, \mathcal{P}(\Omega), \Pr)$.

3.1. Элементы теории возможностей

3.1.0. Предисловие. Хорошо известны два пути построения теории меры и интеграла, образующей математический фундамент теории вероятностей. Первый путь, известный как схема Лебега, начинается с конструкций меры, измеримого множества и измеримой функции и завершается построением интеграла на классе измеримых суммируемых функций. Другой путь, известный как схема Даниэля, начинается с конструкции элементарного интеграла, определенного на классе элементарных функций, затем последний расширяется, а интеграл продолжается на более широкий класс функций (см., например, [24]).

Математическим фундаментом теории возможностей, как и теории вероятностей, является теория меры и интеграла, за основу которой взята конструкция *линейного счетно-аддитивного интеграла* $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$, определенного на классе $\mathcal{L}(X)$ $\mathcal{A}$ -измеримых функций $f(\cdot) : X \rightarrow \mathcal{L}$ и принимающего значения в шкале ¹⁾ $\mathcal{L} = ([0, 1], \leq, +, \bullet)$. Возможность (мера возможности) $P(\cdot) : \mathcal{A} \rightarrow \mathcal{L}$ определена значениями $p(\cdot)$ на классе $\{\chi_A(\cdot), A \in \mathcal{A}\} \subset \mathcal{L}(X)$ индикаторных функций $\mathcal{A}$ -измеримых подмножеств $A \subset X$ (событий): $P(A) = p(\chi_A(\cdot))$. Значения $p(f(\cdot))$ на остальных функциях $f(\cdot) \in \mathcal{L}(X)$ определяют возможности так называемых нечетких событий.

3.1.1. Шкала значений возможности. Интеграл. Определение, свойства. Назовем *шкалой* $\mathcal{L} = ([0, 1], \leq, +, \bullet)$ (*значений возможности*) отрезок $[0, 1]$ с упорядоченностью, определенной неравенством $\leq$, операцией сложения «+», понимаемой как «max», и операцией умножения «•», понимаемой как «min», ²⁾

$$a + b \stackrel{\text{def}}{=} \max(a, b), \quad a \bullet b \stackrel{\text{def}}{=} \min(a, b), \quad a, b \in [0, 1].$$

Так определенные операции коммутативны: $a + b = b + a$, $a \bullet b = b \bullet a$, ассоциативны: $(a + b) + c = a + (b + c)$, $(a \bullet b) \bullet c = a \bullet (b \bullet c)$, взаимно дистрибутивны:

$$\begin{aligned} a \bullet (b + c) &\stackrel{\text{def}}{=} \min(a, \max(b, c)) = \\ &= \max(\min(a, b), \min(a, c)) = (a \bullet b) + (a \bullet c); \end{aligned} \quad (1.1)$$

$$\begin{aligned} a + (b \bullet c) &\stackrel{\text{def}}{=} \max(a, \min(b, c)) = \\ &= \min(\max(a, b), \max(a, c)) = (a + b) \bullet (a + c), \end{aligned} \quad (1.2)$$

¹⁾ Вообще говоря, $f(\cdot) : X \rightarrow [0, 1]$, $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$. Запись, использованная в тексте, призвана отметить законы преобразования $f(\cdot) \in \mathcal{L}(X)$ и $p(\cdot) \in \mathcal{L}(\mathcal{L}(X))$ при преобразовании шкалы $\mathcal{L}$.

²⁾ Шкала $\mathcal{L} = ([0, 1], \leq, +, \bullet)$ — полная дистрибутивная решетка, в которой упорядоченность задана естественным отношением $\leq$, а решеточные операции суть $a \vee b \stackrel{\text{def}}{=} a + b$, $a \wedge b \stackrel{\text{def}}{=} a \bullet b$ [1].

и монотонны:

$$a \leq b \Rightarrow \begin{cases} a \bullet c \leq b \bullet c, \\ a + c \leq b + c, \end{cases} \quad a, b, c \in [0, 1]. \quad (1.3)$$

Определим нейтральные элементы шкалы $\mathbf{0}$ и $\mathbf{1}$, положив $\mathbf{0} \stackrel{\text{def}}{=} 0$, $\mathbf{1} \stackrel{\text{def}}{=} 1$, для которых $\mathbf{0} < \mathbf{1}$,

$$\begin{aligned} \mathbf{0} \bullet a &= \min(0, a) = \mathbf{0}, \quad \mathbf{0} + a = \max(0, a) = a, \\ \mathbf{1} \bullet a &= \min(1, a) = a, \quad \mathbf{1} + a = \max(a, 1) = \mathbf{1}, \quad a \in [0, 1]. \end{aligned} \quad (1.4)$$

Последовательность $\{a_n\} \subset \mathcal{L}$ назовем сходящейся, если $\varliminf_{n \rightarrow \infty} a_n \stackrel{\text{def}}{=} \sup_N \inf_{n \geq N} a_n = \inf_N \sup_{n \geq N} a_n \stackrel{\text{def}}{=} \varlimsup_{n \rightarrow \infty} a_n = a$, число a назовем ее пределом, $a = \lim_{n \rightarrow \infty} a_n$.

Обозначим $\mathcal{L}(X)$ класс функций, определенных на X , со значениями в $\mathcal{L}$, содержащий:

- 1) вместе с каждой функцией $f(\cdot)$ все функции $(a \bullet f)(\cdot)$, $a \in [0, 1]$, $(a \bullet f)(x) \stackrel{\text{def}}{=} \min(a, f(x)) = a \bullet f(x)$, $x \in X$;
- 2) вместе с каждой парой функций $f_1(\cdot)$ и $f_2(\cdot)$ их сумму $(f_1 + f_2)(\cdot)$ и произведение $(f_1 \bullet f_2)(\cdot)$,

$$\begin{aligned} (f_1 + f_2)(x) &\stackrel{\text{def}}{=} f_1(x) + f_2(x) = \max(f_1(x), f_2(x)), & x \in X, \\ (f_1 \bullet f_2)(x) &\stackrel{\text{def}}{=} f_1(x) \bullet f_2(x) = \min(f_1(x), f_2(x)), \end{aligned} \quad (1.5)$$

и, следовательно, любую их «линейную комбинацию» $((a_1 \bullet f_1) + (a_2 \bullet f_2))(\cdot)$, $a_1, a_2 \in [0, 1]$;

- 3) вместе с любой последовательностью функций $f_1(\cdot), f_2(\cdot), \dots$ функции

$$\begin{aligned} \left(\bigoplus_{n=1}^{\infty} f_n \right)(x) &\stackrel{\text{def}}{=} \bigoplus_{n=1}^{\infty} f_n(x) \stackrel{\text{def}}{=} \sup_n f_n(x), \\ \left(\bigodot_{n=1}^{\infty} f_n \right)(x) &\stackrel{\text{def}}{=} \bigodot_{n=1}^{\infty} f_n(x) \stackrel{\text{def}}{=} \inf_n f_n(x), \quad x \in X, \end{aligned} \quad (1.6)$$

и, следовательно, — ее верхний $\varlimsup_{n \rightarrow \infty} f_n(x) \stackrel{\text{def}}{=} \inf_N \sup_{n \geq N} f_n(x)$ и нижний

$\varliminf_{n \rightarrow \infty} f_n(x) \stackrel{\text{def}}{=} \sup_N \inf_{n \geq N} f_n(x)$, $x \in X$, пределы, а также — ее предел $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, $x \in X$, если последний существует;

- 4) вместе с каждой функцией $f(\cdot)$
 - функцию $\gamma \circ f(\cdot)$, $\gamma \circ f(x) \stackrel{\text{def}}{=} \gamma(f(x))$, $x \in X$, где $\gamma(\cdot) : [0, 1] \rightarrow [0, 1]$ — любая функция из класса Γ непрерывных строго моно-

тонных функций, удовлетворяющих условиям $\gamma(0) = 0$, $\gamma(1) = 1$, называемых *изотонными* ¹⁾;

— функцию $\vartheta \circ f(\cdot)$, $\vartheta \circ f(x) \stackrel{\text{def}}{=} \vartheta(f(x))$, $x \in X$, где $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1]$ — любая функция из класса Θ непрерывных, строго монотонных функций, удовлетворяющих условиям $\vartheta(0) = 1$, $\vartheta(1) = 0$. В ряде случаев удобно считать, что $\vartheta(\vartheta(x)) = x$, $x \in [0, 1]$, этим свойством обладает, например, $\vartheta(x) = (1 - x^\alpha)^{1/\alpha}$, $x \in [0, 1]$, $\alpha > 0$. Такие функции $\vartheta(\cdot)$ называются *антитонными*, а, в силу последнего их свойства, — *инволюциями*.

Определим на $\mathcal{L}(X)$ отношение упорядоченности, считая, что $f_1(\cdot) \leq f_2(\cdot)$, если $f_1(x) \leq f_2(x)$, $x \in X$ ²⁾.

Замечание 3.1.1. В условиях 2) и 3) достаточно ограничиться требованиями лишь к одной из операций $+$ или $\bullet$, например, — к $+$, так как для любых функций $f_1(\cdot)$ и $f_2(\cdot)$ из $\mathcal{L}(X)$ и любой функции $\vartheta(\cdot) \in \Theta$, $\vartheta^{-1}(\cdot) \in \Theta$, $\vartheta^{-1} \circ (\vartheta \circ f_1 + \vartheta \circ f_2)(\cdot) = (f_1 \bullet f_2)(\cdot) \in \mathcal{L}(X)$, и включение $\{f_n(\cdot)\} \subset \mathcal{L}(X)$ влечет $\vartheta^{-1} \circ \left(\bigoplus_{n=1}^{\infty} \vartheta \circ f_n \right)(\cdot) = \left(\bigoplus_{n=1}^{\infty} f_n \right)(\cdot) \in \mathcal{L}(X)$. Заметим также, что условие $\forall \vartheta(\cdot) \in \Theta \ f(\cdot) \in \mathcal{L}(X) \Rightarrow \vartheta \circ f(\cdot) \in \mathcal{L}(X)$ влечет $\forall \gamma(\cdot) \in \Gamma \ f(\cdot) \in \mathcal{L}(X) \Rightarrow \gamma \circ f(\cdot) \in \mathcal{L}(X)$, ибо любую функцию $\gamma(\cdot) \in \Gamma$ можно представить композицией $\gamma(\cdot) = \vartheta_1 \circ \vartheta_2(\cdot)$, $\vartheta_1, \vartheta_2 \in \Theta$.

Определение 3.1.1. Интеграл $p(\cdot)$ определим как *линейную счетно-аддитивную функцию* на $\mathcal{L}(X)$, принимающую значения в $\mathcal{L}$, т.е. такую, что $\forall a_1, a_2 \in \mathcal{L}$, $\forall f_1(\cdot), f_2(\cdot) \in \mathcal{L}(X)$

$$p\left(\left((a_1 \bullet f_1) + (a_2 \bullet f_2)\right)(\cdot)\right) = (a_1 \bullet p(f_1(\cdot))) + (a_2 \bullet p(f_2(\cdot))) \quad (1.7)$$

и $\forall \{f_n(\cdot)\} \subset \mathcal{L}(X)$

$$p\left(\sup_n f_n(\cdot)\right) \stackrel{\text{def}}{=} p\left(\bigoplus_{n=1}^{\infty} f_n(\cdot)\right) = \bigoplus_{n=1}^{\infty} p(f_n(\cdot)) \stackrel{\text{def}}{=} \sup_n p(f_n(\cdot)). \quad (1.8)$$

Класс всех интегралов $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$ обозначим $\mathcal{L}(\mathcal{L}(X))$.

Рассмотрим свойства интеграла $p(\cdot)$. Пусть $f_1(\cdot) \geq f_2(\cdot)$ и, следовательно, $f_1(\cdot) = \max(f_1, f_2)(\cdot)$. Поэтому согласно (1.7)

$$p(f_1(\cdot)) = p(\max(f_1, f_2)(\cdot)) = \max(p(f_1(\cdot)), p(f_2(\cdot))) \geq p(f_2(\cdot)), \quad (1.9)$$

т.е. $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$ — *монотонно неубывающая функция*.

¹⁾ Класс Γ является группой относительно композиции $\gamma \circ \gamma'(a) \stackrel{\text{def}}{=} \gamma(\gamma'(a))$, $a \in [0, 1]$, шкала $\mathcal{L}$, т.е. четверка $([0, 1], \leq, +, \bullet)$, инвариантна относительно преобразований $\gamma(\cdot) \in \Gamma$.

²⁾ Класс $\mathcal{L}(X)$, как частично упорядоченное множество функций $f(\cdot) : X \rightarrow [0, 1]$, является полной дистрибутивной решеткой относительно решеточных операций $(f_1 \vee f_2)(\cdot) \stackrel{\text{def}}{=} (f_1 + f_2)(\cdot)$, $(f_1 \wedge f_2)(\cdot) \stackrel{\text{def}}{=} (f_1 \bullet f_2)(\cdot)$ [1].

Пусть $\{f_n(\cdot)\} \subset \mathcal{L}(X)$ — монотонно неубывающая последовательность, т. е. $f_n(x) \leq f_{n+1}(x)$, $n = 1, 2, \dots$, $x \in X$. Она поточечно сходится, ее предел $f(x) = \lim_{n \rightarrow \infty} f_n(x) = \sup_n f_n(x)$, $x \in X$, содержится в $\mathcal{L}(X)$, а условия (1.8) и (1.9) фиксируют непрерывность $p(\cdot)$ относительно такой сходимости: $p(f(\cdot)) = p(\lim_{n \rightarrow \infty} f_n(\cdot)) = p(\sup_n f_n(\cdot)) = \sup_n p(f_n(\cdot)) = \lim_{n \rightarrow \infty} p(f_n(\cdot))$. Последнее равенство следует из монотонности $p(\cdot)$ (1.9).

Пусть $\{f_n(\cdot)\}$ — произвольная последовательность из $\mathcal{L}(X)$. Поскольку согласно (1.9) для любого $k \geq N$ $p(f_k(\cdot)) \geq p(\inf_{n \geq N} f_n(\cdot))$, $N = 1, 2, \dots$, то

$$\inf_{k \geq N} p(f_k(\cdot)) \geq p(\inf_{n \geq N} f_n(\cdot)), \quad N = 1, 2, \dots, \quad (1.10)$$

и, следовательно, в силу (1.8) и (1.10), $\sup_N p(\inf_{n \geq N} f_n(\cdot)) = p(\sup_N \inf_{n \geq N} f_n(\cdot)) \equiv$

$$\equiv p(\lim_{n \rightarrow \infty} f_n(\cdot)) \leq \sup_N \inf_{n \geq N} p(f_n(\cdot)) = \lim_{n \rightarrow \infty} p(f_n(\cdot)).$$

Аналогично, $\forall N = 1, 2, \dots$ $\sup_{k \geq N} p(f_k(\cdot)) = p(\sup_{k \geq N} f_k(\cdot)) \geq p(\inf_N \sup_{k \geq N} f_k(\cdot))$ и, следовательно, $\overline{\lim}_{n \rightarrow \infty} p(f_n(\cdot)) \geq p(\overline{\lim}_{n \rightarrow \infty} f_n(\cdot))$.

В частности, для всякой сходящейся последовательности $f_n(\cdot) \in \mathcal{L}(X)$, $n = 1, 2, \dots$,

$$p(\lim_{n \rightarrow \infty} f_n(\cdot)) \leq \lim_{n \rightarrow \infty} p(f_n(\cdot)), \quad (1.11)$$

т. е. $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$ — полунепрерывная снизу функция.

Итак, интеграл $p(\cdot)$ обладает следующими свойствами.

Теорема 3.1.1. Функция $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$ $\forall \{f_n(\cdot)\} \subset \mathcal{L}(X)$

1. монотонно не убывает: если $f_1(\cdot) \geq f_2(\cdot)$ означает, что $f_1(x) \geq f_2(x)$, $x \in X$, то $f_1(\cdot) \geq f_2(\cdot) \Rightarrow p(f_1(\cdot)) \geq p(f_2(\cdot))$; в частности, $p((f_1 \bullet f_2)(\cdot)) \leq p(f_1(\cdot)) \bullet p(f_2(\cdot))$, хотя $p((f_1 \bullet f_2)(\cdot)) = f_1 \bullet p(f_2(\cdot))$, если $f_1(x) = f_1 = \text{const}$, $x \in X$;
2. непрерывна относительно сходимости монотонно неубывающей последовательности: $f_{n+1}(\cdot) \geq f_n(\cdot)$, $n = 1, 2, \dots, \Rightarrow \Rightarrow p(\lim_{n \rightarrow \infty} f_n(\cdot)) = \lim_{n \rightarrow \infty} p(f_n(\cdot))$;
3. полунепрерывна снизу: $p(\lim_{n \rightarrow \infty} f_n(\cdot)) \leq \lim_{n \rightarrow \infty} p(f_n(\cdot))$; $p(\overline{\lim}_{n \rightarrow \infty} f_n(\cdot)) \leq \overline{\lim}_{n \rightarrow \infty} p(f_n(\cdot))$; если, в частности, $f(\cdot) = \lim_{n \rightarrow \infty} \{f_n(\cdot)\}$, то $p(f(\cdot)) \leq \lim_{n \rightarrow \infty} p(f_n(\cdot))$;
4. непрерывна в «точке» $f(\cdot) = 1(\cdot)$, $1(x) = 1$, $x \in X$, : если $1(\cdot) = \lim_{n \rightarrow \infty} \{f_n(\cdot)\}$, то $\lim_{n \rightarrow \infty} p(f_n(\cdot)) = p(1(\cdot))$; далее будем считать, что $p(1(\cdot)) = 1$ (условие нормировки).

Доказательство. Следует проверить лишь непрерывность $p(\cdot)$ в «точке» $1(\cdot)$. Пусть при $n \rightarrow \infty$ $f_n(\cdot) \rightarrow 1(\cdot)$, тогда в силу монотонности $p(\cdot)$ $f_n(\cdot) \leq 1(\cdot) \Rightarrow p(f_n(\cdot)) \leq p(1(\cdot))$ и, следовательно $\overline{\lim}_{n \rightarrow \infty} p(f_n(\cdot)) \leq p(1(\cdot))$. С другой стороны, в силу полунепрерывности $p(\cdot)$ снизу $\underline{\lim}_{n \rightarrow \infty} p(f_n(\cdot)) \geq p(1(\cdot))$, поэтому $\underline{\lim}_{n \rightarrow \infty} p(f_n(\cdot)) = \overline{\lim}_{n \rightarrow \infty} p(f_n(\cdot)) = \lim_{n \rightarrow \infty} p(f_n(\cdot)) = p(1(\cdot))$. ■

Пример 3.1.1. Пусть $\mathcal{L}(X) \stackrel{\text{def}}{=} \overline{\mathcal{L}}(X)$ — класс всех функций $X \rightarrow [0, 1]$. Поскольку согласно (1.6), (1.5) $\sup$ — символ суммирования, а $\min$ — символ умножения, определим интеграл $p(f(\cdot)) \stackrel{\text{def}}{=} p_g(f(\cdot))$, как «скалярное произведение» фиксированной функции $g(\cdot) \in \overline{\mathcal{L}}(X)$ на $f(\cdot) \in \overline{\mathcal{L}}(X)$:

$$p_g(f(\cdot)) \stackrel{\text{def}}{=} \sup_{x \in X} \min(f(x), g(x)) = \bigoplus_{x \in X} (f(x) \bullet g(x)), \quad f(\cdot) \in \mathcal{L}(X). \quad (1.12)$$

$p_g(\cdot)$ — линейная функция, ибо $\forall a_1, a_2 \in \mathcal{L}, \forall f_1(\cdot), f_2(\cdot) \in \overline{\mathcal{L}}(X)$

$$\begin{aligned} p_g(((a_1 \bullet f_1) \oplus (a_2 \bullet f_2))(\cdot)) &= \sup_{x \in X} \min\{\max[\min(a_1, f_1(x)), \\ &\min(a_2, f_2(x))], g(x)\} = \max\{\min(a_1, \sup_{x \in X} \min(f_1(x), g(x))), \\ &\min(a_2, \sup_{x \in X} \min(f_2(x), g(x)))\} \stackrel{\text{def}}{=} (a_1 \bullet p_g(f_1(\cdot))) \oplus (a_2 \bullet p_g(f_2(\cdot))), \end{aligned}$$

и вполне аддитивная, поскольку $p_g\left(\bigoplus_{j \in J} f_j(x)\right) = \sup_{x \in X} \min\left(\sup_{j \in J} f_j(x), g(x)\right) = \sup_{j \in J} \sup_{x \in X} \min(f_j(x), g(x)) = \sup_{j \in J} p_g(f_j(\cdot)) \stackrel{\text{def}}{=} \bigoplus_{j \in J} p_g(f_j(\cdot))$, где $f_j(\cdot) \in \mathcal{L}(X)$, $j \in J$, — произвольное семейство функций.

Наконец, согласно условию нормировки интеграла $p(\cdot)$ в Теореме 3.1.1 функция $g(\cdot)$ в (1.12) должна удовлетворять условию $\sup_{x \in X} g(\cdot) = 1$.

В [13] показано, что для продолженного на класс $\overline{\mathcal{L}}(X)$ интеграла $p(\cdot) : \mathcal{L}(X) \rightarrow \mathcal{L}$ равенство (1.12) представляет его общее выражение, см. также пример 3.1.3.

Заметим, что $p \bigoplus_{j \in J} g_j(f(\cdot)) = \bigoplus_{j \in J} p_{g_j}(f(\cdot))$, $p \bullet_{j \in J} g_j(f(\cdot)) \leq \bullet_{j \in J} p_{g_j}(f(\cdot))$.

3.1.2. Мера возможности. Определение, свойства. Чтобы определить меру возможности, следует конкретизировать содержимое класса $\mathcal{L}(X)$, включив в него индикаторные функции некоторых подмножеств X .

Лемма 3.1.1. Пусть $\mathcal{A}_0$ — конечная или счетная совокупность некоторых подмножеств X , $\mathcal{A} \triangleq \sigma\mathcal{A}_0$ — минимальная σ -алгебра подмножеств X , содержащая все подмножества из $\mathcal{A}_0$ ($\mathcal{A}$ —

σ -замыкание $\mathcal{A}_0$), $\mathcal{L}(X)$ — минимальный по включению класс функций $f(\cdot) : X \rightarrow [0, 1]$, содержащий индикаторные функции $\chi_A(\cdot) : X \rightarrow \{0, 1\}$ всех подмножеств $A \in \mathcal{A}_0$. Тогда $\mathcal{L}(X)$ класс всех $\mathcal{A}$ -измеримых функций $f(\cdot) : X \rightarrow [0, 1]$.

Доказательство. См. в [13]. ■

Всякое множество $A \in \mathcal{A}$ назовем *событием*; последнее можно задать его и. ф. $\chi_A(\cdot) \in \mathcal{L}(X)$.

Определение 3.1.2. Мерой возможности, или, короче, *возможностью*, назовем функцию $P(\cdot) : \mathcal{A} \rightarrow \mathcal{L}$, определенную равенством $P(A) = p(\chi_A(\cdot))$, $A \in \mathcal{A}$, ее значение $P(A) \stackrel{\text{def}}{=} p(\chi_A(\cdot)) \in \mathcal{L}$ назовем *возможностью события* $A \in \mathcal{A}$, шкалу $\mathcal{L} = ([0, 1], \leq, +, \bullet)$ назовем *шкалой значений возможности* P .

Теорема 3.1.2. Возможность $P(A)$, $A \in \mathcal{A}$, обладает следующими свойствами.

1. Для любых $A, B \in \mathcal{A}$

$$P(A \cup B) = p((\chi_A + \chi_B)(\cdot)) = P(A) + P(B), \quad (1.13)$$

и, как следствие $P(X) = P((X \setminus A) \cup A) = \max(P(X \setminus A), P(A))$, $A \in \mathcal{A}$. Кроме того, $P(A) \leq P(B)$, если $A \subset B$ (монотонность возможности), и, как следствие, $P(A \cap B) = p((\chi_A \bullet \chi_B)(\cdot)) \leq \min(P(A), P(B)) = P(A) \bullet P(B)$, $A, B \in \mathcal{A}$.

2. Возможность $P(\cdot)$ счетно-аддитивна: для любой последовательности событий $A_1, A_2, \dots$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sup_n P(A_n) \stackrel{\text{def}}{=} \bigoplus_{n=1}^{\infty} P(A_n). \quad (1.14)$$

Если $A_1 \subset A_2 \subset \dots$, и $A = \bigcup_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} A_n$, то, как следствие счетной аддитивности и монотонности $P(\cdot)$, получаем непрерывность $P(\cdot)$ относительно такой сходимости: $P(A) = \sup_n P(A_n) = \lim_{n \rightarrow \infty} P(A_n)$.

- 3.

$$P(\overline{A}) \leq \inf_N P\left(\bigcup_{n \geq N} A_n\right) = \lim_{N \rightarrow \infty} P\left(\bigcup_{n \geq N} A_n\right) = \lim_{N \rightarrow \infty} \sup_{n \geq N} P(A_n),$$

где $\overline{A} \stackrel{\text{def}}{=} \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n$ — верхний предел последовательности $A_1, A_2, \dots$; $\overline{A}$ — событие, каждая точка которого принадлежит бесконечно многим из $A_1, A_2, \dots$;

$$P(\underline{A}) = \sup_N P\left(\bigcap_{n \geq N} A_n\right) = \lim_{N \rightarrow \infty} P\left(\bigcap_{n \geq N} A_n\right) \leq \lim_{N \rightarrow \infty} \inf_{n \geq N} P(A_n),$$

где $\underline{A} \stackrel{\text{def}}{=} \bigcup_{N \geq N} \bigcap_{n \geq N} A_n$ — нижний предел последовательности $A_1, A_2, \dots$; $\underline{A}$ — событие, каждая точка которого принадлежит всем $A_1, A_2, \dots$, исключая, быть может, конечное их число.

Так как $\underline{A} \subset \overline{A}$, то $P(\underline{A}) \leq P(\overline{A})$.

Если последовательность $A_1, A_2, \dots$ сходится¹⁾ и $A = \underline{A} = \overline{A} \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} A_n$, то $P(A) \leq \varliminf_{n \rightarrow \infty} P(A_n)$, то есть возможность $P(\cdot) : \mathcal{A} \rightarrow \mathcal{L}$ полунепрерывна снизу. В частности, если $A_1 \supset A_2 \supset \dots$, то $A = \bigcap_{n=1}^{\infty} A_n$ и $P(A) \leq \lim_{n \rightarrow \infty} P(A_n)$, но если $A_1 \subset A_2 \subset \dots$, то $A = \bigcup_{n=1}^{\infty} A_n$ и $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.

4. Для любой сходящейся к X последовательности $A_1, A_2, \dots$ (т.е. такой, что $X = \bigcap_{N=1}^{\infty} \bigcap_{n \geq N} A_n = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n$) последовательность $P(A_1), P(A_2), \dots$ сходится и ее предел равен $P(X)$, т.е. возможность $P(\cdot)$ непрерывна в X , $P(X) = \lim_{n \rightarrow \infty} P(A_n)$, ибо согласно свойству 3. $P(X) \leq \varliminf_{n \rightarrow \infty} P(A_n)$, а в силу монотонности 1. $P(\cdot)$ $P(A_n) \leq P(X)$, $n = 1, 2, \dots$, и поэтому $P(X) \geq \limsup_{n \rightarrow \infty} P(A_n)$.

В соответствии с соглашением о нормировке $p(\cdot)$, принятым в теореме 3.1.1, $P(X) = p(1(\cdot)) = 1$.

Все утверждения 1. – 4. этой теоремы следуют соответственно из утверждений 1. – 4. теоремы 3.1.1

■

Третье свойство возможности означает, что значение $P(\emptyset)$ нельзя определить по непрерывности, поскольку возможность $P(A)$, $A \in \mathcal{A}$, не непрерывна при $A = \emptyset$; если $\emptyset = \lim_{n \rightarrow \infty} A_n$, то $P(\emptyset) \leq \varliminf_{n \rightarrow \infty} P(A_n)$, а при условии $A_1 \supset A_2 \supset \dots$, и $A_n \downarrow \emptyset$, $n \rightarrow \infty$, получим $P(\emptyset) \leq \lim_{n \rightarrow \infty} P(A_n)$. Значение $P(\emptyset)$ можно определить как произвольное число из $[0, \inf_{A \in \mathcal{A}} P(A)]$. При этом для любого события A $P(A \cup \emptyset) = \max(P(A), P(\emptyset)) = P(A)$, $P(A \cap \emptyset) \leq \min(P(A), P(\emptyset)) = P(\emptyset)$. Далее, если не оговорено противное, полагаем $P(\emptyset) = 0$.

¹⁾ Последовательность $A_1, A_2, \dots$ называется сходящейся, если ее нижний предел $\underline{A} \stackrel{\text{def}}{=} \bigcup_{N \geq N} \bigcap_{n \geq N} A_n$ совпадает с ее верхним пределом $\overline{A} \stackrel{\text{def}}{=} \bigcap_{N \geq N} \bigcup_{n \geq N} A_n$, множество $A = \underline{A} = \overline{A}$ называется ее пределом. Это определение эквивалентно определению поточечной сходимости последовательности индикаторных функций $\chi_{A_1}(\cdot), \chi_{A_2}(\cdot), \dots : \liminf_{n \rightarrow \infty} \chi_{A_n}(\cdot) \equiv \chi_{\underline{A}}(\cdot) = \chi_{\overline{A}}(\cdot) \equiv \limsup_{n \rightarrow \infty} \chi_{A_n}(\cdot)$, $\lim_{n \rightarrow \infty} \chi_{A_n}(\cdot) = \chi_A(\cdot)$.

Любую функцию $P(A)$, $A \in \mathcal{A}$, обладающую свойствами, перечисленными в теореме 3.1.2, будем называть *возможностью*.

Тройку $(X, \mathcal{A}, P)$ назовем *пространством с возможностью*.

Пример 3.1.2. Возвращаясь к интегралу $p(\cdot) = p_g(\cdot)$ (1.12), заметим, что в этом случае $\mathcal{A} = \mathcal{P}(X)$ — класс всех подмножеств X , и возможность события $A \in \mathcal{P}(X)$, $A \neq \emptyset$, задается равенством

$$P_g(A) \stackrel{\text{def}}{=} p_g(\chi_A(\cdot)) = \sup_{x \in A} g(x), \quad P_g(\emptyset) = \sup_{x \in \emptyset} g(x) \stackrel{\text{def}}{=} 0. \quad (1.15)$$

Здесь $g(\cdot) \in \overline{\mathcal{L}}(X)$, причем $\sup_{x \in X} g(x) = 1$. Согласно (1.15) $g(\cdot)$ естественно назвать *распределением (вполне аддитивной) возможности* $P_g(\cdot)$, поскольку $g(x) = P_g(\{x\})$, $x \in X$. В [13] показано, что *возможность* $P(\cdot)$ всегда может быть продолжена на алгебру $\mathcal{P}(X)$ всех подмножеств X и задана как $P_g(\cdot)$ в (1.15).

Пример 3.1.3. Если $X = \{x_1, x_2, \dots\}$, то для любого $A \subset X$ $P(A) = P(\bigcup_{i: x_i \in A} \{x_i\}) = \sup_{i: x_i \in A} P(\{x_i\})$, а так как любую функцию $f(\cdot) : X \rightarrow [0, 1]$ можно представить в виде $f(x) = \bigoplus_{i=1}^{\infty} (f(x_i) \bullet \chi_{\{x_i\}}(x)) = \sup_i \min(f(x_i), \chi_{\{x_i\}}(x))$, $x \in X$, то согласно определению 3.1.1

$$\begin{aligned} p(f(\cdot)) &= \bigoplus_{i=1}^{\infty} (f(x_i) \bullet p(\chi_{\{x_i\}}(\cdot))) = \\ &= \bigoplus_{i=1}^{\infty} (f(x_i) \bullet P(\{x_i\})) \stackrel{\text{def}}{=} \sup_i \min(f(x_i), P(\{x_i\})) \end{aligned} \quad (1.16)$$

— общий вид интеграла $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$, см. пример 3.1.1.

3.1.3. Нечеткие множества, элементы, события. В 1965 г. Л. А. Заде предложил новый подход к моделированию нечеткости, основанный на понятии нечеткого множества, [38]. Как известно, любое подмножество A множества X можно задать с помощью его индикаторной функции (и. ф.) $\chi_A(\cdot) : X \rightarrow \{0, 1\}$, определив $\chi_A(x) = 1$ для $x \in A$ и $\chi_A(x) = 0$ для $x \in X \setminus A$, $x \in X$. Нечеткое (под)множество A в [38] также определяется его и. ф. $\mu_A(\cdot) : X \rightarrow [0, 1]$, названной *функцией принадлежности*, значение которой $\mu_A(x) \in [0, 1]$ интерпретируется как «степень включения» $x \in X$ в A . Тот факт, что любой элемент $x \in X$ может принадлежать нечеткому множеству A лишь «отчасти», позволяет моделировать сложные объекты в терминах характеристик, значения которых свойственны им лишь «до некоторой степени», «частично».

На связь теории нечетких множеств с теорией возможностей впервые указал Л. Заде в работе [39].

В теории нечетких множеств Л. А. Заде операции над нечеткими множествами определены как операции над их и.ф. согласно следующим правилам:

$$\begin{array}{ll}
 \text{а) } \mu_A(x) = \mu_B(x) \Leftrightarrow A = B; & \text{д) } \mu_{\bigcup_{\alpha} A_{\alpha}}(x) = \sup_{\alpha} \mu_{A_{\alpha}}(x); \\
 \text{б) } \mu_A(x) \leq \mu_B(x) \Leftrightarrow A \subset B; & \text{е) } \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)); \\
 \text{в) } \mu_{X \setminus A}(x) = 1 - \mu_A(x); & \text{ж) } \mu_{\bigcap_{\alpha} A_{\alpha}}(x) = \inf_{\alpha} \mu_{A_{\alpha}}(x), \quad x \in X. \\
 \text{г) } \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)); &
 \end{array}
 \quad (1.17)$$

Здесь A, B — нечеткие (под)множества X , и значения функций принадлежности, записанных в левом столбце и *определяющих теоретико-множественные операции* $X \setminus A, A \cup B, \dots, \bigcap_{\alpha} A_{\alpha}$, расшифровываются в правом. Заменив в определениях (1.17) $\mu(\cdot)$ на «обычную» и.ф. $\chi(\cdot)$, получим соответствующие операции над «обычными», «четкими» множествами и над их и.ф., согласно которым $\max(\chi_A, \chi_B)(\cdot)$ и $\min(\chi_A, \chi_B)(\cdot)$ суть и.ф. $A \cup B$ и $A \cap B$, условие $\chi_A(\cdot) \leq \chi_B(\cdot)$ эквивалентно $A \subset B$ и т. д.

Заметим, что в теории нечетких множеств Л. А. Заде принимаются следующие условия: $\mu_X(x) = 1, \mu_{\emptyset}(x) = 0, x \in X$, хотя согласно правилам (1.17) $\mu_{(X \setminus A) \cup A}(x) \leq 1, \mu_{(X \setminus A) \cap A}(x) \geq 0, x \in X$, т. е. для некоторого A , вообще говоря, $(X \setminus A) \cup A \neq X, (X \setminus A) \cap A \neq \emptyset$.

Обсудим вопрос о моделировании нечеткости в теории возможностей.

3.1.3.1 Нечеткие множества. Нечеткость может быть охарактеризована в терминах возможности, если значение $\mu_A(x)$ интерпретировать как величину $g^A(x)$ «возможности покрытия» элемента $x \in X$ нечетким множеством A , т. е. если считать, что $g^A(x) \stackrel{\text{def}}{=} P(x \in A)$ — возможность нечеткого события $x \in A$, а $g^A(\cdot) : X \rightarrow [0, 1]$ — индикаторная функция (одноточечного покрытия (и.ф.о.п.)). При такой интерпретации нечеткое множество следует рассматривать как возможностный аналог случайного множества в теории вероятностей. В общих чертах эту точку зрения можно пояснить следующим образом. Пусть $(Y, \mathcal{B}, P_Y)$ — пространство с возможностью, $A^* : Y \rightarrow \mathcal{P}(X)$ — отображение, ставящее в соответствие каждому $y \in Y$ множество $A^y \subset X$, причем так, что для любого $x \in X$ множество $A_x = \{y \in Y, x \in A^y\} \in \mathcal{B}$ (измеримо); A_x — множество тех $y \in Y$, при которых элемент $x \in X$ покрывается множеством ¹⁾ $A^y, g^A(x) \stackrel{\text{def}}{=} P_Y(A_x)$ — возможность покрытия этого $x \in X$.

Поскольку, с одной стороны, множество тех $x \in X$, при которых $A_x \in \mathcal{B}$, вообще говоря, может оказаться пустым, а с другой стороны,

¹⁾ Отображение $A^* : X \rightarrow \mathcal{P}(Y)$ называется обратным к A^* , разумеется, отображение $A^* : Y \rightarrow \mathcal{P}(X)$ — обратное к A , и включения $x \in A^y$ и $y \in A_x, x \in X, y \in Y$, эквивалентны.

как показано в [13], возможность P_Y в $(Y, \mathcal{B}, P_Y)$ всегда может быть продолжена на класс $\mathcal{P}(Y)$ всех подмножеств Y , далее, рассматривая нечеткие множества, условимся считать, что $\mathcal{A} = \mathcal{P}(X)$, $\mathcal{B} = \mathcal{P}(Y)$ и, как следствие, не будем касаться вопросов измеримости множеств $A_x \subset Y$, $x \in X$, и $A^y \subset X$, $y \in Y$.

Отображение $A : Y \rightarrow \mathcal{P}(X)$ назовем нечетким множеством A , определенным на $(Y, \mathcal{P}(Y), P_Y)$, принимающим значения в $\mathcal{P}(X)$, возможность¹⁾ $g^A(x) = P_Y(A_x)$, как функцию $x \in X$, назовем его индикаторной функцией (и. ф.), ее значение $g^A(x)$ будем интерпретировать как возможность события $x \in A$, т. е. как возможность покрытия элемента $x \in X$ нечетким множеством $A \subset X$.

Пусть $B : Y \rightarrow \mathcal{P}(X)$ нечеткое множество B и $g^B(x) = P(B_x)$, $x \in X$, — его и. ф., тогда $A \cup B \stackrel{\text{def}}{=} C : Y \rightarrow \mathcal{P}(X)$, $C^y \stackrel{\text{def}}{=} A^y \cup B^y$, $y \in Y$, $A \cap B \stackrel{\text{def}}{=} D : Y \rightarrow \mathcal{P}(X)$, $D^y \stackrel{\text{def}}{=} A^y \cap B^y$, $y \in Y$, суть, по определению, нечеткие множества $C = A \cup B$ и, соответственно, $D = A \cap B$, причем $C_x = \{y \in Y, x \in A^y \cup B^y\} = A_x \cup B_x$, $D_x = \{y \in Y, x \in A^y \cap B^y\} = A_x \cap B_x$, $x \in X$. Следовательно, как в (1.17) г) $g^C(x) = P(A_x \cup B_x) = \max(P(A_x), P(B_x)) = \max(g^A(x), g^B(x))$, но, в отличие от (1.17) е), $g^D(x) = P(A_x \cap B_x) \leq \min(P(A_x), P(B_x)) = \min(g^A(x), g^B(x))$, $x \in X$.

Нечеткое множество A содержится в нечетком множестве B , $A \subset B$, если для каждого $y \in Y$ $A^y \subset B^y$. Так как при этом для каждого $x \in X$ $A_x = \{y \in Y, x \in A^y\} \subset \{y \in Y, x \in B^y\} = B_x$, то $g^A(x) \leq g^B(x)$, $x \in X$. Обратное, разумеется, неверно: неравенства $g^A(x) \leq g^B(x)$, $x \in X$, не означают, что $A \subset B$.

Определим нечеткое множество $X \setminus A$, дополнительное к A , как отображение $Y \rightarrow \mathcal{P}(X)$, ставящее в соответствие каждому $y \in Y$ множество $X \setminus A^y$. Тогда $(X \setminus A)_x = \{y \in Y, x \in X \setminus A^y\} = Y \setminus A_x$, $x \in X$, и, следовательно, $g^{X \setminus A}(x) = P_Y(Y \setminus A_x)$, $g^{A \cup (X \setminus A)}(x) = \max(P_Y(A_x), P_Y(Y \setminus A_x)) = P(Y) = g^X(x) = 1$, $x \in X$, причем в отличие от (1.17) в) и. ф. $g^A(\cdot)$, вообще говоря, не определяет и. ф. $g^{X \setminus A}(\cdot)$.

Для так определенного нечеткого множества $X \setminus A$, дополнительного к A , условие (1.17) в), вообще говоря, не выполняется.

Наконец, разность $C = A \setminus B$ нечетких множеств A и B определим как функцию $C : Y \rightarrow \mathcal{P}(X)$ условием $C^y = A^y \setminus B^y$, $y \in Y$. Это нечеткое множество можно представить и как $A^y \cap (X \setminus B^y)$, $y \in Y$.

Множество $\Gamma_A \stackrel{\text{def}}{=} \{(y, x) \in Y \times X, y \in A_x\} \equiv \{(y, x) \in Y \times X, x \in A^y\}$ называется графиком нечеткого множества $A : Y \rightarrow \mathcal{P}(X)$, см. рис. 3.4.

¹⁾ Далее, чтобы «степень $\mu_A(x)$ включения» x в A отличать от «возможности покрытия» x нечетким множеством A , последнюю будем обозначать $g^A(x)$.

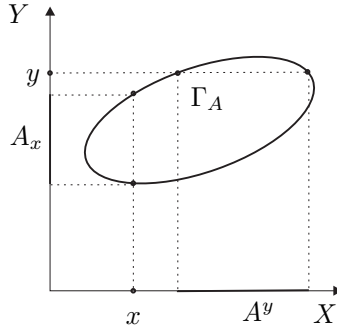


Рис. 3.4. Отображение $A: Y \rightarrow \mathcal{P}(X)$, его график $\Gamma_A \subset Y \times X$ и обратное к A отображение $A^*: X \rightarrow \mathcal{P}(Y)$.

3.1.3.2 Нечеткие элементы. Еще одна точка зрения на нечеткость, также позволяющая охарактеризовать ее в терминах возможности, опирается на понятие *нечеткого элемента*, аналогичное понятию случайного элемента в теории вероятностей.

Пусть опять $(Y, \mathcal{B}, P_Y)$ — пространство с возможностью, $(X, \mathcal{A})$ — измеримое пространство, $\mathcal{A}$ — σ -алгебра подмножеств X , и задана функция $q(\cdot): Y \rightarrow X$, такая, что для любого множества $A \in \mathcal{A}$, $q^{-1}(A) \stackrel{\text{def}}{=} \{y \in Y, q(y) \in A\} \in \mathcal{B}$; такая функция $q(\cdot)$ называется $(\mathcal{B}, \mathcal{A})$ -измеримой и задает то, что называется *нечетким элементом*, обозначим его ξ , определенным на $(Y, \mathcal{B}, P_Y)$ и принимающим значения в $(X, \mathcal{A})$. Отображение $q(\cdot): Y \rightarrow X$ индуцирует на $\mathcal{A}$ возможность $P_X: P_X(A) \stackrel{\text{def}}{=} P_Y(q^{-1}(A))$, $A \in \mathcal{A}$, и тем самым определяет пространство с возможностью $(X, \mathcal{A}, P_X)$ как образ $(Y, \mathcal{B}, P_Y)$.

Нечеткий элемент ξ определяет на $(X, \mathcal{A})$ возможность $P^\xi: P^\xi(\xi \in A) \stackrel{\text{def}}{=} P_X(A)$, $A \in \mathcal{A}$, назовем его *каноническим для пространства с возможностью* $(X, \mathcal{A}, P_X) = (X, \mathcal{A}, P^\xi)$.

Чтобы не касаться вопросов $\mathcal{A}, \mathcal{B}$ -измеримости условимся далее, как в § 3.1.3.1, считать, что $\mathcal{A} = \mathcal{P}(X)$, $\mathcal{B} = \mathcal{P}(Y)$ и определим функцию $g^\xi(x) \triangleq P^\xi(\xi = x) = P_Y(q^{-1}(\{x\}))$, $x \in X$, называемую *распределением возможностей значений нечеткого элемента ξ* , или короче — *распределением нечеткого элемента ξ* ; по определению значение $g^\xi(x)$ есть возможность равенства $\xi = x$, а $P^\xi(\xi \in A) = P^\xi(\xi \in \bigcup_{x \in A} \{x\}) = \sup_{x \in A} g^\xi(x)$.

Итак, нечеткое множество A , заданное на $(Y, \mathcal{P}(Y), P_Y)$, определено как многозначное отображение $A: Y \rightarrow \mathcal{P}(X)$, $g^A(x) = P_Y(\{y \in Y, x \in A^y\})$ — возможность покрытия $x \in X$ этим нечетким множеством, нечеткий элемент ξ , заданный на $(Y, \mathcal{P}(Y), P_Y)$ со значениями в $(X, \mathcal{P}(X))$, определен как функция $q(\cdot): Y \rightarrow X$, $g^\xi(x) = P_Y(\{y \in Y, q(y) = x\})$ — возможность равенства $\xi = x \in X$.

Выберем в качестве $(X, \mathcal{P}(X))$ измеримое пространство $(Y, \mathcal{P}(Y))$ и обозначим η нечеткий элемент, канонический для $(Y, \mathcal{P}(Y), P_Y)$. Его распределение $g^\eta(\cdot) : Y \rightarrow [0, 1]$ определим равенством $g^\eta(y) = P^\eta(\eta = y) \stackrel{\text{def}}{=} P_Y(\{y\})$, $y \in Y$, согласно которому $P_Y(B) = P_Y(\bigcup_{y \in B} \{y\}) = \sup_{y \in B} P_Y(\{y\}) = \sup_{y \in B} g^\eta(y) \stackrel{\text{def}}{=} P^\eta(\eta \in B)$, $B \in \mathcal{P}(Y)$, см. пример 3.1.2.

Понятие канонического нечеткого элемента позволяет единообразно охарактеризовать конструкции нечеткого элемента и нечеткого множества, определенных на $(Y, \mathcal{P}(Y), P_Y)$, а именно,

- нечеткий элемент ξ есть образ $q(\eta)$ канонического нечеткого элемента η при отображении $q(\cdot) : Y \rightarrow X$;
- нечеткое множество A есть образ A^η канонического нечеткого элемента η при отображении $A^\cdot : Y \rightarrow \mathcal{P}(X)$, см. рис. 3.5;

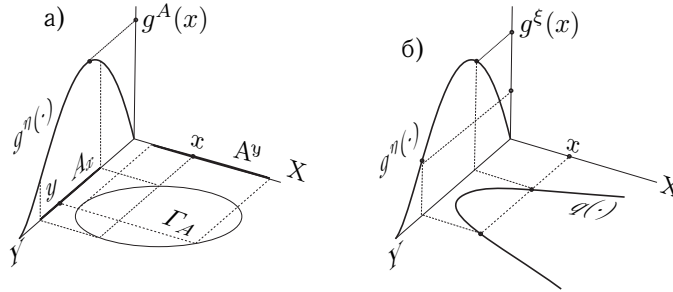


Рис. 3.5. а) Нечеткое множество A^η , его график Γ_A , его индикаторная функция $g^A(\cdot)$, множества A_x, A_y , распределение $g^\eta(\cdot)$ нечеткого элемента η . б) Нечеткий элемент $\xi = q(\eta)$, значение его распределения $g^\xi(\cdot)$ в точке $x \in X$.

Пусть $(Y, \mathcal{P}(Y), P_Y)$ и $(X, \mathcal{P}(X), P_X)$ — пространства с возможностями, для которых каноническими являются нечеткие элементы η и соответственно $\xi = q(\eta)$, $P^\xi(\xi \in A) \stackrel{\text{def}}{=} P^\eta(q(\eta) \in A) = \sup_{y \in q^{-1}(A)} g^\eta(y)$, $A \in \mathcal{P}(X)$. Тогда $P^\xi(A) = \sup_{x \in A} g^\xi(x)$, $A \in \mathcal{P}(X)$,

где $g^\xi(x) \stackrel{\text{def}}{=} P^\xi(\xi = x) = P^\eta(\eta \in q^{-1}(x)) = \sup_{y \in q^{-1}(x)} g^\eta(y)$, $x \in X$,

а $g^A(x) = P^\eta(x \in A^\eta) = P^\eta(\eta \in A_x) \equiv P_Y(A_x) = \sup_{y \in A_x} g^\eta(y)$ — возможность покрытия $x \in X$ нечетким множеством $A^\cdot : Y \rightarrow \mathcal{L}(X)$.

На рис. 3.6 приведены примеры нечетких элементов ξ_s — «короткий», ξ_m — «средний», ξ_l — «длинный» и нечетких множеств A_s^η «коротких элементов X », A_m^η — «средних», A_l^η — «длинных». Нечеткие элементы ξ_s, ξ_l и нечеткие множества $A_s^\eta, A_m^\eta, A_l^\eta$ определены на пространстве с возможностью $(Y, \mathcal{P}(Y), P_Y)$, в котором $Y = [0, 1]$ и возможность P_Y задана распределением $g^\eta(y) = y$, $y \in [0, 1]$, канонического для $(Y, \mathcal{P}(Y), P_Y)$ нечеткого элемента η . Нечеткий элемент ξ_m ,

распределенный согласно рис. 3.6, б), не может быть определен на $(Y, \mathcal{P}(Y), P_Y)$ (не является функцией η).

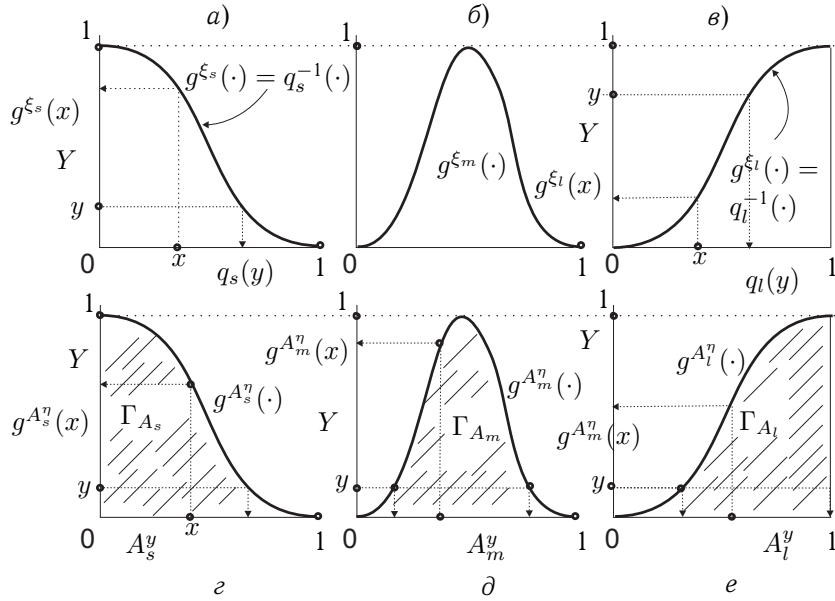


Рис. 3.6. Нечеткие элементы $\xi_s = q_s(\eta)$, $\xi_l = q_l(\eta)$ заданы функциями $q_s(\cdot) : Y \rightarrow X = [0, 1]$, $q_l(\cdot) : Y \rightarrow X = [0, 1]$, графики которых приведены на рис. а) и в), где показаны и распределения $g^{\xi_s}(\cdot) = q_s^{-1}(\cdot)$, $g^{\xi_l}(\cdot) = q_l^{-1}(\cdot)$. Функция $q_m(\cdot) : Y \rightarrow X$, которая определила бы нечеткий элемент ξ_m , распределенный как показано на рис. б), не существует. На рис. з), д), е) приведены графики Γ_{A_s} , Γ_{A_m} , Γ_{A_l} и индикаторные функции $g^{A_s}(\cdot)$, $g^{A_m}(\cdot)$, $g^{A_l}(\cdot)$ одно-точечного покрытия нечетких множеств A_s, A_m, A_l соответственно. В данном случае $g^{A_s^\eta}(x) = g^{\xi_s}(x) = q_s^{-1}(x)$ — возможность покрытия $x \in X = [0, 1]$ нечетким множеством A_s^η и возможность равенства $\xi_s = x$. Аналогично $g^{A_m^\eta}(x) = g^{\xi_m}(x)$ и $g^{A_l^\eta}(x) = g^{\xi_l}(x)$.

3.1.3.3 Нечеткие события. Во многих нечетких моделях обе точки зрения на нечеткость можно согласованно представить, используя интеграл $p_{g^\xi}(\cdot)$ (1.12). Если A — «четкое» подмножество X , $\chi_A(\cdot)$ — его и.ф., то возможность включения $\xi \in A$ (события $\xi \in A$) определяется равенством $P^\xi(\xi \in A) = p_{g^\xi}(\chi_A(\cdot)) = \sup_{x \in A} g^\xi(x) = \sup_{x \in X} \min(g^\xi(x), \chi_A(x))$. Условие нормировки $p_{g^\xi}(\cdot)$ означает, что $P^\xi(\xi \in X) = \sup_{x \in X} g^\xi(x) = 1$. Соответственно значение $p_{g^\xi}(g^A(\cdot)) = \sup_{x \in X} \min(g^\xi(x), g^A(x))$ определит возможность покрытия ξ нечетким множеством A с и.ф. $g^A(\cdot)$, т.е. возможность нечеткого

события, состоящего в том, что нечеткий элемент ξ покрывается нечетким множеством A (значение $\min$ равно возможности того, что $\xi = x$ и $x \in A$, величина $\sup$ определяет возможность включения $\xi \in A$; события $\xi = x$ и $x \in A$ считаются *независимыми* при любом $x \in X$, см. определение 3.1.16). Если нечеткий элемент ξ таков, что $x_0 \in X$ — единственное «возможное» его значение, т. е. если $g^\xi(x) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases}$, $x \in X$, то ξ , по существу, — «четкий» элемент $x_0 \in X$, и возможность его покрытия нечетким множеством A , равная $\sup_{x \in X} \min(g^\xi(x), f^A(x)) = f^A(x_0)$, совпадает с возможностью покрытия элемента $x_0 \in X$ нечетким множеством A .

Если в нечеткой модели нечеткий элемент $\xi = q(\eta)$ ненаблюдаем и, следовательно, нечеткие события $\xi \in B$, $B \in \mathcal{P}(X)$, не могут быть зарегистрированы, но наблюдаемо нечеткое множество A^η и могут быть зарегистрированы, например, события $A^\eta \cap B \neq \emptyset$ и $A^\eta \subset B$, $A^\eta \neq \emptyset, \Leftrightarrow A^\eta \cap (X \setminus B) = \emptyset, A^\eta \neq \emptyset$, то возможность $P^\eta(A^\eta \cap B \neq \emptyset) \stackrel{\text{def}}{=} P_Y(\{y \in Y, A^y \cap B \neq \emptyset\}) = P_Y(\{y \in Y, \bigcup_{x \in B} (A^y \cap \{x\}) \neq \emptyset\}) = P_Y(\bigcup_{x \in B} \{y \in Y, A^y \cap \{x\} \neq \emptyset\}) = \sup_{x \in B} P_Y(\{y \in Y, A^y \cap \{x\} \neq \emptyset\}) \stackrel{\text{def}}{=} \sup_{x \in B} g^A(x)$ выражается через и.ф.о.п. A^η , а возможность

$$\begin{aligned} P^\eta(A^\eta \cap (X \setminus B) = \emptyset, A^\eta \neq \emptyset) &\stackrel{\text{def}}{=} P_Y(\{y \in A_X, A^y \cap (X \setminus B) = \emptyset\}) = \\ &= P_Y(A_X \setminus \{y \in A_X, A^y \cap (X \setminus B) \neq \emptyset\}) = \\ &= P_Y(A_X \setminus \bigcup_{x \in X \setminus B} \{y \in A_X, A^y \cap \{x\} \neq \emptyset\}) = P_Y(A_X \setminus A_{X \setminus B}), \end{aligned} \quad (1.18)$$

где

$$A_X \stackrel{\text{def}}{=} \bigcup_{x \in X} A_x = \{y \in Y, A^y \neq \emptyset\}, \quad A_{X \setminus B} \stackrel{\text{def}}{=} \bigcup_{x \in X \setminus B} A_x, \quad (1.19)$$

— не выражается.

На самом деле возможности многих событий, в которых «участвуют» нечеткие множества, не выражаются через их и.ф.о.п. Характерным примером могут служить события, связанные с последовательностями нечетких множеств и их пределами.

Пусть $A_1, A_2, \dots$ — последовательность нечетких множеств $A_i: Y \rightarrow \mathcal{P}(X)$, $i = 1, 2, \dots$. Нечеткие множества $\bar{A} \stackrel{\text{def}}{=} \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$ и $\underline{A} \stackrel{\text{def}}{=} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k$ называются *верхним и, соответственно, нижним её пределами* последовательности $\{A_i\}$; если $\bar{A} = \underline{A}$, то последовательность $A_1, A_2, \dots$ назовем *сходящейся*, а нечеткое множество $A = \bar{A} = \underline{A}$ — *её пределом*.

Для и.ф. нечетких множеств $\bar{A}$ и $\underline{A}$ найдем: $g^{\bar{A}}(x) = P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_{kx}) \leq \inf_n \sup_{k \geq n} g^{A_k}(x)$, $g^{\underline{A}}(x) = P(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_{kx}) \leq \sup_n \inf_{k \geq n} g^{A_k}(x)$, $x \in X$, и лишь для монотонно возрастающей последовательности $A_1 \subset A_2 \subset \dots$ $g^{\bar{A}}(x) = g^{\underline{A}}(x) = g^A(x) = \sup_n g^{A_n}(x)$, $x \in X$, поскольку в этом случае $\bar{A} = \underline{A} = A = \bigcup_{n=1}^{\infty} A_n$.

3.1.4. Мера необходимости. Определение, свойства. Рассмотрим шкалу $\tilde{\mathcal{L}} = ([0, 1], \tilde{\leq}, \tilde{+}, \tilde{\bullet})$, дуально изоморфную¹⁾ $\mathcal{L} = ([0, 1], \leq, +, \bullet)$, с операцией сложения « $\tilde{+}$ », определенной как « $\min$ », и операцией умножения « $\tilde{\bullet}$ », определенной как « $\max$ », с нейтральными элементами $\tilde{\mathbf{0}} = 1$ и $\tilde{\mathbf{1}} = 0$ и упорядоченностью, определенной отношением $\tilde{\leq}$, обратным естественному. При таком определении все свойства операций « $\tilde{+}$ » и « $\tilde{\bullet}$ » и нейтральных элементов могут быть сформулированы как соотношения (1.1), (1.2), (1.4), (1.3), если всюду $+$, $\bullet$ и $\leq$ заменить соответственно на $\tilde{+}$, $\tilde{\bullet}$ и на $\tilde{\leq}$, а именно: $a \tilde{\bullet} (b \tilde{+} c) = (a \tilde{\bullet} b) \tilde{+} (a \tilde{\bullet} c)$, $a \tilde{+} (b \tilde{\bullet} c) = (a \tilde{+} b) \tilde{\bullet} (a \tilde{+} c)$, $a \tilde{\bullet} \tilde{\mathbf{0}} = \tilde{\mathbf{0}}$, $a \tilde{+} \tilde{\mathbf{0}} = a$, $a \tilde{\bullet} \tilde{\mathbf{1}} = a$, $a \tilde{+} \tilde{\mathbf{1}} = \tilde{\mathbf{1}}$, $a \tilde{\leq} b \Rightarrow \begin{cases} a \tilde{\bullet} c \tilde{\leq} b \tilde{\bullet} c, \\ a \tilde{+} c \tilde{\leq} b \tilde{+} c \end{cases}$, $a, b, c \in \tilde{\mathcal{L}}$, $\tilde{\mathbf{0}} \tilde{\leq} \tilde{\mathbf{1}}$.

Класс $\tilde{\mathcal{L}}(X)$ функций на X со значениями в $\tilde{\mathcal{L}}$ определим аналогично тому, как определен класс $\mathcal{L}(X)$, но, сохранив обозначения $\max$, $\min$, $\sup$ и $\inf$, отвечающие естественной упорядоченности на $[0, 1]$, вместо условий (1.5) будем считать, что $(f_1 \tilde{+} f_2)(x) \stackrel{\text{def}}{=} f_1(x) \tilde{+} f_2(x) \stackrel{\text{def}}{=} \min(f_1(x), f_2(x))$; $(f_1 \tilde{\bullet} f_2)(x) \stackrel{\text{def}}{=} f_1(x) \tilde{\bullet} f_2(x) \stackrel{\text{def}}{=} \max(f_1(x), f_2(x))$, $x \in X$, и соответственно вместо (1.6) $\bigcap_{n=1}^{\infty} f_n(\cdot) \stackrel{\text{def}}{=} \inf_n f_n(\cdot) \in \tilde{\mathcal{L}}(X)$, $\bigcup_{n=1}^{\infty} f_n(\cdot) \stackrel{\text{def}}{=} \sup_n f_n(\cdot) \in \tilde{\mathcal{L}}(X)$, если $\{f_n(\cdot)\} \subset \tilde{\mathcal{L}}(X)$.

Определение 3.1.3. Определим интеграл $n(\cdot)$ как линейную счетно-аддитивную функцию на $\tilde{\mathcal{L}}(X)$, принимающую значения в $\tilde{\mathcal{L}}$, т. е. такую, что

$$\forall a_1, a_2 \in \tilde{\mathcal{L}}, \forall f_1(\cdot), f_2(\cdot) \in \tilde{\mathcal{L}}(X)$$

¹⁾ Упорядоченное множество $\tilde{\mathcal{L}}$ называется дуальным (или двойственным) $\mathcal{L}$, если оно определено на тех же, что и $\mathcal{L}$, элементах отношением порядка $\tilde{\leq}$, обратным отношению $\leq$, определившему $\mathcal{L}$, а именно, $a \leq b \Leftrightarrow b \tilde{\leq} a$, $a, b \in [0, 1]$, $[1]$.

$$\begin{aligned}
n(((a_1 \bullet f_1) \dot{+} (a_2 \bullet f_2))(\cdot)) &= (a_1 \bullet n(f_1(\cdot))) \dot{+} (a_2 \bullet n(f_2(\cdot))), \\
\forall \{f_n(\cdot)\} &\subset \tilde{\mathcal{L}}(X) \\
n(\inf_n f_n(\cdot)) &\stackrel{\text{def}}{=} n(\dot{+}_{n=1}^{\infty} f_n(\cdot)) = \dot{+}_{n=1}^{\infty} n(f(\cdot)) \stackrel{\text{def}}{=} \inf_n n(f_n(\cdot)).
\end{aligned} \tag{1.20}$$

Мерой необходимости, или просто необходимостью, назовем функцию $N(\cdot) : \mathcal{A} \rightarrow \tilde{\mathcal{L}}$, ее значение $N(A) \stackrel{\text{def}}{=} n(\chi_A(\cdot))$ назовем мерой необходимости события $A \in \mathcal{A}$, или, короче, — необходимостью A . Шкалу $\tilde{\mathcal{L}} = ([0, 1], \lesssim, \dot{+}, \bullet)$ назовем шкалой значений необходимости.

Теорема 3.1.3. Функция $n(\cdot) : \tilde{\mathcal{L}}(X) \rightarrow \tilde{\mathcal{L}} \quad \forall \{f_t(\cdot)\} \subset \tilde{\mathcal{L}}(X)$

1. монотонно не убывает: если $f_1(\cdot) \lesssim f_2(\cdot)$, то $n(f_1(\cdot)) \lesssim n(f_2(\cdot))$;
2. непрерывна относительно сходимости монотонной последовательности: $f_t(\cdot) \lesssim f_{t+1}(\cdot)$, $t = 1, 2, \dots$, $\Rightarrow n(\lim_{t \rightarrow \infty} f_t(\cdot)) = \lim_{t \rightarrow \infty} n(f_t(\cdot))$;
3. полунепрерывна сверху: $n(\overline{\lim}_{t \rightarrow \infty} f_t(\cdot)) \lesssim \overline{\lim}_{t \rightarrow \infty} n(f_t(\cdot))$,
 $n(\underline{\lim}_{t \rightarrow \infty} f_t(\cdot)) \lesssim \underline{\lim}_{t \rightarrow \infty} n(f_t(\cdot))$;
4. непрерывна в точке $f(\cdot) = 0(\cdot)$, $0(x) \equiv 0 = \tilde{\mathbf{1}}$, $x \in X$; т. е. для любой последовательности $\{f_t(\cdot)\}$, сходящейся к $0(\cdot)$, $\lim_{t \rightarrow \infty} n(f_t(\cdot)) = n(0(\cdot))$, причем далее будем считать, что $n(0(\cdot)) = 0 = \tilde{\mathbf{1}}$ (условие нормировки).

Доказательство аналогично доказательству теоремы 3.1.1. ■

Теорема 3.1.4. Имеют место следующие свойства $N(\cdot)$:

- $N(A \cap B) = \min(N(A), N(B)) = N(A) \dot{+} N(B)$,
1. $N(A \cup B) \lesssim \max(N(A), N(B)) = N(A) \bullet N(B)$,
 $A \subset B \Rightarrow N(A) \gtrsim N(B)$, $A, B \in \mathcal{A}$.
2. Необходимость $N(\cdot)$ счетно аддитивна:
 $N(\dot{+}_{n=1}^{\infty} A_n) \stackrel{\text{def}}{=} N(\bigcap_{n=1}^{\infty} A_n) = \inf_n N(A_n) \stackrel{\text{def}}{=} \dot{+}_{n=1}^{\infty} N(A_n)$.
Если $A_1 \supset A_2 \supset \dots$ и, следовательно, $A = \bigcap_{n=1}^{\infty} A_n \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} A_n$,
то из счетной аддитивности и монотонности $N(\cdot)$ следует непрерывность $N(\cdot)$ относительно такой сходимости: $N(A) = \inf_n N(A_n) = \lim_{n \rightarrow \infty} N(A_n)$.
3. В общем случае, если $A = \lim_{n \rightarrow \infty} A_n$, то

$$N(A) \lesssim \inf_N \sup_{n \geq N} N(A_n) \stackrel{\text{def}}{=} \overline{\lim}_{n \rightarrow \infty} N(A_n), \tag{1.21}$$

т. е. $N(\cdot)$ полунепрерывна сверху. Значение $N(X)$ можно определить любым числом из промежутка $[\sup_{A \in \mathcal{A}} N(A), 1]$.

4. Для любой сходящейся к пустому множеству $\emptyset$ последовательности $A_1, A_2, \dots$ (т.е. такой, что $\bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n = \emptyset$) предел последовательности $N(A_1), N(A_2), \dots$ существует и равен $N(\emptyset)$, т.е. необходимость $N(\cdot)$ непрерывна в $\emptyset$, $N(\emptyset) = \lim_{n \rightarrow \infty} N(A_n)$. Далее будем считать, что $N(\emptyset) = 0$ (условие нормировки), и, следовательно, для любого $A \in \mathcal{A}$ $\min(N(A), N(X \setminus A)) = N(A \cap (X \setminus A)) = N(\emptyset) = 0$.

Доказательство. аналогично доказательству теоремы 3.1.2. ■
Любую функцию $N(A)$, $A \in \mathcal{A}$, обладающую свойствами, отмеченными в теореме 3.1.4, будем называть необходимостью.

Тройку $(X, \mathcal{A}, N)$ назовем пространством с необходимостью.

Пример 3.1.4. Пусть $\tilde{\mathcal{L}}(X) = \overline{\mathcal{L}}(X)$ — класс всех функций $X \rightarrow [0, 1]$, $\mathcal{A} = \mathcal{P}(X)$ — класс всех подмножеств X . Тогда функция $n_h(\cdot) : \tilde{\mathcal{L}}(X) \rightarrow [0, 1]$,

$$n_h(f(\cdot)) = \inf_{x \in X} \max(f(x), h(x)), \quad f(\cdot) \in \tilde{\mathcal{L}}(X) = \overline{\mathcal{L}}(X), \quad (1.22)$$

где $h(\cdot)$ — произвольная фиксированная функция $X \rightarrow [0, 1]$, удовлетворяет условиям определения 3.1.3 и условию нормировки, принятому в теореме 3.1.4, т.е. является интегралом, если $\inf_{x \in X} h(x) = 0$; $n_h(1(\cdot)) = 1$. Соответственно необходимость

$$N_h(A) = \inf \max(\chi_A(x), h(x)) = \begin{cases} \inf_{x \in X \setminus A} h(x), & \text{если } A \neq X, \\ 1, & \text{если } A = X, \end{cases} \quad A \in \mathcal{P}(X), \quad (1.23)$$

если определить $\inf_{x \in \emptyset} h(x) = 1$. Согласно (1.23) $h(x) = N_h(X \setminus \{x\})$, $x \in X$; функцию $h(\cdot)$ естественно назвать распределением $N_h(\cdot)$.

В самом деле, пусть заданы пространство с необходимостью $(Y, \mathcal{P}(Y), N_Y)$, его образ $(X, \mathcal{P}(X), N_X)$ при отображении $q(\cdot) : Y \rightarrow X$, где

$$N_X(A) \stackrel{\text{def}}{=} N_Y(\{y \in Y, q(y) \in A\}) \equiv N_Y(q^{-1}(A)), \quad A \in \mathcal{P}(X), \quad (1.24)$$

нечеткие элементы η и ξ , $\xi = q(\eta)$, канонические для $(Y, \mathcal{P}(Y), N_Y)$ и, соответственно, для $(X, \mathcal{P}(X), N_X)$, их распределения

$$h^\eta(y) \stackrel{\text{def}}{=} N_Y(Y \setminus \{y\}), \quad y \in Y, \quad h^\xi(x) \stackrel{\text{def}}{=} N_X(X \setminus \{x\}), \quad x \in X, \quad (1.25)$$

т. е. — необходимости неравенств $\eta \neq y$ и $\xi \neq x$. Поскольку согласно (1.25)

$$\begin{aligned} N^\eta(\eta \in B) &\stackrel{\text{def}}{=} N_Y(B) = N_Y\left(\bigcup_{y \in B} \{y\}\right) = N_Y\left(\bigcap_{y \in Y \setminus B} (Y \setminus \{y\})\right) = \\ &= \inf_{y \in Y \setminus B} h^\eta(y), \quad B \in \mathcal{P}(Y), \\ N^\xi(\xi \in A) &\stackrel{\text{def}}{=} N_X(A) = \inf_{x \in X \setminus A} h^\xi(x), \quad a \in \mathcal{P}(X), \end{aligned} \quad (1.26)$$

то $h^\eta(\cdot)$ и $h^\xi(\cdot)$ суть распределения $N^\eta(\cdot)$ и $N^\xi(\cdot)$.

Если A^η — нечеткое множество со значениями в $\mathcal{P}(X)$, образ η при многозначном отображении $A^\cdot : Y \rightarrow \mathcal{P}(X)$, то согласно (1.26)

$$\begin{aligned} h^A(x) &\stackrel{\text{def}}{=} N^\eta(x \in A^\eta) \stackrel{\text{def}}{=} N_Y(\{y \in Y, x \in A^y\}) = \\ &= N^\eta(\eta \in A_x) = \inf_{y \in Y \setminus A_x} h^\eta(y), \quad x \in X, \end{aligned} \quad (1.27)$$

— и. ф. о. п. A^η , значение $h^A(x)$ — необходимость покрытия $x \in X$ A^η .

Наконец, если $\xi = q(\eta)$ и A^η независимы, см. § 3.1.6.6, то

$$N^\eta(q(\eta) \in A^\eta) = \inf_{x \in X} \max(h^A(x), h^\xi(x)) = n_{h^\xi}(h^A(\cdot)) \quad (1.28)$$

— необходимость покрытия нечеткого элемента $\xi = q(\eta)$ нечетким множеством A^η , где $h^\xi(x) = N^\eta(q(\eta) \neq x)$, $h^A(x) = N^\eta(x \in A^\eta)$, $x \in X$, см. (1.22).

Пример 3.1.5. Если $X = \{x_1, x_2, \dots\}$, то для любого $A \subset X$ согласно (1.25), (1.26) $N(A) = \inf_{i: x_i \in X \setminus A} N(X \setminus \{x_i\}) \stackrel{\text{def}}{=} \inf_{i: x_i \in X \setminus A} h(x_i)$, а так как для любой функции $f(\cdot) : X \rightarrow [0, 1]$ $f(x) = \inf_i \max(f(x_i), \chi_{X \setminus \{x_i\}}(x))$, $x \in X$, то согласно определению 3.1.3

$$\begin{aligned} n(f(\cdot)) &= \inf_i \max(f(x_i), n(\chi_{X \setminus \{x_i\}}(\cdot))) = \inf_i \max(f(x_i), N(X \setminus \{x_i\})) = \\ &= \inf_i \max(f(x_i), h(x_i)) \end{aligned} \quad (1.29)$$

— общий вид интеграла $n(\cdot) : \tilde{\mathcal{L}}(X) \rightarrow [0, 1]$ в этом случае.

На самом деле $n_h(\cdot)$ в (1.22) и $N_h(\cdot)$ в (1.23) суть общие выражения для интеграла $n(\cdot)$ и меры $N(\cdot)$, продолженных на класс $\tilde{\mathcal{L}}(X)$ всех функций $X \rightarrow [0, 1]$ и — на класс $\mathcal{P}(X)$ всех подмножеств X , соответственно, [13].

Замечание 3.1.2. В общем случае нечеткая модель определяется как нечеткое пространство $(X, \mathcal{A}, P, N)$ с двумя мерами, связь между которыми определяется свойствами моделируемого объекта или явления. Если объект *стохастический*, то P и N *дуально согласованы* (дуальны), т. е. $\exists \vartheta(\cdot) \in \Theta \quad \forall A \in \mathcal{A} \quad N(A) = \vartheta(P(X \setminus A))$, где Θ — класс функций $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1]$, определенный в § 3.1.1, [13]. Если преобразование $\vartheta(\cdot)$ интерпретировать как нечеткое отрицание, то

"необходимость A " = "невозможность не A ", в частности, событие A необходимо, если противоположное невозможно. Дуальность P и N обозначим как $P < \approx > N$.

3.1.5. Интегрирование по возможности и по необходимости.

Пример $p_g(f(\cdot)) = \sup_{x \in X} \min(f(x), g(x))$ (1.12) одного из вариантов интеграла $p(\cdot)$, определенного в § 3.1.1 как линейная счетно-аддитивная функция $\mathcal{L}(X) \rightarrow \mathcal{L}$, естественно интерпретировать как интеграл функции $f(\cdot) \in \mathcal{L}(X)$ по возможности $P(\cdot)$, заданной распределением $g(\cdot) \in \mathcal{L}(X)$.

В § 3.1.5.1 рассмотрена схема Лебега построения интеграла по возможности $P(\cdot) : \mathcal{A} \rightarrow [0, 1]$, определенного на классе $\mathcal{L}(X)$ $\mathcal{A}$ -измеримых функций $f(\cdot) : X \rightarrow [0, 1]$. Показано, что интеграл $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$ является интегралом Лебега по возможности $P(\cdot)$, определенной в § 3.1.2 равенством $P(A) = p(\chi_A(\cdot))$, $A \in \mathcal{A}$.

В § 3.1.5.2 рассмотрено интегрирование по необходимости $N(\cdot)$.

3.1.5.1 Схема Лебега интегрирования по возможности; $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$ как интеграл Лебега. Заметим, что всякую $\mathcal{A}$ -измеримую функцию $f(\cdot)$ можно представить в виде предела равномерно сходящейся последовательности кусочно-постоянных $\mathcal{A}$ -измеримых функций. А именно, последовательности

$$\underline{f}_n(x) = \bigoplus_{k=1}^n (a_k^{(n)} \bullet \chi_{A_k^{(n)}}(x)), \quad \bar{f}_n(x) = \bigoplus_{k=1}^n (a_{k-1}^{(n)} \bullet \chi_{A_k^{(n)}}(x)),$$

$$\underline{f}_n(x) \leq f(x) \leq \bar{f}_n(x), \quad n = 1, 2, \dots, \quad x \in X, \quad (1.30)$$

равномерно сходятся к $f(x)$, $x \in X$, если $A_1^{(n)} = \{x \in X, a_1^{(n)} \leq f(x) \leq a_0^{(n)}\}$, $A_k^{(n)} = \{x \in X, a_k^{(n)} \leq f(x) < a_{k-1}^{(n)}\}$, $k = 2, 3, \dots, n$, $1 = a_0^{(n)} > a_1^{(n)} > \dots > a_n^{(n)} = 0$, и $\varepsilon^{(n)} = \max_{1 \leq k \leq n} (a_{k-1}^{(n)} - a_k^{(n)}) \rightarrow 0$ при $n \rightarrow \infty$, поскольку $\sup_{x \in X} (f(x) - \underline{f}_n(x)) \leq \varepsilon^{(n)}$, $\sup_{x \in X} (\bar{f}_n(x) - f(x)) \leq \varepsilon^{(n)}$.

Согласно равенствам (1.30) для $f(\cdot) \in \mathcal{L}(X)$ и определению 3.1.2 вследствие линейности и монотонности $p(\cdot)$

$$\begin{aligned} \bigoplus_{k=1}^n (a_k^{(n)} \bullet P(A_k^{(n)})) &= p(\underline{f}_n(\cdot)) \leq p(f(\cdot)) \leq p(\bar{f}_n(\cdot)) = \\ &= \bigoplus_{k=1}^n (a_{k-1}^{(n)} \bullet P(A_k^{(n)})), \quad n = 1, 2, \dots, \end{aligned} \quad (1.31)$$

где выражения слева и справа назовем *лебеговскими Р-интегральными суммами функции $f(\cdot) \in \mathcal{L}(X)$* .

Определение 3.1.4. Функцию $f(\cdot) \in \mathcal{L}(X)$ назовем *Р-интегрируемой на X по Лебегу*, если для любой последовательности разбиений $X = \bigcup_{k=1}^n A_k^{(n)}, n = 1, 2, \dots$, порожденной последователь-

ностью $1 = a_0^{(n)} > a_1^{(n)} > \dots > a_n^{(n)} = 0$, $n = 1, 2, \dots$, такой, что $\varepsilon^{(n)} = \max_{1 \leq k \leq n} (a_{k-1}^{(n)} - a_k^{(n)}) \rightarrow 0$ при $n \rightarrow \infty$, левая и правая части (1.31) стремятся к одному пределу. Этот предел назовем *интегралом Лебега функции* $f(\cdot) \in \mathcal{L}(X)$ по возможности $P(\cdot)$ (Р-интегралом Лебега); согласно (1.31) последний, разумеется, равен $p(f(\cdot))$.

Теорема 3.1.5. *Любая функция $f(\cdot) \in \mathcal{L}(X)$ Р-интегрируема по Лебегу.*

Доказательство. Действительно,

$$\begin{aligned} 0 &\leq \max_{1 \leq k \leq n} \min(a_{k-1}^{(n)}, P(A_k^{(n)})) - \max_{1 \leq k \leq n} \min(a_k^{(n)}, P(A_k^{(n)})) \leq \\ &\leq \max_{1 \leq k \leq n} \min(a_{k-1}^{(n)} - a_k^{(n)}, P(A_k^{(n)})) \leq \max_{1 \leq k \leq n} (a_{k-1}^{(n)} - a_k^{(n)}) = \varepsilon^{(n)} \rightarrow 0 \end{aligned} \quad (1.32)$$

при $n \rightarrow \infty$. Дело в том, что для любых $x, y, z \geq 0$ $\min(x + y, z) \leq \min(x, z) + \min(y, z)$, поэтому $\min(a_{k-1}^{(n)}, P(A_k^{(n)})) \leq \min(a_k^{(n)}, P(A_k^{(n)})) + \min(a_{k-1}^{(n)} - a_k^{(n)}, P(A_k^{(n)}))$ и тем более $\max_{1 \leq k \leq n} \min(a_{k-1}^{(n)}, P(A_k^{(n)})) \leq \max_{1 \leq k \leq n} \min(a_k^{(n)}, P(A_k^{(n)})) + \max_{1 \leq k \leq n} \min(a_{k-1}^{(n)} - a_k^{(n)}, P(A_k^{(n)}))$, что и записано в (1.32); в формуле (1.32) и следующих за (1.32) «+» и «-» символизируют обычные операции сложения и вычитания. ■

Таким образом, между функциями $p(f(\cdot))$, $f(\cdot) \in \mathcal{L}(X)$, и $P(A)$, $A \in \mathcal{A}$, установлено взаимно однозначное соответствие: $p(f(\cdot))$ есть интеграл Лебега функции $f(\cdot)$ по $P(\cdot)$, $P(A) = p(\chi_A(\cdot))$, $\chi_A(\cdot)$ — и. ф. $A \in \mathcal{A}$.

3.1.5.2 Схема Лебега интегрирования по необходимости.

Определим теперь понятие интеграла Лебега по необходимости $N(\cdot)$. Для любой функции $f(\cdot) \in \tilde{\mathcal{L}}(X)$ последовательно-

$$f \underset{\sim n}{(x)} = \bigoplus_{k=1}^n (b_{k-1}^{(n)} \tilde{\bullet} \chi_{B_k^{(n)}}(x)) \stackrel{\text{def}}{=} \min_{1 \leq k \leq n} \max(b_{k-1}^{(n)}, \chi_{B_k^{(n)}}(x)),$$

$$\tilde{f}_n(x) = \bigoplus_{k=1}^n (b_k^{(n)} \tilde{\bullet} \chi_{B_k^{(n)}}(x)) \stackrel{\text{def}}{=} \min_{1 \leq k \leq n} \max(b_k^{(n)}, \chi_{B_k^{(n)}}(x)),$$

$x \in X$, $n = 1, 2, \dots$, в которых $0 = b_0^{(n)} < b_1^{(n)} < \dots < b_{n-1}^{(n)} < b_n^{(n)} = 1$, $B_k^{(n)} = (x \in X, f(x) \leq b_{k-1}^{(n)}) \cup (x \in X, f(x) > b_k^{(n)})$, $k = 1, 2, \dots, n - 1$, $B_n^{(n)} = (x \in X, f(x) \leq b_{n-1}^{(n)}) \cup (x \in X, f(x) \geq b_n^{(n)})$, $n = 1, 2, \dots$,

удовлетворяют условию $f_n(x) \leq f(x) \leq \tilde{f}_n(x)$, $x \in X$, $n = 1, 2, \dots$, и при

$n \rightarrow \infty$ равномерно сходятся к $f(\cdot)$, если $\max_{1 \leq k \leq n} (b_k^{(n)} - b_{k-1}^{(n)}) = \varepsilon^{(n)} \rightarrow 0$

при $n \rightarrow \infty$. Определение интеграла по необходимости $N(\cdot)$ основано на соотношениях

$$\begin{aligned}
\widetilde{+}_{k=1}^n (b_{k-1}^{(n)} \bullet N(B_k^{(n)})) &= n(f_n(\cdot)) \leq n(f(\cdot)) \leq n(\widetilde{f}_n(\cdot)) = \\
&= \widetilde{+}_{k=1}^n (b_k^{(n)} \bullet N(B_k^{(n)})), \quad n = 1, 2, \dots, \quad (1.33)
\end{aligned}$$

вытекающих из свойств необходимости $n(\cdot)$, отмеченных в теореме 3.1.3.

Определение 3.1.5. Функцию $f(\cdot) \in \widetilde{\mathcal{L}}(X)$ назовем *N-интегрируемой по Лебегу*, если для любой последовательности разбиений $[0, 1] = \bigcup_{k=1}^n B_k^{(n)}$, $0 = b_0^{(n)} < b_1^{(n)} < \dots < b_{n-1}^{(n)} < b_n^{(n)} = 1$, $n = 1, 2, \dots$, такой, что $\varepsilon^{(n)} = \max_{1 \leq k \leq n} (b_k^{(n)} - b_{k-1}^{(n)}) \rightarrow 0$ при $n \rightarrow \infty$, левая и правая части (1.33) стремятся к одному пределу. Этот предел, равный, очевидно, $n(f(\cdot))$, назовем *N-интегралом Лебега функции $f(\cdot) \in \widetilde{\mathcal{L}}(X)$* .

Теорема 3.1.6. Любая функция $f(\cdot) \in \widetilde{\mathcal{L}}(X)$ N-интегрируема по Лебегу.

Доказательство этой теоремы не отличается от доказательства теоремы 3.1.5. ■

3.1.6. Независимость. Условные возможность и необходимость. В этом параграфе речь пойдет о теоретико-возможностных аналогах теоретико-вероятностных понятий независимости и условной вероятности.

3.1.6.1 Р-независимость. Условная возможность. Понятия условной вероятности и независимости, играющие важную роль в теории вероятностей, имеют формальные аналоги в теории возможностей. Поскольку в теории возможностей независимость событий может быть охарактеризована как в терминах их возможностей, так и в терминах их необходимостей, в этом параграфе рассмотрены три понятия независимости: по возможности (Р-независимости), по необходимости (N-независимости) и полной независимости. Альтернативные характеристики нечеткой связи событий определены как условная возможность и условная необходимость.

Определение 3.1.6. Пусть $(X, \mathcal{A}, P)$ — пространство с возможностью. События $A, B \in \mathcal{A}$ назовем Р-независимыми, если $P(A \cap B) = \min(P(A), P(B))$ ($= P(A) \bullet P(B)$ в шкале $\mathcal{L}$). События $A_i \in \mathcal{A}$, $i = 1, 2, \dots$, назовем Р-независимыми в совокупности, или взаимно Р-независимыми, если для любых $n = 2, 3, \dots$, и целых $i_1 < i_2 < \dots < i_n$ $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \min(P(A_{i_1}), \dots, P(A_{i_n}))$.

Отметим, что в случае взаимной Р-независимости событий $A_1, A_2, \dots$ возможности любых событий из σ -алгебры $\sigma(A_1, A_2, \dots)$ определяются возможностями событий $A_1, A_2, \dots$, а сам факт Р-независимости инвариантен относительно выбо-

ра шкалы $\mathcal{L}$ значений возможности; в этом случае в шкале $\mathcal{L}$ $P(\bigcup_{k=1}^n A_{i_k}) = \bigoplus_{k=1}^n P(A_{i_k})$, $P(\bigcap_{k=1}^n A_{i_k}) = \bigodot_{k=1}^n P(A_{i_k})$.

Замечание 3.1.3. Из попарной P -независимости событий $A_1, A_2, \dots$ не следует их взаимная P -независимость. Действительно, пусть $P(A) = \sup_{x \in A} f(x)$, $A \in \mathcal{A}$, $f(x_1) = f(x_2) = f(x_3) = 1$, $A_1 = \{x_1\} \cup \{x_2\}$, $A_2 = \{x_2\} \cup \{x_3\}$, $A_3 = \{x_3\} \cup \{x_1\}$. Тогда $P(A_1) = P(A_2) = P(A_3) = 1$, $P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = 1$, т.е. события A_1, A_2, A_3 попарно P -независимы, но $P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 < \min(P(A_1), P(A_2), P(A_3))$.

С другой стороны, если $P(A_1 \cap A_2 \cap A_3) = \min(P(A_1), P(A_2), P(A_3))$, то это еще не означает, что A_1, A_2 и A_3 P -независимы в совокупности. Действительно, пусть $P(\{x_1\}) = P(\{x_2\}) = 1$, $P(\{x_3\}) = 1/2$, $A_1 = \{x_1\} \cup \{x_2\}$, $A_2 = \{x_2\}$, $A_3 = \{x_3\} \cup \{x_2\}$. Тогда $P(A_1) = P(A_3) = 1$, $P(A_2) = 1/2$, $P(A_1 \cap A_2 \cap A_3) = 1/2 = \min(P(A_1), P(A_2), P(A_3))$, но $P(A_1 \cap A_3) = 1/2 < \min(P(A_1), P(A_3)) = 1$.

Определение 3.1.7. Вариантом условной относительно события $C \in \mathcal{A}$ возможности события $A \in \mathcal{A}$ назовем любое решение $P(A|C)$ уравнения

$$P(A|C) \bullet P(C) \stackrel{\text{def}}{=} \min(P(A|C), P(C)) = P(A \cap C), \quad A, C \in \mathcal{A}. \quad (1.34)$$

Любые два варианта $P(A|C)$ и $P'(A|C)$ назовем эквивалентными; эквивалентность будем отмечать знаком равенства: $P(A|C) = P'(A|C)$.

Условной относительно события $C \in \mathcal{A}$ возможностью события $A \in \mathcal{A}$ назовем класс (эквивалентности) всех ее вариантов; для условной возможности будем также использовать обозначение $P(A|C)$.

Каждый вариант условной относительно C возможности может быть определен формулой

$$P_q(A|C) = \begin{cases} P(A \cap C), & \text{если } P(A \cap C) < P(C), \\ q = q(A, C), & \text{если } P(A \cap C) = P(C), \end{cases} \quad A \in \mathcal{A}, \quad (1.35)$$

в которой согласно (1.34) формально $1 \geq q(A, C) \geq P(C)$, $A \in \mathcal{A}$.

Вообще говоря, среди вариантов условной возможности (1.35) может и не быть возможности, т.е. аддитивной: $P_q(A \cup B|C) = \max(P_q(A|C), P_q(B|C))$, $A, B \in \mathcal{A}$,

счетно-аддитивной: $P_q(\bigcup_{j=1}^{\infty} A_j|C) = \sup_{1 \leq j \leq \infty} P_q(A_j|C)$ и нормированной: $P_q(X|C) = 1$ функции $P_q(\cdot|C) : \mathcal{A} \rightarrow [0, 1]$.

Лемма 3.1.2. ([13]) Пусть $0 < P(C) < 1$. Тогда

1. любой вариант (1.35) условной относительно C возможности аддитивен, причем счетно-аддитивен, но не нормирован, если $q(A, C) = P(C)$, и нормирован, но, вообще говоря, не счетно-аддитивен, если, например, $q(A, C) = 1$;

2. если $C \in \mathcal{A}$ таково, что существует последовательность $B_j \subset C$, $j = 1, 2, \dots$, такая, что $P(B_j) < P(C)$, $j = 1, 2, \dots$, и $\bigcup_{j=1}^{\infty} B_j = C$,

то любой вариант (1.35) аддитивен, но либо счетно-аддитивен и не нормирован, либо не счетно-аддитивен, т. е. любой вариант $P_q(\cdot | C)$ не есть возможность;

3. если $C \in \mathcal{A}$ таково, что $P(A \cap C) = P(C)$ для любого $A \in \mathcal{A}$, такого что $A \cap C \neq \emptyset$, то любой вариант (1.35) определяется равенством $P_q(A | C) = \begin{cases} 0, & \text{если } A \cap C = \emptyset, \\ q \geq P(C), & \text{если } A \cap C \neq \emptyset, \end{cases}$ аддитивен, счетно-аддитивен, а при $q = 1$ и нормирован.

4. Наконец, если $P(C) = 1$, то $P(A | C) = P(A \cap C)$ — единственное решение (1.34), аддитивное, счетно-аддитивное и нормированное, а если $P(C) = 0$, то этим свойством обладает $P(A | C) = P(A)$, $A \in \mathcal{A}$.

«Проблемы условной возможности», отмеченные в лемме 3.1.2 при $P(C) < 1$, обусловлены тем, что в (1.35) условная возможность $P(\cdot | C)$ определена в шкале значений возможности $P(\cdot)$, в которой $C \in \mathcal{A}$, вообще говоря, не достоверное событие и, следовательно, условие нормировки $P_q(X | C) = 1$ в шкале значений $P(\cdot)$ должно быть заменено на $P_q(X | C) = P(C)$, см. пункт 4. леммы 3.1.2. «Учет условия C », эквивалентный «переходу» в шкалу значений возможности, в которой возможность C равна единице, рассмотрен в § 3.1.6.2.

Замечание 3.1.4. Если в (1.35) $P(A \cap C) = P(C)$, то A и C P -независимы, ибо $P(A \cap C) \leq \min(P(A), P(C)) \leq P(C) = P(A \cap C)$ и, следовательно, $P(A \cap C) = \min(P(A), P(C))$. Поэтому кажется естественным вместо (1.35) считать, что

$$\tilde{P}(A | C) = \begin{cases} P(A \cap C), & \text{если } P(A \cap C) < P(C), \\ P(A), & \text{если } P(A \cap C) = P(C), \end{cases} \quad A \in \mathcal{A},$$

($P(A) \geq P(A \cap C) = P(C)$!) Покажем, что если $0 < P(C) < 1$, то при таком определении вариант $\tilde{P}(\cdot | C) : \mathcal{A} \rightarrow [0, 1]$ не аддитивен. Заметим для этого, что существуют $A, B \in \mathcal{A}$ такие, что $P(A) < P(B)$, но $P(A \cap C) = P(C) > P(B \cap C)$, например, для $A = C$, $B = X \setminus C$, $P(A) = P(C) < P(\Omega \setminus C) = P(B) = 1$, $P(A \cap C) = P(C) > P(B \cap C) = P(\emptyset) = 0$. Для таких A и B $\tilde{P}(A | C) = P(A) \geq P(C)$, $\tilde{P}(B | C) = P(B \cap C) < P(C)$, $\max(\tilde{P}(A | C), \tilde{P}(B | C)) = P(A)$, но $\tilde{P}(A \cup B | C) = P(A \cup B) = \max(P(A), P(B)) = P(B)$, т. е. $\tilde{P}(A \cup B | C) = P(B) > \max(\tilde{P}(A | C), \tilde{P}(B | C)) = P(A)$.

В связи с понятием условной возможности естественно рассмотреть другое определение независимости.

Определение 3.1.7*. Событие A назовем P -независимым* от B , если существует вариант условной возможности $P(A|B) = P(A)$;

события A и B назовем P -независимыми*, если A P -независимо* от B и B P -независимо* от A .

Лемма 3.1.3. *Следующие утверждения эквивалентны: 1. A P -независимо* от B , 2. A и B P -независимы*, 3. A и B P -независимы.*

Доказательство. Действительно, пусть A P -независимо* от B . Тогда $P(A \cap B) = \min(P(A|B), P(B)) = \min(P(A), P(B))$, т.е. A и B P -независимы ($1 \Rightarrow 3$). Если же $P(A \cap B) = \min(P(A), P(B))$, то существуют как вариант $P(A|B) = P(A)$, так и вариант $P(B|A) = P(B)$, ибо в этом случае $\min(P(A|B), P(B)) = \min(P(A), P(B)) = \min(P(B|A), P(A))$, т.е. A и B P -независимы* ($3 \Rightarrow 2$). Импликация $2 \Rightarrow 1$ очевидна. ■

В заключение приведем аналог формулы полной вероятности — формулу полной возможности, ср. с (1.4.7). Пусть $X = \bigcup_{n=1}^{\infty} B_n$, тогда

$$\begin{aligned} P(A) &= P(A \cap X) = P\left(\bigcup_{n=1}^{\infty} (A \cap B_n)\right) = \bigoplus_{n=1}^{\infty} P(A \cap B_n) = \\ &= \bigoplus_{n=1}^{\infty} (\min(P(A|B_n), P(B_n))) = \bigoplus_{n=1}^{\infty} (P(A|B_n) \bullet P(B_n)), \quad A \in \mathcal{A}. \end{aligned} \quad (1.36)$$

Другая точка зрения на условную возможность, которой не свойственны трудности, отмеченные в Лемме 3.1.2, представлена в следующем параграфе.

3.1.6.2 Условная возможность: альтернативный подход. Обратимся теперь к условной при условии C возможности $P(A | C)$, $A \in \mathcal{A}$, определив её как возможность $P(A \cap C)$ в преобразованной шкале, в которой истинность C абсолютна. Обозначим $\tilde{G}_C$ класс функций $\tilde{g}_C(\cdot) : [0, P(C)] \rightarrow [0, 1]$, непрерывных, строго монотонных и удовлетворяющих условиям $\tilde{g}_C(0) = 0$, $\tilde{g}_C(P(C)) = 1$. Выберем $\tilde{g}_C(\cdot) \in \tilde{G}_C$ и рассмотрим возможность $P_{\tilde{g}_C}(\cdot | C) : \mathcal{A} \rightarrow [0, 1]$, определенную равенством

$$P_{\tilde{g}_C}(A | C) = \tilde{g}_C(P(A \cap C)), \quad A \in \mathcal{A}, \quad (1.37)$$

которую назовем *вариантом* условной относительно C возможности, отвечающим функции $\tilde{g}_C(\cdot) \in \tilde{G}_C$.

Тот факт, что формула (1.37) при любой функции $\tilde{g}_C(\cdot) \in \tilde{G}_C$ определяет возможность, проверяется непосредственно. В частности, $P_{\tilde{g}_C}(C | C) = P_{\tilde{g}_C}(X | C) = 1$ и для любых $A_j \in \mathcal{A}$, $j = 1, 2, \dots$,

$$\begin{aligned} P_{\tilde{g}_C}\left(\bigcup_j A_j | C\right) &= \tilde{g}_C\left(P\left(\bigcup_j (A_j \cap C)\right)\right) = \tilde{g}_C\left(\sup_j P(A_j \cap C)\right) = \\ &= \sup_j \tilde{g}_C(P(A_j \cap C)) = \sup_j P_{\tilde{g}_C}(A_j | C). \end{aligned}$$

Функция $\tilde{g}_C(\cdot) \in \tilde{G}_C$ определяет гомоморфизм шкалы $\mathcal{L}_C \stackrel{\text{def}}{=} ([0, \text{Pr}(C)], \leq, \max, \min)$ значений меры $\text{P}(A \cap C)$, $A \in \mathcal{A}$, в шкалу $\mathcal{L} = ([0, 1], \leq, \max, \min)$ значений возможности, в данном случае — условной относительно C , поскольку, как нетрудно убедиться, $\forall a, b \in [0, \text{Pr}(C)]$

$$a \leq b \Rightarrow \tilde{g}_C(a) \leq \tilde{g}_C(b), \quad \tilde{g}_C(\max(a, b)) = \max(\tilde{g}_C(a), \tilde{g}_C(b)), \\ \tilde{g}_C(\min(a, b)) = \min(\tilde{g}_C(a), \tilde{g}_C(b)), \quad \tilde{g}_C(0) = 0, \quad \tilde{g}_C(\text{P}(C)) = 1.$$

Заметим, что, когда A и C P -независимы, $\text{P}(A \cap C) = \min(\text{P}(A), \text{P}(C)) = \text{P}(A)$, и в преобразованной шкале вариант $\text{P}_{\tilde{g}_C}(A | C) = \tilde{g}_C(\min(\text{P}(A), \text{P}(C))) = \min(\tilde{g}_C(\text{P}(A)), \tilde{g}_C(\text{P}(C))) = \tilde{g}_C(\text{P}(A))$.

3.1.6.3 *N-независимость. Условная необходимость.*

Определение 3.1.8. События $A, B \in \mathcal{A}$ назовем N -независимыми, если $\text{N}(A \cup B) = \max(\text{N}(A), \text{N}(B)) (= \text{N}(A) \bullet \text{N}(B))$.

События $A_i \in \mathcal{A}$, $i = 1, 2, \dots$ назовем *взаимно* N -независимыми, если для любых $n = 2, 3, \dots$ и целых $i_1 < i_2 < \dots < i_n$

$$\text{N}(A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_n}) = \max(\text{N}(A_{i_1}), \dots, \text{N}(A_{i_n})) \triangleq \text{N}(A_{i_1}) \bullet \dots \bullet \text{N}(A_{i_n}).$$

Для дуальных N и P N -независимость событий A и B эквивалентна P -независимости событий $X \setminus A$ и $X \setminus B$, см. определение 3.1.6.

Определение 3.1.9. Вариантом условной относительно события $B \in \mathcal{A}$ необходимости события $A \in \mathcal{A}$ назовем любое решение $\text{N}(A|B)$ уравнения $\max(\text{N}(A|B), \text{N}(B)) = \text{N}(A \cup B)$, $A, B \in \mathcal{A}$.

Любые его решения удовлетворяют условиям

$$\text{N}(A|B) = \text{N}(A \cup B), \quad \text{если } \text{N}(B) < \text{N}(A \cup B), \\ \text{N}(A|B) \leq \text{N}(A \cup B), \quad \text{если } \text{N}(B) = \text{N}(A \cup B), \quad A, B \in \mathcal{A},$$

и считаются эквивалентными; эквивалентность будем отмечать знаком равенства. Класс (эквивалентности) всех решений назовем условной необходимостью (со значениями в шкале $\mathcal{L}$ значений $\text{N}(\cdot)$) и, как любое решение, будем обозначать $\text{N}(A|B)$.

Определение 3.1.8*. Событие A назовем N -независимым* от B , если существует вариант условной необходимости $\text{N}(A|B) = \text{N}(A)$; события A и B назовем N -независимыми*, если A N -независимо* от B и B N -независимо* от A .

Имеет место дуальный аналог леммы 3.1.3.

Лемма 3.1.4. Следующие утверждения эквивалентны: 1. A N -независимо* от B . 2. A и B N -независимы*. 3. A и B N -независимы.

Доказательство См. доказательство леммы 3.1.3. ■

Среди свойств условной необходимости выделим *формулу полной необходимости*, дуальную формуле полной возможности (1.36). Пусть

$\bigcap_{n=1}^{\infty} A_j = \emptyset$ и A — любое событие. Поскольку $A = A \cup (\bigcap_{j=1}^{\infty} A_j) =$
 $= \bigcap_{j=1}^{\infty} (A \cup A_j)$, то согласно определениям операций сложения и умноже-

ния в шкале $\tilde{\mathcal{L}}$, $N(A) = \inf_j N(A \cup A_j) = \bigoplus_{j=1}^{\infty} (N(A|A_j) \tilde{\bullet} N(A_j))$, $A \in \mathcal{A}$.

Определение 3.1.10. События A и B назовем P , N -независимыми, если они P - и N -независимы.

Для дуальных N и P , см. условия (0.11), события A и B независимы, если и только если как A и B , так и $X \setminus A$ и $X \setminus B$ P -независимы (или N -независимы), вполне независимы, если P -независимы A и B , A и $X \setminus B$, B и $X \setminus A$, $X \setminus A$ и $X \setminus B$.

3.1.6.4 Независимость нечетких элементов. Условное распределение. Пусть η — нечеткий элемент, канонический для $(Y, \mathcal{B}, P_Y)$, нечеткие элементы $\xi_1 = q_1(\eta), \dots, \xi_n = q_n(\eta)$ принимают значения соответственно в $(X_1, \mathcal{A}_1), \dots, (X_n, \mathcal{A}_n)$, где $q_1(\cdot) : Y \rightarrow X_1, \dots, q_n(\cdot) : Y \rightarrow X_n$ суть $\mathcal{B}, \mathcal{A}_1, \dots, \mathcal{B}, \mathcal{A}_n$ -измеримые функции¹⁾. Взаимная P_Y -независимость $\xi_1, \dots, \xi_n$ определяется следующим образом.

Определение 3.1.11. Нечеткие элементы $\xi_1, \dots, \xi_n$, определенные на $(Y, \mathcal{B}, P_Y)$ и принимающие значения соответственно в $(X_1, \mathcal{A}_1), \dots, \dots, (X_n, \mathcal{A}_n)$, называются взаимно P_Y -независимыми, если для любых множеств $A_j \in \mathcal{A}_j$, $j = 1, \dots, n$,

$$P_Y(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \min(P_Y(\xi_1 \in A_1), \dots, \\ P_Y(\xi_n \in A_n)) \stackrel{\text{def}}{=} \bigbullet_{i=1}^n P_Y(\xi_i \in A_i). \quad (1.38)$$

Если $\mathcal{B} = \mathcal{P}(Y)$, $\mathcal{A}_j = \mathcal{P}(X_j)$, $j = 1, \dots, n$, то определение 3.1.11 эквивалентно следующему определению.

Определение 3.1.12. Нечеткие элементы $\xi_i = q_i(\eta)$, $i = 1, \dots, n$, со значениями соответственно в X_i , $i = 1, \dots, n$, назовем взаимно P_Y -независимыми, или — P_Y -независимыми в совокупности, если для любых $x_i \in X_i$, $i = 1, \dots, n$, возможность системы равенств $\xi_i = x_i$, $i = 1, \dots, n$,

$$g^{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \min_{1 \leq i \leq n} g^{\xi_i}(x_i) \stackrel{\text{def}}{=} \bigbullet_{i=1}^n g^{\xi_i}(x_i), \quad (1.39)$$

где $g^{\xi_i}(x_i) = \sup\{g^{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) | x_1 \in X_1, \dots, x_{i-1} \in X_{i-1}, x_{i+1} \in X_{i+1}, \dots, x_n \in X_n\}$, $x_i \in X_i$, — маргинальное распределение ξ_i , $i = 1, \dots, n$.

Действительно, в таком случае возможности в левой и правой частях равенства (1.38) могут быть заданы распределениями, и если равенство (1.39) выполнено, то $P_Y(\xi_1 \in A_1, \dots, \xi_n \in A_n) =$

¹⁾ Функция $q_j(\cdot) : Y \rightarrow X_j$ $\mathcal{B}, \mathcal{A}_j$ -измерима, т. е. $\forall A_j \in \mathcal{A}_j \quad q_j^{-1}(A_j) \in \mathcal{B}$, $j = 1, \dots, n$.

$= \sup\{\min_{1 \leq i \leq n} g^{\xi_i}(x_i) \mid x_1 \in A_1, \dots, x_n \in A_n\} = \min_{1 \leq i \leq n} \sup_{x_i \in A_i} g^{\xi_i}(x_i) =$
 $= \min_{1 \leq i \leq n} P_Y(\xi_i \in A_i)$, т. е. выполнено условие (1.38). Наоборот, если для любых $A_j \in \mathcal{P}(X_j)$ выполнено условие (1.38), то для произвольных $x_i \in X_i$, $i = 1, \dots, n$, выбрав в (1.38) $A_i = \{x_i\}$, $i = 1, \dots, n$, получим (1.39).

Замечание 3.1.5. Если считать, что меры $P_Y(\cdot)$ и $N_Y(\cdot)$ дуально согласованные (дуальные), $P_Y(\cdot) \approx N_Y(\cdot)$, см. условие (0.11), то определение 3.1.11 независимости $\xi_1, \xi_2, \dots, \xi_n$ означает, что $\xi_1, \xi_2, \dots, \xi_n$ P_Y -, N_Y - и вполне независимы. Действительно, N_Y -независимость $\xi_1, \xi_2, \dots, \xi_n$ согласно определению 3.1.11 означала бы, что для любых $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$

$$N_Y\left(\bigcup_{i=1}^n \{y \in Y, q_i(y) \in A_i\}\right) = \max_{1 \leq i \leq n} N_Y(\{y \in Y, q_i(y) \in A_i\}), \quad (1.40)$$

что в силу (0.11) эквивалентно условию: для любых $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$ $P_Y\left(\bigcap_{i=1}^n \{y \in Y, q_i(y) \in X_i \setminus A_i\}\right) = \min_{1 \leq i \leq n} P_Y(\{y \in Y, q_i(y) \in X_i \setminus A_i\})$.

Отметим также, что для взаимно независимых $\xi_1, \dots, \xi_n$ $P_Y\left(\bigcup_{i=1}^n \{y \in Y, q_i(y) \in A_i\}\right) = \bigoplus_{i=1}^n P_Y(\{y \in Y, q_i(y) \in A_i\})$, $P_Y\left(\bigcap_{i=1}^n \{y \in Y, q_i(y) \in A_i\}\right) = \bigodot_{i=1}^n P_Y(\{y \in Y, q_i(y) \in A_i\})$, т. е. возможности событий в левых частях этих равенств определяются возможностями событий $\{y \in Y, q_i(y) \in A_i\}$, $i = 1, \dots, n$.

Если же $P_Y(\cdot)$ и $N_Y(\cdot)$ не дуальны, т. е. если $\forall \vartheta(\cdot) \in \Theta \exists A \in \mathcal{P}(Y)$ $N_Y(A) \neq \vartheta(P_Y(Y \setminus A))$, то условие (1.40) взаимной N_Y -независимости с условием (1.38) P_Y -независимости непосредственно не связано. Далее для упрощения формулировок будем считать, что $P_Y(\cdot) \approx N_Y(\cdot)$, а P_Y -независимость будем называть независимостью.

Читателю предлагается проверить, что, как и в теории вероятностей, см. теорему 8.1 гл. 1, имеют место следующие факты.

Лемма 3.1.5. Пусть $z_i(\cdot) : X_i \rightarrow Z_i$, $i = 1, \dots, n$, — произвольные функции, и нечеткие элементы ξ_i , $i = 1, \dots, n$, взаимно независимы. Тогда взаимно независимы и нечеткие элементы $\zeta_i = z_i(\xi_i)$, $i = 1, \dots, n$, причем для любых $z_i \in Z_i$, $i = 1, \dots, n$, $g^{\zeta_1, \dots, \zeta_n}(z_1, \dots, z_n) = \sup\{\min_{1 \leq j \leq n} g^{\xi_j}(x_j) \mid x_i \in X_i, z_i(x_i) = z_i, i = 1, \dots, n\} = \min_{1 \leq i \leq n} \sup\{g^{\xi_i}(x_i) \mid x_i \in X_i, z_i(x_i) = z_i\} = \min_{1 \leq i \leq n} g^{\zeta_i}(z_i)$, где $g^{\zeta_i}(z_i) = \sup\{g^{\xi_i}(x_i) \mid x_i \in X_i, z_i(x_i) = z_i\}$, $z_i \in Z_i$, $i = 1, \dots, n$.

В частности, если $Y = Y_1 \times \dots \times Y_n$, $y = (y_1, \dots, y_n)$, $\eta = (\eta_1, \dots, \eta_n)$ и

$$g^\eta(y) = \min_{1 \leq i \leq n} g_i^\eta(y_i), \quad y_i \in Y_i, \quad i = 1, \dots, n, \quad (1.41)$$

то при любых функциях $q_i(\cdot) : Y_i \rightarrow X_i$, $i = 1, \dots, n$, нечеткие элементы $\xi_i = q_i(\eta_i)$, $i = 1, \dots, n$, взаимно независимы, причем для любых $x_i \in X_i$, $i = 1, \dots, n$,

$$g^{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \min_{1 \leq i \leq n} g^{\xi_i}(x_i), \quad (1.42)$$

где

$$g^{\xi_i}(x_i) = \sup\{g^{\eta_i}(y_i) \mid y_i \in Y_i, q_i(y_i) = x_i\}, \quad x_i \in X_i, \quad i = 1, \dots, n. \quad (1.43)$$

В дальнейшем, как правило, используется модель независимости (1.41–1.43), верная при любых функциях $q_i(\cdot) : Y_i \rightarrow X_i$, $i = 1, \dots, n$.

Определение 3.1.13. Вариантом условного распределения нечетких элементов $\xi_1, \dots, \xi_k$ при условии $\xi_{k+1} = x_{k+1}, \dots, \xi_n = x_n$ называется любое решение $g^{\xi_1, \dots, \xi_k | \xi_{k+1}, \dots, \xi_n}(x_1, \dots, x_k | x_{k+1}, \dots, x_n)$ уравнения

$$\begin{aligned} \min(g^{\xi_1, \dots, \xi_k | \xi_{k+1}, \dots, \xi_n}(x_1, \dots, x_k | x_{k+1}, \dots, x_n), g^{\xi_{k+1}, \dots, \xi_n}(x_{k+1}, \dots, x_n)) = \\ = g^{\xi_1, \dots, \xi_n}(x_1, \dots, x_n), \quad x_1 \in X_1, \dots, x_n \in X_n, \end{aligned} \quad (1.44)$$

где

$$g^{\xi_{k+1}, \dots, \xi_n}(x_{k+1}, \dots, x_n) = \sup_{\substack{x_i \in X_i, \\ i=1, \dots, k}} g^{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \quad (1.45)$$

— маргинальное распределение $\xi_{k+1}, \dots, \xi_n$.

Для упрощения формулировок обозначим $(\xi_1, \dots, \xi_k) \sim \zeta_1$, $(\xi_{k+1}, \dots, \xi_n) \sim \zeta_2$, $X_1 \times \dots \times X_k \sim Z_1$, $X_{k+1} \times \dots \times X_n \sim Z_2$, $(x_1, \dots, x_k) \sim z_1$, $(x_{k+1}, \dots, x_n) \sim z_2$ и рассмотрим уравнение (1.44) при условии (1.45), переписав их в виде

$$\min(g^{\zeta_1 | \zeta_2}(z_1 | z_2), g^{\zeta_2}(z_2)) = g^{\zeta_1, \zeta_2}(z_1, z_2), \quad z_1 \in Z_1, \quad z_2 \in Z_2, \quad (1.46)$$

где

$$g^{\zeta_2}(z_2) = \sup_{z_1 \in Z_1} g^{\zeta_1, \zeta_2}(z_1, z_2). \quad (1.47)$$

Поскольку в (1.46), (1.47) $g^{\zeta_2}(z_2) \geq g^{\zeta_1, \zeta_2}(z_1, z_2)$, $z_1 \in Z_1$, $z_2 \in Z_2$, уравнение (1.46) всегда имеет решение. Любой вариант условного распределения ζ_1 при условии $\zeta_2 = z_2$ можно определить равенством

$$g^{\zeta_1 | \zeta_2}(z_1 | z_2) = \begin{cases} g^{\zeta_1, \zeta_2}(z_1, z_2), & \text{если } g^{\zeta_1, \zeta_2}(z_1, z_2) < g^{\zeta_2}(z_2), \\ q(g^{\zeta_2}(z_2)), & \text{если } g^{\zeta_1, \zeta_2}(z_1, z_2) = g^{\zeta_2}(z_2), \end{cases} \quad z_1 \in Z_1, \quad z_2 \in Z_2, \quad (1.48)$$

где $q(\cdot) : [0, 1] \rightarrow [0, 1]$ — произвольная функция, удовлетворяющая условию $q(a) \geq a$, $a \in [0, 1]$.

Заметим, что при некоторых $z_2 \in Z_2$ среди вариантов условного распределения $g^{\zeta_1 | \zeta_2}(\cdot | z_2)$ может и не быть распределения возможности. Действительно, если в (1.47) при $\bar{z}_2 \in Z_2$ точная верхняя грань не до-

стигается, то $g^{\zeta_1, \zeta_2}(z_1, \bar{z}_2) < g^{\zeta_2}(\bar{z}_2)$, $z_1 \in Z_1$, и, следовательно, в (1.46) (см. (1.48))

$$\min(g^{\zeta_1, \zeta_2}(z_1 | \bar{z}_2), g^{\zeta_2}(\bar{z}_2)) = g^{\zeta_1 | \zeta_2}(z_1 | \bar{z}_2) = g^{\zeta_1, \zeta_2}(z_1, \bar{z}_2). \quad (1.49)$$

Если при этом $g^{\zeta_2}(\bar{z}_2) < 1$, то согласно (1.47), (1.49) $\sup_{z_1 \in Z_1} g^{\zeta_1, \zeta_2}(z_1 | \bar{z}_2) = g^{\zeta_2}(\bar{z}_2) < 1$. ■

3.1.6.5 Переходные возможность и необходимость. Для определения условного распределения $g^{\zeta_1 | \zeta_2}(\cdot | z_2)$ при каждом $z_2 \in Z_2$ необходимо знать совместное распределение $g^{\zeta_1, \zeta_2}(z_1, z_2)$, $z_1 \in Z_1$, $z_2 \in Z_2$, нечетких элементов ζ_1 , ζ_2 . Формулами (1.46), (1.47) можно воспользоваться для определения $g^{\zeta_1, \zeta_2}(\cdot, \cdot)$, если ввести понятие переходной возможности и ее распределения¹⁾.

Определение 3.1.14. Отображение $P(\cdot | \cdot) : \mathcal{P}(Z_1) \times Z_2 \rightarrow [0, 1]$ называется *переходной возможностью на $(Z_2, \mathcal{P}(Z_2))$, $(Z_1, \mathcal{P}(Z_1))$* , если при каждом $z_2 \in Z_2$ $P(\cdot | z_2)$ есть возможность²⁾ на $(Z_1, \mathcal{P}(Z_1))$; функция $g(\cdot | \cdot) : Z_1 \times Z_2 \rightarrow [0, 1]$ называется *распределением переходной возможности* $P(\cdot | \cdot) : \mathcal{P}(Z_1) \times Z_2 \rightarrow [0, 1]$, если при каждом $z_2 \in Z_2$ $g(\cdot | z_2) : Z_1 \rightarrow [0, 1]$ — распределение возможности $P(\cdot | z_2) : \mathcal{P}(Z_1) \rightarrow [0, 1]$, т. е. если для любого $A \in \mathcal{P}(Z_1)$ $P(A | z_2) = \sup_{z_1 \in A} g(z_1 | z_2)$, и, в частности,

$$\sup_{z_1 \in Z_1} g(z_1 | z_2) = 1, \quad z_2 \in Z_2.$$

Непосредственным следствием определения 3.1.14 является

Лемма 3.1.6. Для произвольной возможности $P_{Z_2}(\cdot) : \mathcal{P}(Z_2) \rightarrow [0, 1]$ на $(Z_2, \mathcal{P}(Z_2))$, заданной распределением $g(\cdot) : Z_2 \rightarrow [0, 1]$,

$$g(z_1, z_2) = \min(g(z_1 | z_2), g(z_2)), \quad z_1 \in Z_1, \quad z_2 \in Z_2, \quad (1.50)$$

— *распределение возможности* $P_{Z_1 \times Z_2}(\cdot) : \mathcal{P}(Z_1 \times Z_2) \rightarrow [0, 1]$ на $(Z_1 \times Z_2, \mathcal{P}(Z_1 \times Z_2))$, определенной формулой

$$P_{Z_1 \times Z_2}(C) = \sup_{(z_1, z_2) \in C} g(z_1, z_2), \quad C \in \mathcal{P}(Z_1 \times Z_2), \quad (1.51)$$

и при этом в (1.50)

$$g(z_2) = \sup_{z_1 \in Z_1} g(z_1, z_2), \quad z_2 \in Z_2. \quad (1.52)$$

Доказательство. Действительно, аддитивность и счетная аддитивность

$$P_{Z_1 \times Z_2}(\cdot) \quad (\text{см.} \quad (1.51)) \quad \text{очевидны} \quad \text{и} \quad P_{Z_1 \times Z_2}(Z_1 \times Z_2) =$$

¹⁾ Распределение переходной возможности обозначим так же, как обозначено условное распределение.

²⁾ т. е. вполне аддитивная и нормированная мера на $(Z_1, \mathcal{P}(Z_1))$; если рассматривать отображение $P(\cdot | \cdot) : \mathcal{A}_1 \times Z_2 \rightarrow [0, 1]$ как переходную возможность на $(Z_2, \mathcal{A}_2)$, $(Z_1, \mathcal{A}_1)$, то $\forall z_2 \in Z_2$ $P(\cdot | z_2)$ — возможность на $(Z_1, \mathcal{A}_1)$ и $\forall A \in \mathcal{A}_1$ $P(A | \cdot) — \mathcal{A}_2$ -измеримая функция $Z_2 \rightarrow [0, 1]$.

$$= \sup_{z_2 \in Z_2} \sup_{z_1 \in Z_1} \min(g(z_1|z_2), g(z_2)) = \sup_{z_2 \in Z_2} g(z_2) = 1, \text{ ибо } \sup_{z_1 \in Z_1} \min(g(z_1|z_2), g(z_2)) = \min(\sup_{z_1 \in Z_1} g(z_1|z_2), g(z_2)) = g(z_2), z_2 \in Z_2. \blacksquare$$

Замечание 3.1.6. Пусть $g(\cdot|\cdot) : Z_1 \times Z_2 \rightarrow [0, 1]$ — распределение некоторой переходной возможности, $g(\cdot) : Z_2 \rightarrow [0, 1]$ — распределение некоторой возможности, и $g(\cdot, \cdot)$ — отвечающая им левая часть равенства (1.50). Тогда, считая, что в (1.50) $g^{\zeta_1, \zeta_2}(z_1, z_2) \stackrel{\text{def}}{=} g(z_1, z_2)$ и соответственно в силу (1.52) $g^{\zeta_2}(z_2) \stackrel{\text{def}}{=} g(z_2)$, $z_1 \in Z_1$, $z_2 \in Z_2$, и рассматривая (1.50) как уравнение относительно $g^{\zeta_1|\zeta_2}(z_1|z_2) \stackrel{\text{def}}{=} g(z_1|z_2)$, $z_1 \in Z_1$, $z_2 \in Z_2$, найдем, что в таком случае существует вариант условного распределения $g^{\zeta_1|\zeta_2}(\cdot|z_2)$, являющийся распределением возможности при каждом $z_2 \in Z_2$. Таким вариантом является, в частности, переходная возможность, рассматриваемая как функция $g(\cdot|z_2) : Z_1 \rightarrow [0, 1]$ при каждом фиксированном $z_2 \in Z_2$. Далее, говоря об условном распределении, будем иметь в виду такую его конструкцию.

В заключение — коротко о переходной необходимости.

Определение 3.1.15. Переходной необходимостью на $(Z_2, \mathcal{P}(Z_2))$, $(Z_1, \mathcal{P}(Z_1))$ назовем отображение $N(\cdot|\cdot) : \mathcal{P}(Z_1) \times Z_2 \rightarrow [0, 1]$, которое при каждом $z_2 \in Z_2$ есть необходимость на $(Z_1, \mathcal{P}(Z_1))$; функция $h(\cdot|\cdot) : Z_1 \times Z_2 \rightarrow [0, 1]$ называется *распределением* $N(\cdot|\cdot) : \mathcal{P}(Z_1) \times Z_2 \rightarrow [0, 1]$, если при каждом $z_2 \in Z_2$ $h(\cdot|z_2) : Z_1 \rightarrow [0, 1]$ — распределение $N(\cdot|z_2) : \mathcal{P}(Z_1) \rightarrow [0, 1]$, т. е. $\forall A \in \mathcal{P}(Z_1) \forall z_2 \in Z_2 N(A|z_2) = \inf_{z_1 \in Z_1 \setminus A} h(z_1|z_2)$.

Если $h(\cdot) : Z_2 \rightarrow [0, 1]$ — распределение некоторой необходимости $N_{z_2}(\cdot) : \mathcal{P}(Z_2) \rightarrow [0, 1]$, то

$$N_{Z_1 \times Z_2}(C) = \inf_{(z_1, z_2) \in (Z_1 \times Z_2) \setminus C} \max(h(z_1|z_2), h(z_2)), C \in \mathcal{P}(Z_1 \times Z_2),$$

— необходимость на $(Z_1 \times Z_2, \mathcal{P}(Z_1 \times Z_2))$,

$$h(z_1, z_2) = \max(h(z_1|z_2), h(z_2)), (z_1, z_2) \in Z_1 \times Z_2, \quad (1.53)$$

— её распределение.

Равенство (1.53), записанное для распределения необходимости неравенства $(\zeta_1, \zeta_2) \neq (z_1, z_2)$ нечеткого элемента (ζ_1, ζ_2)

$$h^{\zeta_1, \zeta_2}(z_1, z_2) = \max(h^{\zeta_1|\zeta_2}(z_1|z_2), h^{\zeta_2}(z_2)), (z_1, z_2) \in Z_1 \times Z_2, \quad (1.54)$$

при заданной левой части определяет варианты условного распределения необходимостей неравенств $\zeta_1 \neq z_1$ при условии $\zeta_2 = z_2$ как решений уравнения (1.54) относительно ¹⁾ $h^{\zeta_1, \zeta_2}(z_1|z_2)$, где $h^{\zeta_2}(z_2) = \inf_{z_1 \in Z} h^{\zeta_1, \zeta_2}(z_1, z_2)$, $(z_1, z_2) \in Z_1 \times Z_2$.

¹⁾ $(\zeta_1, \zeta_2) \neq (z_1, z_2) \Leftrightarrow \zeta_1 \neq z_1 \text{ и } \zeta_2 \neq z_2, \text{ либо } \zeta_1 = z_1 \text{ и } \zeta_2 \neq z_2, \text{ либо } \zeta_1 \neq z_1 \text{ и } \zeta_2 = z_2 \Leftrightarrow \zeta_1 \neq z_1, \text{ если } \zeta_2 = z_2, \text{ либо } \zeta_2 \neq z_2.$

3.1.6.6 Независимость нечетких множеств и нечетких элементов. Пусть $A_i : Y \rightarrow \mathcal{P}(X_i)$, $i = 1, \dots, n$, — многозначные отображения, $q_j(\cdot) : Y \rightarrow X'_j$, $j = 1, \dots, m$, — функции и $P^\eta(\cdot) \approx N^\eta(\cdot)$.

Определение 3.1.16. Нечеткие множества A_i^η , $i = 1, \dots, n$, и нечеткие элементы $\xi_j = q_j(\eta)$, $j = 1, \dots, m$, назовем

► взаимно независимыми, если для любых множеств $A_i \in \mathcal{P}(X_i)$, $i = 1, \dots, n$, и $A'_j \in \mathcal{P}(X'_j)$, $j = 1, \dots, m$,

$$\begin{aligned} P^\eta(A_1^\eta \cap A_1 = \emptyset, A_1^\eta \neq \emptyset; A_2^\eta \cap A_2 \neq \emptyset; \dots; A_n^\eta \cap A_n = \emptyset, A_n^\eta \neq \emptyset; \\ \xi_1 \in A'_1, \dots, \xi_m \in A'_m) = \min(P^\eta(A_1^\eta \cap A_1 = \emptyset, A_1^\eta \neq \emptyset), \\ P^\eta(A_2^\eta \cap A_2 \neq \emptyset), \dots, P^\eta(A_n^\eta \cap A_n = \emptyset, A_n^\eta \neq \emptyset), \\ P^\eta(\xi_1 \in A'_1), \dots, P^\eta(\xi_m \in A'_m)) \quad (1.55) \end{aligned}$$

при любых (непротиворечивых) комбинациях « $= \emptyset$ », « $\neq \emptyset$ »¹⁾;

► взаимно независимыми в смысле одноточечного покрытия, если для любых $x_i \in X_i$, $i = 1, 2, \dots, n$, и $x'_j \in X'_j$, $j = 1, \dots, m$,

$$\begin{aligned} P^\eta(x_1 \in A_1^\eta, x_2 \notin A_2^\eta, \dots, x_n \in A_n^\eta, \xi_1 = x'_1, \dots, \xi_m = x'_m) = \\ \min(g^{A_1^\eta}(x_1), g^{X_2 \setminus A_2^\eta}(x_2), \dots, g^{A_n^\eta}(x_n), g^{\xi_1}(x'_1), \dots, g^{\xi_m}(x'_m)), \quad (1.56) \end{aligned}$$

или, что в силу условия $P^\eta(\cdot) \approx N^\eta(\cdot)$ эквивалентно (1.56), $N^\eta(x_1 \notin A_1^\eta, x_2 \in A_2^\eta, \dots, x_n \notin A_n^\eta, \xi_1 \neq x'_1, \dots, \xi_m \neq x'_m) = \max(h^{X_1 \setminus A_1^\eta}(x_1), h^{A_2^\eta}(x_2), \dots, h^{X_n \setminus A_n^\eta}(x_n), h^{\xi_1}(x'_1), \dots, h^{\xi_m}(x'_m))$, где $h^{A_i^\eta}(x_i) = N^\eta(x_i \in A_i^\eta)$, $h^{\xi_j}(x'_j) = N^\eta(\xi_j \neq x'_j)$, $x_i \in X$, $i = 1, \dots, n$, $x'_j \in X'_j$, $j = 1, \dots, m$, см. (1.25), (1.27), для любых комбинаций « $\in$ » и « $\notin$ », см. замечание 3.1.5.

Ясно, что взаимно независимые нечеткие множества и нечеткие элементы взаимно независимы и в смысле одноточечного покрытия.

Теорема 3.1.7. Пусть $q(\cdot) : Y \rightarrow X$, $A : Y \rightarrow \mathcal{P}(X)$ и $\xi = q(\eta)$. Тогда $P^\eta(\xi \in A^\eta) = \sup\{g^\eta(y) \mid y \in Y_{(1)}, q(y) \in A^\eta\}$, а если нечеткий элемент $\xi = q(\eta)$ и нечеткое множество A^η независимы в смысле одноточечного покрытия, то

$$P^\eta(\xi \in A^\eta) = \sup_{x \in X} \min(g^\xi(x), g^{A^\eta}(x)) \stackrel{\text{def}}{=} p_{g^\xi}(g^{A^\eta}(\cdot)). \quad (1.57)$$

¹⁾ Заметим, что $\{y \in Y, A^y \cap B = \emptyset, A^y \neq \emptyset\} = Y \setminus \bigcup_{x \in B} A_x$, $\{y \in Y, A^y \cap B \neq \emptyset\} = \bigcup_{x \in B} A_x$, где $A_x = \{y \in Y, x \in A^y\}$, $x \in X$.

Доказательство. Действительно, событие $\{\xi \in A^\eta\} = \bigcup_{x \in X} \{\xi = x, x \in A^\eta\}$, так как $\{y \in Y, q(y) \in A^\eta\} = \bigcup_{x \in X} \{y \in Y, q(y) = x, x \in A^\eta\}$. Отсюда следует равенство (1.57). ■

Пример 3.1.1. Этот пример содержательно иллюстрирует свойство линейности интеграла $p_{g^\xi}(g^{A^\eta}(\cdot))$ в (1.57), см. определение 3.1.1, формулы (1.7), (1.8).

Пусть $g^{\eta, \varkappa}(y, k) = \min(g^{\eta| \varkappa}(y|k), g^{\varkappa}(k))$, $y \in Y, k = 1, 2, \dots$, — распределение пары $(\eta, \varkappa)$ нечетких элементов η и $\varkappa$, A^η — нечеткое множество, определенное на $(Y, \mathcal{P}(Y), P^\eta)$ и принимающее значения в $\mathcal{P}(X)$, $g^\xi(x)$, $x \in X$, — распределение нечеткого элемента ξ , причем $(\eta, \varkappa)$ и ξ — независимы. В таком случае

$g^{\varkappa}(k) \stackrel{\text{def}}{=} \sup_{y \in Y} g^{\eta, \varkappa}(y, k)$ — возможность равенства $\varkappa = k$, $k = 1, 2, \dots$;
 $g^{A^\eta}(x|k) = \sup_{y \in A_x} g^{\eta| \varkappa}(y|k)$ — условная при условии $\varkappa = k$ возможность покрытия $x \in A^\eta$, $k = 1, 2, \dots$, $x \in X$, где $A_x \stackrel{\text{def}}{=} \{y \in Y, x \in A^\eta\}$;
 $g^\eta(y) = \sup_k g^{\eta, \varkappa}(y, k) \equiv \bigoplus_{k=1}^{\infty} (g^{\eta| \varkappa}(y|k) \bullet g^{\varkappa}(k))$ — возможность равенства $\eta = y$, $y \in Y$;

$$g^{A^\eta}(x) \stackrel{\text{def}}{=} P^\eta(\eta \in A_x) \stackrel{\text{def}}{=} \sup_{y \in A_x} g^\eta(y) \equiv \bigoplus_{k=1}^{\infty} (g^{A^\eta}(x|k) \bullet g^{\varkappa}(k)) \quad (1.58)$$

— возможность покрытия $x \in A^\eta$, $x \in X$. Согласно (1.57) и (1.58)

$$\begin{aligned} p_{g^\xi}(g^{A^\eta}(\cdot)) &= \sup_k \min(g^{\varkappa}(k), \sup_{x \in X} \min(g^\xi(x), g^{A^\eta}(x|k))) \equiv \\ &\equiv \bigoplus_{k=1}^{\infty} (g^{\varkappa}(k) \bullet p_{g^\xi}(g^{A^\eta}(\cdot|k))) \end{aligned} \quad (1.59)$$

— возможность покрытия $\xi \in A^\eta$, где $p_{g^\xi}(g^{A^\eta}(\cdot|k))$ — условная при условии $\varkappa = k$ возможность покрытия $\xi \in A^\eta$.

Равенства (1.58) и (1.59) являются вариантами формулы полной возможности (1.36) и в терминах полной возможности иллюстрируют свойство линейности интеграла (1.57).

Если $Y = Y_1 \times \dots \times Y_n$, $y = (y_1, \dots, y_n)$, $\eta = (\eta_1, \dots, \eta_n)$, причем нечеткие элементы $\eta_1, \dots, \eta_n$ взаимно независимы, т.е. $g^\eta(y) = \min_{1 \leq i \leq n} g^{\eta_i}(y_i)$, $y \in Y$, то нечеткие множества $A_1^{\eta_1}, \dots, A_n^{\eta_n}$ взаимно независимы, в частности, и в смысле одноточечного покрытия, при любых отображениях $A_i : Y_i \rightarrow \mathcal{P}(X_i)$, $i = 1, \dots, n$, а нечеткие элементы $\xi_1 = q_1(\eta_1), \dots, \xi_n = q_n(\eta_n)$ взаимно независимы при любых функциях $q(\cdot) : Y_i \rightarrow X_i$, $i = 1, \dots, n$.

Более того, если $\eta = (\eta_1, \eta_2)$ и независимы нечеткие элементы η_1 и η_2 , то для любой функции $q(\cdot) : Y_1 \rightarrow X_1$ и любого отображения

$A : Y_2 \rightarrow \mathcal{P}(X_2)$ нечеткий элемент $\xi = q(\eta_1)$ и нечеткое множество A^{η_2} независимы, так как для любого $x \in X_1$ и $A \in \mathcal{P}(X_2)$ $P^\eta(\xi = x, A^{\eta_2} \cap A = \emptyset, A^{\eta_2} \neq \emptyset) = \min(P^\eta(\xi = x), P^\eta(A^{\eta_2} \cap A = \emptyset, A^{\eta_2} \neq \emptyset))$; $P^\eta(\xi = x, A^{\eta_2} \cap A \neq \emptyset) = \min(P^\eta(\xi = x), P^\eta(A^{\eta_2} \cap A \neq \emptyset))$. В частности, если $X_1 = X_2 = X$, то $\forall x \in X$ $P^\eta(\xi = x, x \in A^{\eta_2}) = \min(g^\xi(x), g^{A^{\eta_2}}(x))$ и соответственно ¹⁾ $P^\eta(\xi \in A^{\eta_2}) = P^\eta(\exists x \in X \xi = x, x \in A^{\eta_2}) = \sup_{x \in X} \min(g^\xi(x), g^{A^{\eta_2}}(x))$.

3.2. Стохастические модели возможности

3.2.0. Введение. Рассмотрим вначале возможностьную модель стохастического эксперимента (С.Э.), математической моделью которого является класс $\mathcal{Pr}$ дискретных вероятностных пространств $(\Omega, \mathcal{P}(\Omega), \text{Pr})$, в которых $\Omega = \{\omega_1, \omega_2, \dots\}$ и значения вероятностей $\text{pr}_i = \text{Pr}(\{\omega_i\})$, $i = 1, 2, \dots$, элементарных событий подчинены лишь условиям упорядоченности и нормировки

$$\text{pr}_1 \geq \text{pr}_2 \geq \dots \geq 0, \quad \text{pr}_1 + \text{pr}_2 + \dots = 1. \quad (2.1)$$

Определение 3.2.1. Класс вероятностей $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$,

$$\text{Pr}(A) = \sum_{i: \omega_i \in A} \text{pr}_i, \quad A \in \mathcal{P}(\Omega), \quad (2.2)$$

удовлетворяющих условиям (2.1), обозначим $\mathbb{Pr}$, соответственно $\mathcal{Pr} \stackrel{\text{def}}{=} \{(\Omega, \mathcal{P}(\Omega), \text{Pr}), \text{Pr} \in \mathbb{Pr}\}$ обозначим стохастическую ²⁾ модель С.Э.

Сопоставим С.Э. класс пространств с возможностью $(\Omega, \mathcal{P}(\Omega), P)$, в которых значения возможностей $p_i = P(\{\omega_i\})$, $i = 1, 2, \dots$, элементарных событий, определяющие возможность $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ формулой

$$P(A) = \sup_{i: \omega_i \in A} p_i, \quad A \in \mathcal{P}(\Omega), \quad (2.3)$$

упорядочены подобно значениям вероятностей в (2.1):

$$1 = p_1 \geq p_2 \geq \dots \geq 0. \quad (2.4)$$

Определим класс $\mathbb{P}$ возможностей, согласованный с классом вероятностей $\mathbb{Pr}$, добавив к условию (2.4), согласующему упорядоченность возможностей $\{\omega_i\}$, $i = 1, 2, \dots$, с упорядоченностью их вероятностей (2.1), аналогичные условия для всех тех исходов $A, B \in \mathcal{P}(\Omega)$ С.Э.,

¹⁾ Разумеется, в общем случае $P^\eta(\{\xi \in A^{\eta_2}\}) = \sup_{x \in X} P^\eta(\{\xi = x\} \cap \{x \in A^{\eta_2}\})$.

²⁾ Если $\mathcal{Pr}$ — модель С.Э., то моделью каждого испытания является некоторое вероятностное пространство $(\Omega, \mathcal{P}(\Omega), \text{Pr})$, $\text{Pr} \in \mathbb{Pr}$.

для которых неравенство $\text{Pr}(A) \leq \text{Pr}(B)$ выполняется для любой вероятности $\text{Pr} \in \mathbb{P}$.

Определение 3.2.2. Максимальный по включению класс $\mathbb{P}$ возможностей $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ называется согласованным с классом $\mathbb{P}$ вероятностей $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$, если $\forall A, B \in \mathcal{P}(\Omega) :$

$$\forall \text{Pr} \in \mathbb{P} \quad \text{Pr}(A) \leq \text{Pr}(B) \Rightarrow \forall P \in \mathbb{P} \quad P(A) \leq P(B), \quad (2.5)$$

где вероятность Pr и возможность P определены соответственно формулами (2.2) и (2.3). События A, B , удовлетворяющие условиям (2.5), назовем $(\mathbb{P}, \mathbb{P})$ -упорядоченными.

Класс $\mathbb{P}$ назовем вполне согласованным с классом $\mathbb{P}$, если $\mathbb{P}$ согласован с $\mathbb{P}$ и

$$\forall A \in \mathcal{P}(\Omega) : \forall P \in \mathbb{P} \quad P(A) = 0 \Rightarrow \forall \text{Pr} \in \mathbb{P} \quad \text{Pr}(A) = 0. \quad (2.5^*)$$

Полная согласованность включает условие согласованности нейтральных элементов («0» и «1») шкал значений возможности и вероятности, согласно которому, если исход A невозможен, $P(A) = 0$, то он и невероятен, $\text{Pr}(A) = 0$, а если $\text{Pr}(A) = 1$, то и $P(A) = 1$; последнее условие учтено в (2.4).

Класс $\mathbb{P}$, согласованный (вполне согласованный) с $\mathbb{P}$, определяет возможность модель $\mathcal{P} = \{(\Omega, \mathcal{P}(\Omega), P), P \in \mathbb{P}\}$. С.Э., согласованную (вполне согласованную) с его стохастической моделью Pr условиями (2.1) и (2.5) ((2.1), (2.5) и (2.5*)).

Классы Pr и $\mathbb{P}$ назовем стохастическими моделями классов $\mathcal{P}$ и $\mathbb{P}$ соответственно.

Заметим, что согласно (2.5), если класс $\mathbb{P}$ согласован с классом $\mathbb{P}$ и $\forall \text{Pr} \in \mathbb{P} \quad \text{Pr}(A) = \text{Pr}(B)$, то $\forall P \in \mathbb{P} \quad P(A) = P(B)$.

Заметим также, что условие (2.5) влечет условие (2.4), ибо в качестве A и B в (2.5) можно выбирать множества $\{\omega_i\}$, $i = 1, 2, \dots$, упорядоченные согласно (2.1).

В монографии [13] показано, что согласованность класса $\mathbb{P}$ с классом $\mathbb{P}$ в определенном смысле характеризует

► *соответствие* «тах», как операции «сложения» возможностей элементарных событий (2.3), операции сложения их вероятностей (2.2);

► *согласованность* теоретико-возможностной независимости с независимостью теоретико-вероятностной и, как следствие,

► *соответствие* «min», как операции «умножения» возможностей P -независимых событий, операции умножения вероятностей Pr -независимых событий, а полная согласованность $\mathbb{P}$ с $\mathbb{P}$ определяет

► *полную согласованность* класса $\mathbb{P}^{|C}$ условных относительно $C \in \mathcal{P}(\Omega)$ возможностей с классом $\mathbb{P}^{|C}$ условных относительно C вероятностей.

Однако класс $\mathcal{P}$ содержит несчетное множество неэквивалентных пространств с возможностью $(\Omega, \mathcal{P}(\Omega), P)$, $P \in \mathbb{P}$, и поэтому является

в высшей степени «размытой» моделью, описывающей С.Э. «в общих чертах», без детализации. В этой связи класс $\mathcal{P}$ естественно представить в виде объединения попарно непересекающихся подклассов *эквивалентных пространств с возможностью*, каждый из которых определяет максимально детальную, в известном смысле неприводимую, не допускающую дальнейшей детализации модель С.Э. Поскольку $\forall \gamma \in \Gamma$ включения $P \in \mathbb{P}$ и $\gamma \circ P \in \mathbb{P}$ эквивалентны, и $\forall P \in \mathbb{P}$ множество $\{\gamma \circ P, \gamma \in \Gamma\} \subset \mathbb{P}$ является классом эквивалентных (и эквивалентных P) возможностей, то класс $\mathbb{P}$ можно представить в виде объединения попарно непересекающихся (неприводимых) классов $\mathbb{P}_{(e)}$, $e \in (0, 1)$, эквивалентных возможностей

$$\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}, \quad (2.6)$$

где $e = 0.e_1e_2\dots$ — двоичная запись числа из $(0, 1)$, определяющего *конкретную упорядоченность* распределения $P(\cdot) \in \mathbb{P}_{(e)}$, заданную отношениями¹⁾ « $\sim$ » $\sim$ «0» и « $\succ$ » $\sim$ «1», и $\forall \gamma(\cdot) \in \Gamma$ $\gamma \circ P_{(e)} \stackrel{\text{def}}{=} \{\gamma \circ P(\cdot), P(\cdot) \in \mathbb{P}_{(e)}\} = \mathbb{P}_{(e)}$, $e \in (0, 1)$. Каждый класс $\mathbb{P}_{(e)}$, $e \in (0, 1)$, определяет *неприводимую* модель С.Э. как класс эквивалентных пространств с возможностью $\mathcal{P}_{(e)} = \{(\Omega, \mathcal{P}(\Omega), P), P \in \mathbb{P}_{(e)}\}$, $e \in (0, 1)$. При этом $\mathcal{P} = \bigcup_{e \in (0,1)} \mathcal{P}_{(e)}$.

Чтобы построить *стохастическую* модель класса $\mathbb{P}_{(e)}$, $e \in (0, 1)$, введем понятия *согласованности* и *максимальной согласованности возможности с вероятностью*.

Определение 3.2.3. Возможность P называется *согласованной с вероятностью* Pr , если существует функция $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr}) \subset \tilde{\Gamma}$, такая, что $\tilde{\gamma}(a) = 0 \Leftrightarrow a = 0$, $\tilde{\gamma}(1) = 1$, и для любого $A \in \mathcal{P}(\Omega)$ $P(A) = \tilde{\gamma}(\text{Pr}(A)) \stackrel{\text{def}}{=} \tilde{\gamma} \circ \text{Pr}(A)$. Поэтому, в частности, если $P(A) = 0$, $\text{Pr}(B) = 1$, то $\text{Pr}(A) = 0$, $P(B) = 1$, $A, B \in \mathcal{P}(\Omega)$. Согласованность P с Pr обозначим $\text{Pr} \sim P$.

Возможность P называется *максимально согласованной с вероятностью* Pr , если $\text{Pr} \sim P$, и для любой возможности $\tilde{P}$, $\text{Pr} \sim \tilde{P} \succ P$, можно указать функцию $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}$, такую, что для любого $A \in \mathcal{P}(\Omega)$ $\tilde{P}(A) = \tilde{\gamma} \circ P(A)$. Максимальную согласованность P с Pr обозначим $\text{Pr} \approx P$.

Если $\text{Pr} \approx P$, то *вероятность* Pr называется *вероятностной моделью возможности* P , *возможность* P называется *Pr-измеримой*, а это её свойство — *вероятностной измеримостью*.

¹⁾ Например, $0.01110\dots$ соответствует $1 = p_1 = p_2 > p_3 > p_4 > p_5 = p_6$, где «ноль» перед «точкой» соответствует всегда верному равенству $1 = p_1$; условимся считать, что $0 < 0.00\dots \leq 0.e_1e_2\dots \leq 0.11\dots < 1$, что будем записывать как $e \in (0, 1)$; «конкретной» назовем упорядоченность, в которой использованы только отношения « $\succ$ » и « $\sim$ ».

Для каждой конкретной вероятности Pr , распределение которой удовлетворяет условию (2.1), распределение любой возможности P , максимально согласованной с Pr , удовлетворяет условию (2.4) с максимальным числом строгих неравенств.

Для любой вероятности $\text{Pr} \in \mathbb{P}\text{r}$ определим класс

$$\mathbb{P}(\text{Pr}) = \{\gamma \circ P, \gamma \in \Gamma, \text{Pr} \approx > P\}, \quad (2.7)$$

возможностей, максимально согласованных с вероятностью¹⁾ Pr . Эта формула определяет многозначное отображение $\mathbb{P}(\cdot) : \mathbb{P}\text{r} \rightarrow \mathcal{P}(\mathbb{P})$, поскольку, согласно определению 3.2.3, если $\text{Pr} \in \mathbb{P}\text{r}$ и $\text{Pr} \approx > P$, то $P \in \mathbb{P}$, и, следовательно, $\mathbb{P}(\text{Pr}) \subset \mathbb{P}$. Обратное к $\mathbb{P}(\cdot)$ отображение $\text{Pr}(\cdot) : \mathbb{P} \rightarrow \mathcal{P}(\mathbb{P}\text{r})$ определяется равенством

$$\text{Pr}(P) = \{\text{Pr} \in \mathbb{P}\text{r}, P \in \mathbb{P}(\text{Pr})\}, \quad P \in \mathbb{P}, \quad (2.8)$$

в котором $\mathbb{P}\text{r}(P)$ — класс всех вероятностей, с каждой из которых возможность $P \in \mathbb{P}$ максимально согласована. Включения $P \in \mathbb{P}(\text{Pr})$ и $\text{Pr} \in \mathbb{P}\text{r}(P)$, по определению, эквивалентны, причем для любой возможности $P \in \mathbb{P}$ $\text{Pr}(P) = \text{Pr}(\gamma \circ P)$, $\gamma \in \Gamma$.

Класс $\mathbb{P}\text{r}(P)$ и каждую вероятность $\text{Pr} \in \mathbb{P}\text{r}(P)$ естественно считать стохастическими моделями класса $\mathbb{P}(\text{Pr})$ и каждой возможности $P \in \mathbb{P}(\text{Pr})$ соответственно.

Каждому классу $\mathbb{P}_{(e)}$ из (2.6) согласно (2.7), (2.8) сопоставлен класс $\mathbb{P}\text{r}_{(e)}$, причем так, что любая возможность $P \in \mathbb{P}_{(e)}$ максимально согласована со всеми вероятностями $\text{Pr} \in \mathbb{P}\text{r}_{(e)}$ и только с ними, поскольку

$$\begin{aligned} \mathbb{P}\text{r}_{(e)} &\stackrel{\text{def}}{=} \{\text{Pr} \in \mathbb{P}\text{r}, \mathbb{P}(\text{Pr}) = \mathbb{P}_{(e)}\}, \\ \mathbb{P}_{(e)} &= \{P \in \mathbb{P}, \text{Pr}(P) = \mathbb{P}\text{r}_{(e)}\}, \quad e \in (0, 1). \end{aligned}$$

Отсюда следует разбиение²⁾

$$\mathbb{P}\text{r} = \bigcup_{e \in (0, 1)} \mathbb{P}\text{r}_{(e)}, \quad (2.9)$$

индуцированное разбиением (2.6), причем в (2.9) $\mathbb{P}\text{r}_{(e)}$ — стохастическая модель $\mathbb{P}_{(e)}$ в (2.6), $e \in (0, 1)$.

На самом деле, как показано в [13], условия согласованности в определении 3.2.3 иногда «слишком нормативны», поскольку удовлетворяющая им возможность P , даже максимально согласованная с вероятностью Pr , зачастую оказывается «тривиальной» в том смысле, что для любых (не пустых) $A, B \in \mathcal{P}(\Omega)$ возможности $P(A) = P(B) = 1$,

¹⁾ Если $\text{Pr} \approx > P$, то $\forall \gamma \in \Gamma \quad \text{Pr} \approx > \gamma \circ P$, и наоборот, если $\exists \gamma \in \Gamma \quad \text{Pr} \approx > \gamma \circ P$, то $\text{Pr} \approx > P$, т.е. класс $\mathbb{P}(\text{Pr})$ (2.7) не зависит от выбора возможности P , максимально согласованной с вероятностью Pr .

²⁾ Действительно, $\mathbb{P}\text{r}_{(e)} \cap \mathbb{P}\text{r}_{(e')} = \emptyset$, $e \neq e'$, и если некоторая вероятность $\text{Pr} \in \mathbb{P}\text{r}$ не содержится ни в одном $\mathbb{P}\text{r}_{(e)}$, то любая возможность $P \in \mathbb{P}(\text{Pr})$ не содержится ни в одном $\mathbb{P}_{(e)}$, $e \in (0, 1)$, что в силу (2.6) невозможно.

в то время как их вероятности $\text{Pr}(A)$ и $\text{Pr}(B)$ могут существенно отличаться. Можно сказать, что в этом случае вероятность Pr характеризует «детали, слишком тонкие» для того, чтобы их могла охарактеризовать возможность P .

Чтобы градации вероятности стали «осознаваемыми» для возможности, «тонкие детали» следует «укрупнить», сделав их более «контрастными» для вероятности и, как следствие, — «различимыми» для возможности. Другими словами, чтобы возможность максимально отражала свойства стохастического объекта, охарактеризованные вероятностью, следует не только «согласовывать» возможность с вероятностью, но и вероятность — с возможностью путём гранулирования пространства элементарных событий. Формально это означает, что требования, сформулированные в определении 3.2.3, следует ослабить так, чтобы они выполнялись не для всех подмножеств Ω , а лишь для подмножеств из некоторой σ -алгебры $\mathcal{A} \subset \mathcal{P}(\Omega)$, т. е. для сужений $\text{Pr}|_{\mathcal{A}}$ и $P|_{\mathcal{A}}$ вероятности Pr и возможности P на $\mathcal{A}$.

Определение 3.2.4. Возможность $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ называется согласованной с вероятностью $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ на σ -алгебре $\mathcal{A}$, если сужение $P(\cdot)|_{\mathcal{A}} : \mathcal{A} \rightarrow [0, 1]$ согласовано с сужением $\text{Pr}(\cdot)|_{\mathcal{A}} : \mathcal{A} \rightarrow [0, 1]$, т. е. если $\text{Pr}|_{\mathcal{A}} \sim^A P|_{\mathcal{A}}$. Последнее означает, что существует функция $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr}) \subset \tilde{\Gamma}$, такая, что $\tilde{\gamma}(a) = 0 \Leftrightarrow a = 0$, $\tilde{\gamma}(1) = 1$, и для любого $A \in \mathcal{A}$ $P(A) = \tilde{\gamma}(\text{Pr}(A))$. Согласованность P с Pr на $\mathcal{A}$ обозначим $\text{Pr} \sim^A P$.

Возможность $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ называется максимально согласованной с вероятностью $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ на $\mathcal{A}$, если $\text{Pr} \sim^A P$ и для любой возможности $\tilde{P}$, $\text{Pr} \sim^A \tilde{P}$, можно указать функцию $\tilde{\tilde{\gamma}}(\cdot) \in \tilde{\Gamma}$, такую, что для любого $A \in \mathcal{A}$ $\tilde{P}(A) = \tilde{\tilde{\gamma}}(\text{Pr}(A))$. Максимальную согласованность P с Pr на $\mathcal{A}$ обозначим $\text{Pr} \approx^A P$ и Pr назовем вероятностной моделью P на $\mathcal{A}$, а P назовем Pr -измеримой на $\mathcal{A}$.

В заключение заметим, что в том случае, когда математическая модель С. Э. определена как вероятностное пространство $(\Omega, \mathcal{P}(\Omega), \text{Pr})$, любая согласованная с ней возможностная модель $(\Omega, \mathcal{P}(\Omega), P)$ не содержит ничего такого, чего нельзя извлечь из модели $(\Omega, \mathcal{P}(\Omega), \text{Pr})$. Но, разумеется, модель $(\Omega, \mathcal{P}(\Omega), P)$ «не обязана» иметь стохастический «прототип», а при его наличии может формулироваться и независимо от последнего. Более того, возможностные модели как раз характерны для «нечетких», «размытых» и т. п. объектов и явлений, не содержащих стохастической компоненты.

Вместе с тем факт существования стохастического «прототипа» открывает путь эмпирического построения его возможностной модели, причем в том числе и тогда, когда его стохастическая модель принципиально не может быть восстановлена эмпирически.

3.2.1. Возможность, максимально согласованная с вероятностью. Этот параграф посвящен построению класса $\tilde{\Gamma}(\text{Pr}) \subset \tilde{\Gamma}$ функций $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$, определяющих класс $\{P = \tilde{\gamma} \circ \text{Pr}, \tilde{\gamma} \in \tilde{\Gamma}(\text{Pr})\}$ взаимно эквивалентных возможностей, максимально согласованных с вероятностью Pr . С этой целью естественно рассматривать вероятности $\text{Pr} \in \mathbb{P}$ и возможности $P \in \mathbb{P}$, тогда $\{P = \tilde{\gamma} \circ \text{Pr}, \tilde{\gamma} \in \tilde{\Gamma}(\text{Pr})\} = \{P \in \mathbb{P}, \text{Pr} \approx > P\} = \mathbb{P}(\text{Pr}), \text{Pr} \in \mathbb{P}$.

3.2.2. Метод построения возможности, максимально согласованной с вероятностью. Напомним, что $\mathbb{P}$ — класс вероятностей, распределения которых выделены условием (2.1), $\mathbb{P}$ — класс возможностей, распределенных согласно (2.4). Задачу построения класса $\mathbb{P}(\text{Pr})$ возможностей, максимально согласованных с вероятностью ¹⁾ $\text{Pr} \in \mathbb{P}$, сформулируем в терминах распределений вероятности $\text{Pr} \in \mathbb{P}$ и возможности $P \in \mathbb{P}(\text{Pr}) \subset \mathbb{P}$, максимально согласованной с Pr . Дело в том, что, как нетрудно заметить, распределение $P \in \mathbb{P}(\text{Pr})$ выделяется среди всех распределений, удовлетворяющих условию (2.4), наличием максимального числа строгих неравенств, отвечающих строгим неравенствам в (2.1); *распределение $P \in \mathbb{P}(\text{Pr})$ будем также называть максимально согласованным с распределением Pr .*

Рассмотрим метод построения функции $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}$, определяющей возможность $P = \tilde{\gamma} \circ \text{Pr}$, максимально согласованную с $\text{Pr} \in \mathbb{P}$, и класса $\tilde{\Gamma}(\text{Pr}) \subset \tilde{\Gamma}$ всех таких функций, удовлетворяющих условиям:

1. распределение $p_i = \tilde{\gamma}(\text{pr}_i), i = 1, 2, \dots$, возможности $P = \tilde{\gamma} \circ \text{Pr}$ удовлетворяет условию (2.4) с максимальным числом строгих неравенств;
2. для любого $A \in \mathcal{P}(\Omega)$

$$P(A) = \sup_{i: \omega_i \in A} \tilde{\gamma}(\text{pr}_i) = \tilde{\gamma}(\sup_{i: \omega_i \in A} \text{pr}_i) = \tilde{\gamma}(\sum_{i: \omega_i \in A} \text{pr}_i) = \tilde{\gamma}(\text{Pr}(A)); \quad (2.10)$$

3. функция $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$ монотонно не убывает и непрерывна на $(0, 1]$, для любого ²⁾ $a \in (0, 1]$ $\tilde{\gamma}(a) > 0$ и $\tilde{\gamma}(0) = 0, \tilde{\gamma}(1) = 1$.
4. Класс $\tilde{\Gamma}(\text{Pr})$ всех таких функций $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$ инвариантен относительно преобразований $\gamma(\cdot) \in \Gamma: \forall \tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr}) \quad \tilde{\Gamma}(\text{Pr}) = \{\gamma \circ \tilde{\gamma}(\cdot), \gamma(\cdot) \in \Gamma\}$.

Пусть $\mathcal{P}_1(\Omega)$ — класс тех $A \in \mathcal{P}(\Omega)$, каждое из которых содержит ω_1 , т.е. пусть $\mathcal{P}_1(\Omega) = \{A \in \mathcal{P}(\Omega), \omega_1 \in A\}$. Для каждого $A \in \mathcal{P}_1(\Omega)$ $P(A) = \sup_{i: \omega_i \in A} p_i = p_1 = 1$. Так как минимальный по включению интервал $\Delta_1 \subset [0, 1]$, содержащий значения всех вероятностей $\text{Pr}(A), A \in \mathcal{P}_1(\Omega)$, есть $\Delta_1 = [\text{pr}_1, 1]$, то равенства (2.10) и условия 3.

¹⁾ если $\text{Pr} \in \mathbb{P}$, то любая согласованная с Pr возможность $P \in \mathbb{P}$.

²⁾ Это условие связано с тем, что, если событие $A \in \mathcal{P}(\Omega)$ невозможно, $P(A) = 0$, то A невероятно, $\text{Pr}(A) = 0$, и если $\text{Pr}(A) = 1$, то $P(A) = 1$.

будут выполнены для всех $A \in \mathcal{P}_1(\Omega)$, если $\tilde{\gamma}(a) = 1$, $a \in [\text{pr}_1, 1]$; при таком определении $\tilde{\gamma}(\cdot)$ $P(A) = \tilde{\gamma}(\text{Pr}(A)) = 1$, $A \in \mathcal{P}_1(\Omega)$.

Пусть $\mathcal{P}_2(\Omega)$ — класс всех подмножеств Ω , содержащих ω_2 и не содержащих ω_1 , т. е. пусть $\mathcal{P}_2(\Omega) = \{A \in \mathcal{P}(\Omega) \setminus \mathcal{P}_1(\Omega), \omega_2 \in A\}$. Минимальный по включению интервал $\Delta_2 \subset [0, 1]$, содержащий все значения вероятностей $\text{Pr}(A)$, $A \in \mathcal{P}_2(\Omega)$, есть $\Delta_2 = [\text{pr}_2, 1 - \text{pr}_1]$, и условия 3. и равенства (2.10) будут выполнены для всех $A \in \mathcal{P}_2(\Omega)$, если $\tilde{\gamma}(a) = \text{pr}_2 \leq \text{pr}_1 = 1$, $a \in \Delta_2$, а именно, для любого $A \in \mathcal{P}_2(\Omega)$ $P(A) = \text{pr}_2 = \tilde{\gamma}(\text{Pr}(A))$, причем, если $\Delta_1 \cap \Delta_2 = \emptyset$, то согласно условию 1. максимальной согласованности и условиям 3. следует определить $\text{pr}_2 < \text{pr}_1$, а если $\Delta_1 \cap \Delta_2 \neq \emptyset$, то $\text{pr}_2 = \text{pr}_1$.

Пусть

$$\mathcal{P}_k(\Omega) = \{A \in \mathcal{P}(\Omega) \setminus (\mathcal{P}_1(\Omega) \cup \mathcal{P}_2(\Omega) \cup \dots \cup \mathcal{P}_{k-1}(\Omega)), \omega_k \in A\}, \quad (2.11)$$

тогда $\Delta_k = [\text{pr}_k, 1 - \text{pr}_1 - \dots - \text{pr}_{k-1}]$ — минимальный интервал, содержащий все значения $\text{Pr}(A)$, $A \in \mathcal{P}_k(\Omega)$, и для любого $A \in \mathcal{P}_k(\Omega)$ $P(A) = \sup_{i: \omega_i \in A} \text{pr}_i = \text{pr}_k = \tilde{\gamma}(\text{Pr}(A))$, если $\tilde{\gamma}(a) = \text{pr}_k$, $a \in \Delta_k$, где

$$\begin{aligned} &\text{либо } \text{pr}_k < \text{pr}_{k-1}, \text{ если } \Delta_k \cap \Delta_{k-1} = \emptyset, \\ &\text{либо } \text{pr}_k = \text{pr}_{k-1}, \text{ если } \Delta_k \cap \Delta_{k-1} \neq \emptyset, \quad k = 1, 2, \dots \end{aligned} \quad (2.12)$$

На k -том шаге функция $\tilde{\gamma}(\cdot)$ определена на $\bigcup_{i=1}^k \Delta_i$ как функция $\tilde{\gamma}_k(\cdot): \bigcup_{i=1}^k \Delta_i \rightarrow [0, 1]$, при этом $\text{pr}_i = \tilde{\gamma}(\text{pr}_i) = \tilde{\gamma}_k(\text{pr}_i)$, $i = 1, \dots, k$, и для любого $A \in \bigcup_{i=1}^k \mathcal{P}_i(\Omega)$ $P(A) = \tilde{\gamma}_k(\text{Pr}(A))$. Следовательно, рассмотренный алгоритм определяет функцию $\tilde{\gamma}_\infty(\cdot): \bigcup_{i=1}^\infty \Delta_i \rightarrow [0, 1]$ так, что $\tilde{\gamma}_\infty(a) = \text{pr}_i$, $a \in \Delta_i$, $\text{pr}_i = \tilde{\gamma}_\infty(\text{pr}_i)$, $i = 1, 2, \dots$, и для любого $A \in \bigcup_{i=1}^\infty \mathcal{P}_i(\Omega) = \mathcal{P}(\Omega)$ $P(A) = \tilde{\gamma}_\infty(\text{Pr}(A))$; равенство $\bigcup_{i=1}^\infty \mathcal{P}_i(\Omega) = \mathcal{P}(\Omega)$ следует из определения (2.11).

Покажем, что функция $\tilde{\gamma}_\infty(\cdot)$ может быть продолжена до функции $\tilde{\gamma}(\cdot): [0, 1] \rightarrow [0, 1]$, непрерывной на $(0, 1]$, монотонно неубывающей на $[0, 1]$ и удовлетворяющей условию $\tilde{\gamma}(0) = 0$. Для этого заметим, что поскольку множество $\bigcup_{i=1}^k \Delta_i$ замкнуто в $[\text{pr}_k, 1]$, то множество

$[\text{pr}_k, 1] \setminus (\bigcup_{i=1}^k \Delta_i)$, на котором не определена функция $\tilde{\gamma}_k(\cdot)$, открыто и либо пусто, либо является объединением конечного числа непересекающихся открытых промежутков. Поэтому функцию $\tilde{\gamma}_k(\cdot)$ всегда можно продолжить до непрерывной монотонно неубывающей функции, определенной на $[\text{pr}_k, 1]$, принимающей значение pr_i на интерва-

ле $\Delta_i \subset [pr_k, 1]$, $i = 1, 2, \dots, k$, $k = 1, 2, \dots$, а $\tilde{\gamma}_\infty(\cdot)$ — до функции $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$, монотонно неубывающей, непрерывной на $(0, 1]$, принимающей значение p_i на интервале Δ_i , $i = 1, 2, \dots$, причем — не единственным способом. Класс всех таких функций $\tilde{\gamma}(\cdot)$ обозначим $\tilde{\Gamma}(\text{Pr})$; по определению $\tilde{\Gamma}(\text{Pr}) \subset \tilde{\Gamma}$, причём, как нетрудно видеть, $\tilde{\Gamma}(\text{Pr}) = \{\gamma \circ \tilde{\gamma}(\cdot), \gamma(\cdot) \in \Gamma\}$, где $\tilde{\gamma}(\cdot)$ — любая функция из $\tilde{\Gamma}(\text{Pr})$.

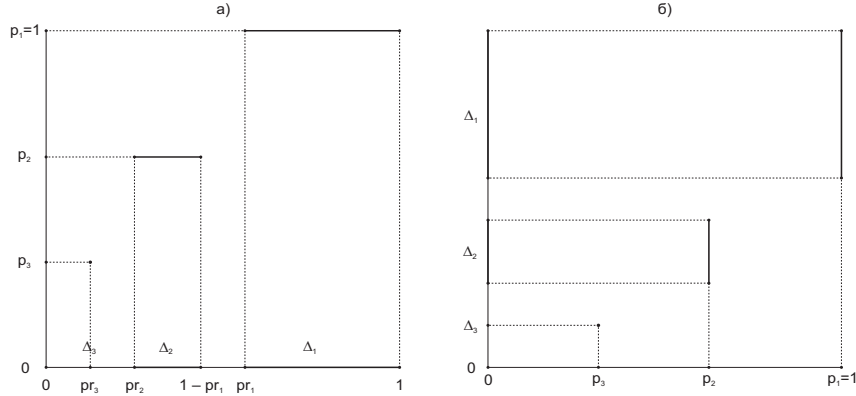


Рис. 3.7. а) Фрагменты графика функции $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr})$, определяющей распределение $p_i = \tilde{\gamma}(pr)$, $pr \in \Delta_i$, $i = 1, 2, 3$, возможности P из класса возможностей, удовлетворяющих условиям $1 = p_1 > p_2 > p_3 > 0$, максимально согласованных с вероятностями Pr из класса вероятностей, удовлетворяющих условиям $2pr_1 > 1$, $pr_1 + 2pr_2 > 1$, $pr_1 + pr_2 + 2pr_3 = 1 + pr_3 > 1$, согласно которым $\Delta_i \cap \Delta_j = \emptyset$, $i \neq j$, $i, j = 1, 2, 3$. Интервал Δ_1 содержит pr_1 , $pr_1 + pr_3$, $pr_1 + pr_2$ и $pr_1 + pr_2 + pr_3 = 1$, интервал Δ_2 содержит pr_2 и $pr_2 + pr_3 = 1 - pr_1$, $\Delta_3 = \{pr_3\}$.

б) Фрагменты графика многозначного отображения $\tilde{\gamma}^{-1}(\cdot) : [0, 1] \rightarrow \mathcal{P}([0, 1])$, определяющего множества $\Delta_i = \tilde{\gamma}^{-1}(p_i)$, $i = 1, 2, 3$, удовлетворяющие условиям $\Delta_i \cap \Delta_j = \emptyset$, $i \neq j$, $i, j = 1, 2, 3$, характеризующим класс вероятностей, с каждой из которых максимально согласованы все возможности из класса возможностей, удовлетворяющих условиям $1 = p_1 > p_2 > p_3 > 0$.

Суммируем полученные результаты, см. рис. 3.7.

Лемма 3.2.1. Пусть $pr_1, pr_2, \dots$ — распределение вероятности $\text{Pr} \in \mathbb{P}$, $\mathbb{P}(\text{Pr})$ — класс возможностей, максимально согласованных с Pr , $\tilde{\Gamma}(\text{Pr})$ — класс функций $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1]$, монотонно неубывающих, непрерывных на $(0, 1]$ и удовлетворяющих условиям: $\tilde{\gamma}(0) = 0$, $\tilde{\gamma}(a) = p_i$, $a \in \Delta_i = [pr_i, 1 - pr_1 - \dots - pr_{i-1}]$, $i = 1, 2, \dots$, где $p_i > p_{i+1}$, если $\Delta_i \cap \Delta_{i+1} = \emptyset$, $p_i = p_{i+1}$, если $\Delta_i \cap \Delta_{i+1} \neq \emptyset$, $i = 1, 2, \dots$. Тогда

1. $\mathbb{P}(\text{Pr}) = \{\tilde{\gamma} \circ \text{Pr}, \tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr})\}$,
2. всякая возможность $P \in \mathbb{P}(\text{Pr})$ может быть задана распределением $p_i = \tilde{\gamma}(pr_i)$, $i = 1, 2, \dots$, где $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr})$,

3. при этом функция $\tilde{\gamma}(\cdot)$ непрерывна на $[0, 1]$, если либо $\Delta_i \cap \Delta_{i+1} = \emptyset$ для бесконечно многих $i \in \{1, 2, \dots\}$, либо для некоторого $n \in \{1, 2, \dots\}$ $\text{pr}_n > 0 = \text{pr}_{n+1} = \text{pr}_{n+2} = \dots$, т. е. $\Omega = \{\omega_1, \dots, \omega_n\}$.

Если же $\Delta_i \cap \Delta_{i+1} = \emptyset$ лишь для конечного числа $i \in \{1, 2, \dots\}$, т. е. начиная с некоторого $n \in \{1, 2, \dots\}$ $\Delta_i \cap \Delta_{i+1} \neq \emptyset$, $i = n, n+1, \dots$, причем $\text{pr}_{n+1} > 0$, то $\tilde{\gamma}(\cdot)$ разрывна в нуле: $\lim_{a \rightarrow 0} \tilde{\gamma}(a) = \tilde{\gamma}(\text{pr}_n) = \text{pr}_n > 0 = \tilde{\gamma}(0)$.

4. В силу непрерывности $\tilde{\gamma}(\cdot)$ на $(0, 1]$ для любого $A \in \mathcal{P}(\Omega)$

$$\begin{aligned} P(A) &= \sup\{\text{pr}_i \mid i = 1, 2, \dots, \omega_i \in A\} = \sup\{\tilde{\gamma}(\text{pr}_i) \mid i = 1, 2, \dots, \omega_i \in A\} = \\ &= \tilde{\gamma}(\sup\{\text{pr}_i \mid i = 1, 2, \dots, \omega_i \in A\}) = \tilde{\gamma}(\text{Pr}(A)), \end{aligned}$$

ибо $\sup\{\text{pr}_i \mid i = 1, 2, \dots, \omega_i \in A\} = \text{pr}_{i(A)}$, где $i(A) = \min\{i \mid i \in \{1, 2, 3, \dots\}, \omega_i \in A\}$, $\tilde{\gamma}(\text{pr}_{i(A)}) = \tilde{\gamma}(\text{Pr}(A))$, так как $\text{Pr}(A) \in \Delta_{i(A)}$, и $\tilde{\gamma}(\text{pr}_{i(A)}) = \text{pr}_{i(A)} = P(A)$; как следствие

5. $\tilde{\Gamma}(\text{Pr})$ можно охарактеризовать как класс гомоморфизмов шкалы $([0, 1], \leq, +, \times)$ значений вероятности в шкалу $([0, 1], \leq, +, \bullet)$ значений возможности, определяющих «взаимную симметрию» значений вероятности Pr и возможности P , связанных условием $\text{Pr} \approx P$, а именно, $\forall \text{Pr} \forall P, \text{Pr} \approx P, \exists \tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr}) \forall A \in \mathcal{P}(\Omega) \tilde{\gamma} \circ \text{Pr}(A) \stackrel{\text{def}}{=} \tilde{\gamma}\left(\text{pr}\left(\left(\bigoplus_{j \in J(A)} \chi_j\right)(\cdot)\right)\right) = \tilde{\gamma}\left(\sum_{j \in J(A)} \text{Pr}(\{\omega_j\})\right) =$
 $= \bigoplus_{j \in J(A)} \tilde{\gamma} \circ \text{Pr}(\{\omega_j\}) = \bigoplus_{j \in J(A)} P(\{\omega_j\}) \stackrel{\text{def}}{=} P(A)$, где $J(A) = \{j, \omega_j \in A\}$,
 $\chi_j(\omega) = \begin{cases} 1, & \omega = \omega_j, \\ 0, & \omega \neq \omega_j, \end{cases} \quad j = 1, 2, \dots$ ■

Пример 3.2.1. Пусть для некоторого $q > 0$ $\text{Pr}(\{\omega_k\}) = \text{pr}_k = q/(1+q)^k$, $k = 1, 2, \dots$, $\sum_{k=1}^{\infty} \text{pr}_k = 1$. Для любого $n = 1, 2, \dots$, как нетрудно проверить, $1/(1+q)^n = 1 - \text{pr}_1 - \dots - \text{pr}_n = \text{pr}_{n+1} + \dots \leq \text{pr}_n$, если и только если $q > 1$. В этом случае все интервалы $\Delta_1 = [\text{pr}_1, 1]$, $\dots$, $\Delta_j = [\text{pr}_j, 1 - \text{pr}_1 - \dots - \text{pr}_{j-1}]$, $j = 2, 3, \dots$, попарно не пересекаются, их суммарная длина $1/q < 1$; функции $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\text{Pr})$ постоянны на каждом интервале: $\tilde{\gamma}(\text{pr}) = \text{pr}_j$, $\text{pr} \in \Delta_j$, причем $1 = \text{pr}_1 > \text{pr}_2 > \dots$, и, если выбрать последовательность $\text{pr}_j \rightarrow 0$ при $j \rightarrow \infty$, то $\tilde{\gamma}(\cdot)$ можно доопределить по непрерывности на всем интервале $[0, 1]$, положив $\tilde{\gamma}(0) = 0$. С другой стороны, если $q \leq 1$, то все интервалы Δ_j и Δ_{j+1} , $j = 1, 2, \dots$, пересекаются, и класс $\tilde{\Gamma}(\text{Pr})$ состоит из единственной функции $\tilde{\gamma}(a) = \begin{cases} 1, & a \in (0, 1], \\ 0, & a = 0. \end{cases}$

3.2.3. Возможность, согласованная с вероятностью на σ -алгебре. Гранулирование Ω . В связи с примером 3.2.1 обратим внимание на «парадокс равновозможности», обусловленный, как нетрудно заметить, неаддитивностью (в обычном понимании) возмож-

ности. Дело в том, что, например, при $q = 1$ $\text{pr}_k = 2^{-k}$, $k = 1, 2, \dots$, но распределение возможности, максимально согласованной с Pr , «тривиально»: $\text{pr}_k = 1$, $k = 1, 2, \dots$, т. е. ω_1 и ω_k равновозможны, хотя их вероятности отличаются в 2^k раз! В этом случае «градации» вероятности $\text{Pr}(\{\omega_k\})$, $k = 1, 2, \dots$, недостаточны, чтобы их могла «различить» возможность P , даже максимально согласованная с Pr .

Для того, чтобы «вероятностные детали» сделать «различимыми» для возможности P , «гранулируем» $\Omega = \{\omega_1, \omega_2, \dots\}$, объявив наблюдаемыми не элементарные события $\omega_1, \omega_2, \dots$, а гранулы $\Omega_1, \Omega_2, \dots$ — попарно непересекающиеся подмножества Ω , образующие разбиение

$$\Omega = \bigcup_{k=1}^{\infty} \Omega_k, \quad (2.13)$$

для которого $\text{Pr}(\Omega_1) > \text{Pr}(\Omega_2) > \dots$, $\sum_{k=1}^{\infty} \text{Pr}(\Omega_k) = 1$.

Обозначим $\mathcal{A}$ σ -алгебру подмножеств Ω , порожденную ¹⁾ разбиением (2.13), $\text{Pr}|_{\mathcal{A}}(\cdot) : \mathcal{A} \rightarrow [0, 1]$ — сужение вероятности $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ на $\mathcal{A}$ и $(\Omega, \mathcal{A}, \text{Pr}|_{\mathcal{A}})$ — соответствующее вероятностное пространство, в котором вероятность $\text{Pr}|_{\mathcal{A}}$ определена своими значениями на гранулах: $\text{Pr}|_{\mathcal{A}}(\Omega_k) \stackrel{\text{def}}{=} \text{Pr}(\Omega_k)$, $k = 1, 2, \dots$. Гранулирование Ω формально означает замену $(\Omega, \mathcal{P}(\Omega), \text{Pr})$ на $(\Omega, \mathcal{A}, \text{Pr}|_{\mathcal{A}})$ и замену требований 1. и 2. к функции $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}$, см. (2.10), на следующие, более слабые:

1*. возможности $P|_{\mathcal{A}}(\Omega_i) \stackrel{\text{def}}{=} \tilde{\gamma}(\text{Pr}(\Omega_i))$, $i = 1, 2, \dots$, должны удовлетворять условию $1 = P|_{\mathcal{A}}(\Omega_1) \geq P|_{\mathcal{A}}(\Omega_2) \geq \dots$ с максимальным числом строгих неравенств, и

2*. для любого $A \in \mathcal{A}$ $P(A) = \sup_{k: \Omega_k \subset A} \tilde{\gamma}(\text{Pr}(\Omega_k)) = \tilde{\gamma}(\sup_{k: \Omega_k \subset A} \text{Pr}(\Omega_k)) = \tilde{\gamma}(\sum_{k: \Omega_k \subset A} \text{Pr}(\Omega_k)) \stackrel{\text{def}}{=} P|_{\mathcal{A}}(A)$.

Так определенную возможность $P|_{\mathcal{A}}(\cdot) : \mathcal{A} \rightarrow [0, 1]$ назовем максимально согласованной с $\text{Pr}|_{\mathcal{A}}(\cdot)$ на $\mathcal{A}$, сохранив принятое ранее обозначение $\text{Pr}|_{\mathcal{A}} \approx^> P|_{\mathcal{A}}$, см. определение 3.2.4, для максимальной согласованности. Если же речь идет о согласованности (максимальной согласованности) $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ с $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ на $\mathcal{A} \subset \mathcal{P}(\Omega)$, то (см. определение 3.2.4) будем пользоваться символом $\text{Pr} \sim^{\mathcal{A}}_> P$ ($\text{Pr} \approx^{\mathcal{A}}_> P$).

Пусть $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ — любое ²⁾ продолжение возможности $P|_{\mathcal{A}}(\cdot) : \mathcal{A} \rightarrow [0, 1]$, $\text{Pr}|_{\mathcal{A}} \approx^> P|_{\mathcal{A}}$, на алгебру $\mathcal{P}(\Omega)$ всех подмножеств

¹⁾ Элементы $\mathcal{A} = \sigma(\Omega_1, \Omega_2, \dots)$ — подмножества Ω , представимые конечными или счетными объединениями Ω_k , $k = 1, 2, \dots$, а также пустое множество.

²⁾ Не обязательно максимальное продолжение $P|_{\mathcal{A}}$, см. §9 гл. 1 в [13].

Ω , удовлетворяющее условию (2.4). Тогда, очевидно, $\text{Pr} \approx^A > \text{P}$. Такое продолжение P возможности $\text{P}|_{\mathcal{A}}$ назовем согласованным с Pr .

Разумеется, смысл гранулирования Ω в том, чтобы в пункте 1* все неравенства были строгими, вероятность каждой гранулы — минимальной, а число гранул — максимальным. При этом условии возможность $\text{P}|_{\mathcal{A}}$ и любое согласованное с Pr ее продолжение P «различат наиболее тонкий узор $\mathcal{A}$, различаемый вероятностью» Pr . В этом смысле возможность и вероятность будут максимально согласованы на $\mathcal{A} = \sigma(\Omega_1, \Omega_2, \dots)$.

Алгоритмы гранулирования, позволяющие *максимально согласовать вероятность с возможностью*, рассмотрены в [13].

3.2.4. Когерентные разбиения $\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}$ и $\text{Pr} = \bigcup_{e \in (0,1)} \text{Pr}_{(e)}$. В этом параграфе взаимно однозначное соответствие между классами $\mathbb{P}_{(e)}$ и $\text{Pr}_{(e)}$, $e \in (0,1)$ в разбиениях $\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}$ (2.6) и $\text{Pr} = \bigcup_{e \in (0,1)} \text{Pr}_{(e)}$ (2.9) охарактеризовано конструктивно.

Пусть $\Omega = \{\omega_1, \omega_2, \dots\}$, $p_i = P(\{\omega_i\})$, $\text{pr}_i = \text{Pr}(\{\omega_i\})$, $i = 1, 2, \dots$, и $e = 0.e_1e_2\dots$ — двоичная запись числа $e \in (0,1)$. Результаты, полученные в § 3.2.1, позволяют для любого $e \in (0,1)$ определить классы $\mathbb{P}_{(e)}$ и $\text{Pr}_{(e)}$ так, что для любой возможности $\text{P} \in \mathbb{P}_{(e)}$ и для любой вероятности $\text{Pr} \in \text{Pr}_{(e)}$:

$$\begin{aligned} \text{либо } e_i = 1 &\Leftrightarrow p_i > p_{i+1} \Leftrightarrow \Delta_i \cap \Delta_{i+1} = \emptyset \Leftrightarrow \\ &\Leftrightarrow f_i \stackrel{\text{def}}{=} \text{pr}_1 + \dots + \text{pr}_{i-1} + 2\text{pr}_i > 1, \\ \text{либо } e_i = 0 &\Leftrightarrow p_i = p_{i+1} \Leftrightarrow \Delta_i \cap \Delta_{i+1} \neq \emptyset \Leftrightarrow f_i \leq 1, \quad i = 1, 2, \dots \end{aligned} \quad (2.14)$$

В (2.14) эквивалентности

$$e_i = 1 \Leftrightarrow p_i > p_{i+1}, \quad e_i = 0 \Leftrightarrow p_i = p_{i+1}, \quad i = 1, 2, \dots, \quad (2.15)$$

означают¹⁾, что каждому $e \in (0,1)$ взаимно однозначно сопоставлен класс $\mathbb{P}_{(e)} \subset \mathbb{P}$ эквивалентных возможностей, распределения которых одинаково упорядочены согласно условиям (2.15).

Среди условий (2.14) эквивалентности

$$\begin{aligned} e_i = 1 &\Leftrightarrow \Delta_i \cap \Delta_{i+1} = \emptyset \Leftrightarrow f_i > 1, \\ e_i = 0 &\Leftrightarrow \Delta_i \cap \Delta_{i+1} \neq \emptyset \Leftrightarrow f_i \leq 1, \quad i = 1, 2, \dots, \end{aligned} \quad (2.16)$$

¹⁾ Эквивалентности в (2.15) удобно выразить правилом: «>» — под единицей, «=» — под нулем, например,

$$e = 0 \quad . \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad \dots$$

$$1 = p_1 > p_2 > p_3 = p_4 = p_5 > \dots$$

означают, что при любом $f_i > 1$ левая граница pr_i интервала $\Delta_i = [\text{pr}_i, 1 - \text{pr}_1 - \dots - \text{pr}_{i-1}]$ правее правой границы $1 - \text{pr}_1 - \dots - \text{pr}_i$ интервала $\Delta_{i+1} = [\text{pr}_{i+1}, 1 - \text{pr}_1 - \dots - \text{pr}_i]$, т. е. что интервалы Δ_i и Δ_{i+1} не пересекаются, а при любом $f_i \leq 1$ — пересекаются, $i = 1, 2, \dots$

Заметим, что условия (2.16) определяют любой класс $\mathbb{P}_{(e)}$, $e \in (0, 1)$, как выпуклое подмножество $\mathbb{P}_{\text{r}}$. Действительно, если распределения $\text{pr}_1^k \geq \text{pr}_2^k \geq \dots$, $k = 1, \dots, m$, удовлетворяют условиям: $\forall k = 1, \dots, m$ либо $e_i = 1 \Leftrightarrow f_i^k \stackrel{\text{def}}{=} \text{pr}_1^k + \dots + \text{pr}_{i-1}^k + 2\text{pr}_i^k > 1$, либо $e_i = 0 \Leftrightarrow f_i^k \leq 1$, $i = 1, 2, \dots$, то для любых $\lambda_1 \geq 0, \dots, \lambda_m \geq 0$, таких, что $\lambda_1 + \dots + \lambda_m = 1$, распределение $\text{pr}_i = \sum_{k=1}^m \lambda_k \text{pr}_i^k$, $i = 1, 2, \dots$, удовлетворяет условию (2.16).

Замечание 3.2.1. Если $\text{pr}_1 \geq \text{pr}_2 \geq \dots \geq \text{pr}_n \geq 0 = \text{pr}_{n+1} = \dots$, или, иначе говоря, если $\Omega = \{\omega_1, \dots, \omega_n\}$, то для $i = 1, 2, \dots, n-1$ выполнены условия (2.14), а для $i = n$

$$\begin{aligned} e_n = 1 &\Leftrightarrow p_n > 0 \Leftrightarrow \text{pr}_1 + \dots + \text{pr}_n > 1 - \text{pr}_n \Leftrightarrow \text{pr}_n > 0, \\ e_n = 0 &\Leftrightarrow p_n = 0 \Leftrightarrow \text{pr}_1 + \dots + \text{pr}_n \leq 1 - \text{pr}_n \Leftrightarrow \text{pr}_n = 0, \end{aligned} \quad (2.17)$$

Заметим, что согласно обозначениям, использованным в (2.14),

$$\begin{aligned} \text{pr}_1 &= \frac{1}{2}f_1, \text{pr}_2 = \frac{1}{2}(f_2 - \frac{1}{2}f_1), \dots, \text{pr}_k = \frac{1}{2}(f_k - \sum_{j=1}^{k-1} (f_j/2^{k-j})), \dots, \\ f_k &\in (0, \frac{k+1}{k}], \quad k = 1, 2, \dots \end{aligned} \quad (2.18)$$

Замечание 3.2.2. Согласно лемме 3.2.1 неравенство $p_i > p_{i+1}$ влечет неравенство $\text{pr}_i > \text{pr}_{i+1}$, но обратное неверно, $i = 1, 2, \dots$. Если в (2.14) неравенство $\text{pr}_1 + \dots + \text{pr}_{i-1} + 2\text{pr}_i > 1$ переписать в виде неравенства

$$\text{Pr}(\{\omega_i\} | \Omega \setminus \{\omega_1, \dots, \omega_{i-1}\}) = \frac{\text{pr}_i}{1 - \text{pr}_1 - \dots - \text{pr}_{i-1}} > \frac{1}{2}$$

для условной вероятности, то согласно (2.14) $p_i > p_{i+1} \Leftrightarrow \Leftrightarrow \text{Pr}(\{\omega_i\} | \Omega \setminus \{\omega_1, \dots, \omega_{i-1}\}) > 1/2$. Это можно интерпретировать так: $p_i > p_{i+1}$, если и только если элементарное событие $\{\omega_i\}$ более вероятно, чем событие $\{\omega_{i+1}, \omega_{i+2}, \dots\}$.

С другой стороны, равенство $\text{pr}_i = \text{pr}_{i+1}$ влечет равенство $p_i = p_{i+1}$, но обратное опять-таки неверно. В частности, если $\text{pr}_i = q/(1+q)^i$, $i = 1, 2, \dots$, то при $q > 1$ $\text{pr}_1 > \text{pr}_2 > \dots$ и $1 = p_1 > p_2 > \dots$, а при $0 < q \leq 1$ $\text{pr}_1 > \text{pr}_2 > \dots$, но $1 = p_1 = p_2 = \dots$, т.е. в этом случае все элементарные события $\{\omega_i, i = 1, 2, \dots\}$ равновозможны, хотя их вероятности экспоненциально убывают при $i \rightarrow \infty$.

Подведем итоги.

Теорема 3.2.1. Пусть $e = 0.e_1e_2\dots \in (0, 1)$, $\mathbb{P}_{(e)} \subset \mathbb{P}$ — класс эквивалентных возможностей, распределенных согласно (2.15), и

$\mathbb{Pr}_{(e)} \subset \mathbb{Pr}$ — выпуклый класс вероятностей, распределенных согласно (2.14). Тогда

- $\forall e \in (0, 1) \forall P \in \mathbb{P}_{(e)} \forall \mathbb{Pr} \in \mathbb{Pr}_{(e)} \mathbb{Pr} \approx_{>} P \text{ и } \exists \tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\mathbb{Pr}) P = \tilde{\gamma} \circ \mathbb{Pr};$
- для различных $e, e' \in (0, 1) \quad \mathbb{P}_{(e)} \cap \mathbb{P}_{(e')} = \emptyset, \mathbb{Pr}_{(e)} \cap \mathbb{Pr}_{(e')} = \emptyset$ и

$$\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}, \quad \mathbb{Pr} = \bigcup_{e \in (0,1)} \mathbb{Pr}_{(e)}. \quad (2.19)$$

Класс $\mathbb{P}_{(e)}$ назовем максимально согласованным с классом $\mathbb{Pr}_{(e)}$, $e \in (0, 1)$. Разбиения в (2.19) назовем когерентными.

Замечание 3.2.3. В одной из задач эмпирического построения возможности, рассмотренных в § 3.3, речь идет о С.Э., исходы которого контролируются вероятностями из неизвестного, но фиксированного условия выполнения С.Э. класса $\mathbb{Pr}_{(e)}$, и требуется по результатам наблюдений за исходами С.Э. определить, какой класс $\mathbb{Pr}_{(e)}$ фиксируют условия выполнения С.Э. Решение этой задачи определяет единственную с точностью до эквивалентности возможность $P \in \mathbb{P}_{(e)}$, максимально согласованную с каждой вероятностью $\mathbb{Pr} \in \mathbb{Pr}_{(e)}$, и возможностную модель $(\Omega, \mathcal{P}(\Omega), P)$ эксперимента, выполняемого при названных условиях.

Своеобразие этой задачи состоит в том, что вероятность, контролирующая каждый исход С.Э., вообще говоря, может произвольно изменяться от испытания к испытанию, оставаясь в пределах некоторого класса $\mathbb{Pr}_{(e)}$. Результаты таких испытаний, вообще говоря, не позволяют восстановить их стохастическую модель, но, как показано в § 3.3, при некоторых ограничениях на класс $\mathbb{Pr}_{(e)}$ возможностная модель С.Э. может быть восстановлена безошибочно на основе результатов п.н. конечного числа испытаний.

Класс $\mathbb{P}_{(e)}$ и любую возможность $P \in \mathbb{P}_{(e)}$, которые могут быть определены эмпирически на основе наблюдений за исходами С.Э., моделью которого является класс вероятностных пространств $(\Omega, \mathcal{P}(\Omega), \mathbb{Pr})$, $\mathbb{Pr} \in \mathbb{Pr}_{(e)}$, естественно назвать стохастически $\mathbb{Pr}_{(e)}$ -измеримыми. Всякая $\mathbb{Pr}_{(e)}$ -измеримая возможность определяется равенством $P(A) = \tilde{\gamma}(\mathbb{Pr}(A))$, $A \in \mathcal{P}(\Omega)$, где $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}(\mathbb{Pr})$, $\mathbb{Pr} \in \mathbb{Pr}_{(e)}$.

3.3. Эмпирическое построение возможности

3.3.0. Введение. Свойства стохастической измеримости возможности, см. замечание 3.2.3, позволяют свести задачу эмпирического построения последней к задаче проверки статистических гипотез о классах $\mathbb{P}_{(e)}$, $e \in (0, 1)$, эквивалентных возможностям, каждый из которых определен той или иной конкретной упорядоченностью значений возможностей элементарных событий. Дело в том, что эти классы образуют разбиение $\mathbb{P} = \bigcup_{e \in (0,1)} \mathbb{P}_{(e)}$ (2.6) класса $\mathbb{P}$ возможностей и

определяют разбиение $\mathbb{Pr} = \bigcup_{e \in (0,1)} \mathbb{Pr}_{(e)}$ (2.9) класса $\mathbb{Pr}$ вероятностей на классы вероятностей $\mathbb{Pr}_{(e)}$, $e \in (0,1)$, причем так, что каждая возможность $P \in \mathbb{Pr}_{(e)}$ максимально согласована с каждой вероятностью $\Pr \in \mathbb{Pr}_{(e)}$ ($\Pr$ -измерима) и определяется¹⁾ последней равенством $P(A) = \tilde{\gamma}(\Pr(A))$, $A \in \mathcal{P}(\Omega)$, см. лемму 3.2.1.

Последнее означает, что задача эмпирического определения стохастически измеримой возможности, как *задача выбора одного из классов эквивалентных возможностей, образующих разбиение* $\mathbb{Pr}$, сводится к задаче проверки статистических гипотез, в которой, в предположении, что результаты испытаний контролируются, вообще говоря, изменяющимися от испытания к испытанию вероятностями $\Pr \in \mathbb{Pr}_{(e')}$, требуется на основе результатов испытаний принять решение о значении $e' \in (0,1)$, фиксированном условиями испытаний.

Заметим, что в то время как все пространства с возможностью $(\Omega, \mathcal{P}(\Omega), P)$, $P \in \mathbb{Pr}_{(e')}$, определяют, по существу, одну и ту же *неприводимую возможностьную модель испытаний*, которая полностью характеризует их в терминах формализма теории возможностей, гипотеза $\Pr \in \mathbb{Pr}_{(e')}$ определяет вероятностную модель испытаний как класс вероятностных пространств $(\Omega, \mathcal{P}(\Omega), \Pr)$, $\Pr \in \mathbb{Pr}_{(e')}$, но такая модель, вообще говоря, *не позволяет использовать стандартный формализм теории вероятностей для интерпретации исходов испытаний*.

3.3.1. Эмпирическое построение $\Pr$ -измеримой возможности, $\Pr \in \mathbb{Pr}$. Рассмотрим задачу эмпирического восстановления возможности P , максимально согласованной с вероятностью $\Pr$ в случае, когда $\Omega = \{\omega_1, \omega_2, \dots\}$, априори известна упорядоченность вероятностей

$$pr_1 \geq pr_2 \geq \dots > 0, \quad pr_1 + pr_2 + \dots = 1, \quad pr_i = \Pr(\{\omega_i\}), i = 1, 2, \dots, \quad (3.1)$$

и с ней согласована упорядоченность возможностей

$$1 = p_1 \geq p_2 \geq \dots > 0, \quad p_i = P(\{\omega_i\}), \quad i = 1, 2, \dots \quad (3.2)$$

В § 3.2.4 показано, что конкретная упорядоченность возможностей в (3.2) при условии $\Pr \approx > P$ определяется следующими условиями, связывающими ее со значениями вероятностей в (3.1),

$$\begin{aligned} e_j = 1 &\iff p_j > p_{j+1} \iff f_j \stackrel{\text{def}}{=} pr_1 + \dots + pr_{j-1} + 2pr_j > 1; \\ e_j = 0 &\iff p_j = p_{j+1} \iff f_j \leq 1, \quad j = 1, 2, \dots \end{aligned} \quad (3.3)$$

Задача состоит в том, чтобы, наблюдая за исходами испытаний, модель которых определена как $(\Omega, \mathcal{P}(\Omega), \Pr)^n$, $n = 1, 2, \dots$, решить, какие из

¹⁾ С точностью до эквивалентности. Напомним, что каждая возможность $P \in \mathbb{Pr}_{(e)}$ называется стохастически $\mathbb{Pr}_{(e)}$ - или $\Pr$ -измеримой, где $\Pr$ — любая вероятность из $\mathbb{Pr}_{(e)}$, класс $\mathbb{Pr}_{(e)}$ эквивалентных возможностей называется стохастически $\mathbb{Pr}_{(e)}$ -измеримым и максимально согласованным с классом $\mathbb{Pr}_{(e)}$, $e \in (0,1)$.

неравенств в (3.3) выполняются для вероятностей в (3.1), тем самым определить фактическую упорядоченность возможностей в (3.2), а, следовательно, — восстановить *возможностную модель испытаний*.

Разумеется, в случае счетного $\Omega = \{\omega_1, \omega_2, \dots\}$ эмпирически в (3.2) может быть восстановлена фактическая упорядоченность возможностей только конечного числа элементарных событий.

Определение 3.3.1. Возможность $P(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ условимся называть эмпирически определенной (безошибочно на основе данных п. н. конечного числа испытаний), если для каждого $s = 1, 2, \dots$ эмпирически определено (безошибочно на основе данных п. н. конечного числа испытаний) сужение¹⁾ $[P]_s(\cdot) : \sigma(\omega_1, \dots, \omega_s) \rightarrow [0, 1]$ возможности P на $\sigma(\omega_1, \dots, \omega_s)$.

Заметим, что если возможность $[P]_s$ известна, то для любых $A, B \in \mathcal{P}(\Omega)$, таких, что $A \cap \{\omega_1, \dots, \omega_s\} \neq \emptyset$, $B \cap \{\omega_1, \dots, \omega_s\} \neq \emptyset$, известно, какое из отношений $P(A) > P(B)$, $P(A) = P(B)$ или $P(A) < P(B)$ выполнено, а если $A \cap \{\omega_1, \dots, \omega_s\} \neq \emptyset$, а $B \cap \{\omega_1, \dots, \omega_s\} = \emptyset$, то $P(A) \geq P(B)$, наконец, если $A, B \subset \Omega \setminus \{\omega_1, \dots, \omega_s\}$, то согласно (3.2) известно лишь какое из неравенств $P(A) \geq P(B)$ или $P(A) \leq P(B)$ выполнено, причем для любого $A \in \sigma(\omega_1, \dots, \omega_s)$ $P(A) \geq P(\Omega \setminus \{\omega_1, \dots, \omega_s\})$.

Пусть $\nu_j^{(n)} \stackrel{\text{def}}{=} \nu^{(n)}(\{\omega_j\})$ — частота элементарного события $\{\omega_j\}$ в последовательности n взаимно независимых испытаний, $j = 1, 2, \dots$, $n = 1, 2, \dots$. Так как для каждого $j = 1, 2, \dots$ при $n \rightarrow \infty$ $\nu_j^{(n)} \xrightarrow{\text{п.н.}} \text{pr}_j = \text{Pr}(\{\omega_j\})$, то для любого $m = 1, 2, \dots$ и любого $\varepsilon > 0$ можно указать номер $N = N(\varepsilon, m)$, такой, что для всех $n \geq N(\varepsilon, m)$

$$\text{Pr}^{(\infty)}(\{ |\nu_1^{(n)} - \text{pr}_1| < \varepsilon, \dots, |\nu_m^{(n)} - \text{pr}_m| < \varepsilon \}) = 1. \quad (3.4)$$

Поскольку в (3.3) $f_j = f_j(\overline{\text{pr}}_{(j)}) = \text{pr}_1 + \dots + \text{pr}_{j-1} + 2\text{pr}_j$, где $\overline{\text{pr}}_{(j)} = (\text{pr}_1, \dots, \text{pr}_j)$, есть непрерывная функция в каждой точке $\overline{\text{pr}}_{(j)} \in \{(\text{pr}_1, \dots, \text{pr}_j), 1 > \text{pr}_1 + \dots + \text{pr}_j > \text{pr}_1 \geq \dots \geq \text{pr}_j \geq 0\}$, $j = 1, \dots, m$, то согласно (3.4) для всякого $\varepsilon > 0$ и любого $m = 1, 2, \dots$ можно указать номер $\overline{N}(\varepsilon, m)$, такой, что для всех $n \geq \overline{N}(\varepsilon, m)$

$$\text{Pr}^{(\infty)}(\{ \max_{1 \leq j \leq m} |\widehat{f}_j^{(n)} - f_j| < \varepsilon \}) = 1, \quad (3.5)$$

где

$$\widehat{f}_j^{(n)} \stackrel{\text{def}}{=} \nu_j^{(n)} + \dots + \nu_{j-1}^{(n)} + 2\nu_j^{(n)}, \quad j = 1, \dots, m, \quad m = 1, 2, \dots, \quad n = 1, 2, \dots \quad (3.6)$$

Этот факт при дополнительном предположении *регулярности* распределения (3.1) позволит охарактеризовать асимптотически при $n \rightarrow \infty$ решение задачи эмпирического восстановления возможности.

¹⁾ $\sigma(\omega_1, \dots, \omega_s)$ — минимальная подалгебра $\mathcal{P}(\Omega)$, содержащая $\{\omega_1\}, \dots, \{\omega_s\}$.

Определение 3.3.2. Вероятность Pr и ее распределение, удовлетворяющее условию (3.1), назовем регулярными, если

$$f_j = \text{pr}_1 + \dots + \text{pr}_{j-1} + 2\text{pr}_j \neq 1, \quad j = 1, 2, \dots, \quad (3.7)$$

а если в (3.1) допустить равенство вероятностей нулю, т. е. если считать, что вместо (3.1)

$$\text{pr}_1 \geq \text{pr}_2 \geq \dots \geq 0, \quad \text{pr}_1 + \text{pr}_2 + \dots = 1, \quad (3.1^*)$$

то вместо (3.7) условие регулярности запишем в следующем виде:

если как в (3.1) $\text{pr}_j > 0, j = 1, 2, \dots$, то $f_j \neq 1, j = 1, 2, \dots$;

если существует $j \in \{2, 3, \dots\}$, такое, что $\text{pr}_1 \geq \dots \geq \text{pr}_{j-1} > 0 = \text{pr}_j = \text{pr}_{j+1} = \dots$, то

$$f_1 \neq 1, \dots, f_{j-1} \neq 1 = f_j = f_{j+1} = \dots \quad (3.7^*)$$

В частности, если $\Omega = \{\omega_1, \dots, \omega_m\}$, то условие регулярности Pr сформулируем так:

если $\text{pr}_m > 0$, то ¹⁾ $f_1 \neq 1, \dots, f_{m-1} \neq 1$;

если существует $j \in \{2, \dots, m\}$, такое, что $\text{pr}_{j-1} > 0 = \text{pr}_j = \dots = \text{pr}_m$, то $f_1 \neq 1, \dots, f_{j-1} \neq 1 = f_j = \dots = f_m$.

Теорема 3.3.1. Если распределение вероятностей элементарных событий удовлетворяет условиям (3.1), (3.7), то возможность восстанавливается безошибочно на основе результатов п. н. конечного числа взаимно независимых испытаний.

Доказательство. Прежде всего заметим, что при условиях теоремы речь идет о выборе в (3.3) одной из альтернатив $f_j < 1$, или $f_j > 1$ для каждого $j = 1, 2, \dots$, основанном на результатах испытаний. А поскольку при любой из альтернатив значение f_j принадлежит одному из интервалов $(0, 1)$ или ²⁾ $(1, 2)$ вместе с некоторой окрестностью, то согласно (3.5) вместе с f_j при достаточно большом $n = n_j$ и все $\hat{f}_j^{(n_j)}, \hat{f}_j^{(n_j+1)}, \dots$ в (3.6) п. н. принадлежат этому же интервалу, а, следовательно, вместе с $f_1, \dots, f_m$ соответствующим интервалам $(0, 1)$ или $(1, 2)$ п. н. принадлежат и все $\hat{f}_1^{(n(m))}, \dots, \hat{f}_m^{(n(m))}; \hat{f}_1^{(n(m)+1)}, \dots, \hat{f}_m^{(n(m)+1)}, \dots$, где $n(m) = \max_{1 \leq j \leq m} \{n_j\}$. Это, в свою очередь, означает, что для любого $m = 1, 2, \dots$ на основе п. н. конечного числа $n \geq n(m)$ испытаний безошибочно восстанавливается упорядоченность $\text{pr}_j > \text{pr}_{j+1}$ или $\text{pr}_j = \text{pr}_{j+1}$ для $j = 1, \dots, m$. ■

Замечание 3.3.1. Условия (3.1), (3.7) в теореме 3.3.1 можно заменить на условия (3.1*), (3.7*). Действительно, если для некоторого j $\text{pr}_{j-1} > 0 = \text{pr}_j = \text{pr}_{j+1} = \dots$, то речь идет о выборе одной из альтернатив $f_i < 1$ или $f_i > 1$ для каждого $i = 1, \dots, j-1$, ибо в этом случае $\hat{f}_j^{(n)} = f_j = 1, \hat{f}_{j+1}^{(n)} = 1, \dots$ для всех $n = 1, 2, \dots$

¹⁾ В этом случае $f_m = \text{pr}_1 + \dots + \text{pr}_{m-1} + 2\text{pr}_m = 1 + \text{pr}_m > 1$.

²⁾ $f_j = \text{pr}_1 + \dots + \text{pr}_{j-1} + 2\text{pr}_j < 1 + \text{pr}_j \leq 1 + 1/j < 2$

Замечание 3.3.2. В случае нерегулярного распределения вероятностей в (3.1), например, если на самом деле $\text{pr}_j > 0$ и $f_j = 1$, то для любого $\varepsilon > 0$ можно указать $N(\varepsilon, j)$, такое, что для всех $n > N(\varepsilon, j)$ $1 - \varepsilon < \hat{f}_j^{(n)} < 1 + \varepsilon$, т. е. все $\hat{f}_j^{(n)}$, $n = 1, 2, \dots$, исключая, быть может, конечное их число, п.н. попадают в сколь угодно малую окрестность единицы, и на этом основании нельзя выбрать одну из альтернатив $f_j \leq 1$ или $f_j > 1$.

3.3.2. Эмпирическое построение $\text{Pr}^1 - \dots, \text{Pr}^k$ -измеримой возможности, $\text{Pr}^s \in \mathbb{P}_{(e)}$, $s = 1, \dots, k$. Пусть в бесконечной последовательности взаимно независимых испытаний, модель которых определена как $(\Omega, \mathcal{P}(\Omega), \text{Pr}_1) \times (\Omega, \mathcal{P}(\Omega), \text{Pr}_2) \times \dots$, среди вероятностей Pr_i , $i = 1, 2, \dots$, конечное число k различных, скажем, $\text{Pr}^1, \dots, \text{Pr}^k$, и $A \in \mathcal{P}(\Omega)$ — некоторое событие. Тогда при любом $n = 1, 2, \dots$ в 3.Б.Ч. (0.2)

$$\frac{1}{n} \sum_{i=1}^n \text{Pr}_i(A) = \sum_{s=1}^k (n_s/n) \text{Pr}^s(A) \in [\min_{1 \leq s \leq k} \text{Pr}^s(A), \max_{1 \leq s \leq k} \text{Pr}^s(A)],$$

где n_s/n — частота, с которой вероятность $\text{Pr}^s(A)$ встречается в последовательности $\text{Pr}_1(A), \dots, \text{Pr}_n(A)$, $s = 1, \dots, k$, $n_1 + \dots + n_k = n$. Поскольку при $n \rightarrow \infty$ частоты n_s/n , $s = 1, \dots, k$, принимают, вообще говоря, произвольные значения¹⁾ из $[0, 1]$, удовлетворяющие лишь условию $n_1/n + \dots + n_k/n = 1$, то при $n = 1, 2, \dots$ значение $\text{Pr}_{(n)}(A) \stackrel{\text{def}}{=} \sum_{s=1}^k (n_s/n) \text{Pr}^s(A)$ произвольно «блуждает» по отрезку $[\min_{1 \leq s \leq k} \text{Pr}^s(A), \max_{1 \leq s \leq k} \text{Pr}^s(A)]$, и согласно равенству (0.2) предисловия с увеличением n за ним все более точно следует частота $\nu^{(n)}(A)$. В этом случае знание вероятностей $\text{Pr}^1(A), \dots, \text{Pr}^k(A)$ не позволяет оценить частоту $\nu^{(n)}(A)$, а по наблюдениям за частотами $\nu^{(n)}(A)$, $n = 1, 2, \dots$, $A \in \mathcal{A}$, не может быть восстановлена вероятностная модель испытаний, но при условии $\exists e \in (0, 1) \text{Pr}^s \in \mathbb{P}_{(e)}$, $s = 1, \dots, k$, может быть восстановлена возможность модель испытаний, причем безошибочно на основании результатов п.н. конечного числа испытаний, если $\text{Pr}^1, \dots, \text{Pr}^k$ — регулярные вероятности.

Действительно, рассмотрим последовательность $(\Omega, \mathcal{A}, \text{Pr}_1) \times (\Omega, \mathcal{A}, \text{Pr}_2) \times \dots$ взаимно независимых испытаний, в моделях

¹⁾ Если вероятности $\text{Pr}^1(A), \dots, \text{Pr}^k(A)$ встречаются в последовательности $\text{Pr}_1(A), \text{Pr}_2(A), \dots$ с определенной закономерностью, например, с вероятностями $\text{pr}^1, \dots, \text{pr}^k$ (байесовская модель!), то при $n \rightarrow \infty$ $n_s/n \xrightarrow{\text{п.н.}} \text{pr}^s$, $s = 1, \dots, k$, и $\sum_{s=1}^k (n_s/n) \text{Pr}^s(A) \xrightarrow{\text{п.н.}} \sum_{s=1}^k \text{Pr}^s(A) \text{pr}^s$. В этом случае при $n \rightarrow \infty$ $\nu^{(n)}(A) \xrightarrow{\text{п.н.}} \text{Pr}(A) \triangleq \sum_{s=1}^k \text{Pr}^s(A) \text{pr}^s$.

$(\Omega, \mathcal{A}, \text{Pr}_i)$, $i = 1, 2, \dots$, которых $\Omega = \{\omega_1, \omega_2, \dots\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ и среди вероятностей $\text{Pr}_1, \text{Pr}_2, \dots$ конечное число k различных $\text{Pr}^1, \dots, \text{Pr}^k$, заданных распределениями $\text{pr}_i^s \stackrel{\text{def}}{=} \text{Pr}^s(\{\omega_i\})$, $s = 1, \dots, k$, $i = 1, 2, \dots$, удовлетворяющими условиям

$$\text{pr}_1^s \geq \text{pr}_2^s \geq \dots > 0, \text{pr}_1^s + \text{pr}_2^s + \dots = 1, \quad (3.8)$$

для каждого $s = 1, \dots, k$. Предположим, что все вероятности $\text{Pr}^1, \dots, \text{Pr}^k$ содержатся в некотором классе $\mathbb{P}_{(e)}$, P — любая стохастически $\mathbb{P}_{(e)}$ -измеримая возможность из класса $\mathbb{P}_{(e)}$, *максимально согласованная с каждой вероятностью $\text{Pr}^1, \dots, \text{Pr}^k$* . Конкретную упорядоченность распределения $p_i = P(\{\omega_i\})$, $i = 1, 2, \dots$,

$$1 = p_1 \geq p_2 \geq \dots > 0, \quad (3.9)$$

или, иначе говоря, — число $e = 0.e_1e_2\dots \in (0, 1)$, надлежит определить на основе результатов взаимно независимых испытаний, каждое из которых контролируется одной из вероятностей $\text{Pr}^1, \dots, \text{Pr}^k$.

Максимальная согласованность распределения (3.9) со всеми k распределениями в (3.8) определяется следующими условиями

$$\begin{aligned} e_i = 1 &\iff p_i > p_{i+1} \iff \forall s = 1, \dots, k \ f_i^s \stackrel{\text{def}}{=} \text{pr}_1^s + \dots + \text{pr}_{i-1}^s + 2\text{pr}_i^s > 1; \\ e_i = 0 &\iff p_i = p_{i+1} \iff \forall s = 1, \dots, k \ f_i^s \leq 1, \ i = 1, 2, \dots \end{aligned} \quad (3.10)$$

При эмпирическом восстановлении возможности $P \in \mathbb{P}_{(e)}$, $e = 0.e_1e_2\dots$, условия $f_i^1 > 1, \dots, f_i^k > 1$, либо $f_i^1 \leq 1, \dots, f_i^k \leq 1$, $i = 1, 2, \dots$, в (3.10), эквивалентные условию включения вероятностей $\text{Pr}^1, \dots, \text{Pr}^k$ в класс $\mathbb{P}_{(e)}$, играют такую же роль, как условие неизменности вероятности при ее эмпирическом восстановлении по выборке, контролируемой этой вероятностью.

Лемма 3.3.1. Пусть $\text{pr}_i = \sum_{s=1}^k \lambda_s \text{pr}_i^s$, $i = 1, 2, \dots$, где $\lambda_1 \geq 0, \dots, \lambda_k \geq 0$, $\lambda_1 + \dots + \lambda_k = 1$. Тогда $f_j = \text{pr}_1 + \dots + \text{pr}_{j-1} + 2\text{pr}_j > 1$ (≤ 1), если для всех $s = 1, \dots, k$ соответственно $f_j^s = \text{pr}_1^s + \dots + \text{pr}_{j-1}^s + 2\text{pr}_j^s > 1$ (≤ 1), $j = 1, 2, \dots$.

Доказательство следует из равенств $f_j = \sum_{s=1}^k \lambda_s f_j^s$, $j = 1, 2, \dots$ ■

Заметим, что $\text{pr}_i = \sum_{s=1}^k \lambda_s \text{pr}_i^s$, $i = 1, 2, \dots$, при любых $\lambda_s \geq 0$, $s = 1, \dots, k$, $\lambda_1 + \dots + \lambda_k = 1$, является распределением $\text{Pr}(\cdot) = \sum_{s=1}^k \lambda_s \text{Pr}^s(\cdot)$, и если $\text{Pr}^s \in \mathbb{P}_{(e)}$, $s = 1, \dots, k$, то, в силу выпуклости класса $\mathbb{P}_{(e)}$, и $\text{Pr} \in \mathbb{P}_{(e)}$.

Введем обозначения

$$\text{pr}_j^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \text{Pr}_i(\{\omega_j\}) = \sum_{s=1}^k \frac{n_s}{n} \text{pr}_j^s, \quad j = 1, 2, \dots, \quad (3.11)$$

где n_s/n – частота, с которой вероятность Pr^s встречается в последовательности $\text{Pr}_1, \text{Pr}_2, \dots, \text{Pr}_n$, $n_1 + \dots + n_k = n$,

$$\nu_j^{(n)} \stackrel{\text{def}}{=} \nu^{(n)}(\omega_j) = \sum_{s=1}^k \frac{n_s}{n} \nu_j^{(n_s)}, \quad j = 1, 2, \dots, \quad (3.12)$$

где $\nu_j^{(n_s)}$ – частота, с которой ω_j встречается в n_s испытаниях, исход которых контролируется вероятностью Pr^s , $s = 1, \dots, k$.

Так как для любых $\varepsilon > 0$, k и m существует число $N = N(\varepsilon, k, m)$, такое, что для всех $s = 1, \dots, k$, $j = 1, \dots, m$, для всех $n \geq N(\varepsilon, k, m)$ $|\nu_j^{(n_s)} - \text{pr}_j^s| \stackrel{\text{п.н.}}{<} \varepsilon$, то для любых $\varepsilon > 0$, $j = 1, \dots, m$ и для всех $n \geq N(\varepsilon, k, m)$

$$|\nu_j^{(n)} - \text{pr}_j^{(n)}| \leq \sum_{s=1}^k (n_s/n) |\nu_j^{(n_s)} - \text{pr}_j^s| \stackrel{\text{п.н.}}{<} \varepsilon \quad (3.13)$$

Наконец, поскольку функция

$$\begin{aligned} f_j(\overline{\text{pr}}_{(j)}) &\stackrel{\text{def}}{=} \text{pr}_1 + \dots + \text{pr}_{j-1} + 2\text{pr}_j, \quad \overline{\text{pr}}_{(j)} \stackrel{\text{def}}{=} (\text{pr}_1, \dots, \text{pr}_j) \in \\ &\in \mathcal{D}_j \stackrel{\text{def}}{=} \{(\text{pr}_1, \dots, \text{pr}_j), \quad 1 > \text{pr}_1 \geq \dots \geq \text{pr}_j > 0\}, \quad j = 1, \dots, m, \end{aligned} \quad (3.14)$$

равномерно непрерывна на множестве $\overline{\text{pr}}_{(j)} = \{\sum_{s=1}^k \lambda_s \overline{\text{pr}}_{(j)}^s, \quad \lambda_s \geq 0, \quad s = 1, \dots, k, \quad \lambda_1 + \dots + \lambda_k = 1\} \subset \mathcal{D}_j$, то для любых $\varepsilon > 0$ и $j = 1, \dots, m$ существует число $N(\varepsilon, m)$, такое, что п. н. для всех $n \geq N(\varepsilon, m)$

$$\max_{1 \leq j \leq m} |\widehat{f}_j^{(n)} - f_j^{(n)}| < \varepsilon, \quad m = 1, 2, \dots, \quad (3.15)$$

где (см. (3.11), (3.12))

$$\begin{aligned} f_j^{(n)} &\stackrel{\text{def}}{=} \text{pr}_1^{(n)} + \dots + \text{pr}_{j-1}^{(n)} + 2\text{pr}_j^{(n)}, \\ \widehat{f}_j^{(n)} &= \nu_1^{(n)} + \dots + \nu_{j-1}^{(n)} + 2\nu_j^{(n)}, \quad j = 1, 2, \dots \end{aligned} \quad (3.16)$$

Предположим, что при условиях (3.8) $\text{Pr}^1, \dots, \text{Pr}^k$ – регулярные вероятности, т. е. в (3.10) $f_j^s = \text{pr}_1^s + \dots + \text{pr}_{j-1}^s + 2\text{pr}_j^s \neq 1$ для всех $j = 1, 2, \dots$ и $s = 1, \dots, k$. Тогда вместо (3.10) *полный набор* различных упорядоченностей в (3.9) определится условиями

$$\begin{aligned} e_i = 1 &\iff \text{pr}_i > \text{pr}_{i+1} \iff \forall s = 1, \dots, k \quad f_i^s > 1; \\ e_i = 0 &\iff \text{pr}_i = \text{pr}_{i+1} \iff \forall s = 1, \dots, k \quad f_i^s < 1, \quad i = 1, 2, \dots, \end{aligned} \quad (3.17)$$

а так как

$$f_j^{(n)} = \sum_{s=1}^k (n_s/n) f_j^s, \quad j = 1, 2, \dots, \quad (3.18)$$

то согласно лемме 3.3.1 и эквивалентностям¹⁾ в (3.17),

$$\text{либо } p_i > p_{i+1} \Leftrightarrow \forall s = 1, \dots, k, \quad f_i^s > 1 \Leftrightarrow f_i^{(n)} > 1,$$

$$\text{либо } p_i = p_{i+1} \Leftrightarrow \forall s = 1, \dots, k, \quad f_i^s < 1 \Leftrightarrow f_i^{(n)} < 1, \quad i = 1, 2, \dots,$$

причём согласно (3.18), если $f_j^s < 1$, $s = 1, \dots, k$, то $\forall n = 1, 2, \dots$ $f_j^{(n)} \leq \max_{1 \leq s \leq k} f_j^s < 1$, а если $f_j^s > 1$, $s = 1, \dots, k$, то $\forall n = 1, 2, \dots$ $f_j^{(n)} \geq \min_{1 \leq s \leq k} f_j^s > 1$, $j = 1, 2, \dots$. Иначе говоря, для любого $n = 1, 2, \dots$

в (3.16) $f_j^{(n)}$ вместе с независимой от n окрестностью либо строго левее единицы, либо строго правее. А отсюда и из (3.15) следует, что существует номер $N(m)$, такой, что для всех $n \geq N(m)$ в (3.16) п. н. либо $\hat{f}_j^{(n)} < 1$, либо $\hat{f}_j^{(n)} > 1$ для каждого $j = 1, \dots, m$.

Верно и обратное утверждение, а именно: если существует номер N , начиная с которого для всех $n = N, N+1, \dots$ п. н. либо $\hat{f}_j^{(n)} < 1$, либо $\hat{f}_j^{(n)} > 1$ для каждого $j = 1, \dots, m$, то соответственно либо $f_j^{(n)} < 1$, либо $f_j^{(n)} > 1$, $j = 1, \dots, m$, $n \geq N$.

Эти рассуждения позволяют сформулировать следующий результат.
Теорема 3.3.2. Пусть в последовательности $(\Omega, \mathcal{P}(\Omega), \text{Pr}_1)$, $(\Omega, \mathcal{P}(\Omega), \text{Pr}_2), \dots$ взаимно независимых испытаний среди вероятностей $\text{Pr}_1, \text{Pr}_2, \dots$ конечное число различных $\text{Pr}^1, \dots, \text{Pr}^k$. Если существует возможность P , максимально согласованная с вероятностями $\text{Pr}^1, \dots, \text{Pr}^k$, каждая из которых распределена согласно (3.8) и регулярна, то возможностьная модель испытаний $(\Omega, \mathcal{P}(\Omega), P)$ восстанавливается безошибочно на основании результатов п. н. конечного числа испытаний.

Доказательство. может быть проведено по схеме доказательства теоремы 3.3.1. ■

Замечание 3.3.3. Если вместо (3.8) исходить из более слабого условия упорядоченности

$$\text{pr}_1^s \geq \text{pr}_2^s \geq \dots \geq 0, \quad \text{pr}_1^s + \text{pr}_2^s + \dots = 1, \quad s = 1, \dots, k, \quad (3.19)$$

то условия регулярности вероятностей $\text{Pr}^1, \dots, \text{Pr}^k$ должны формулироваться следующим образом:

¹⁾ Импликации $\forall s = 1, \dots, k \quad f_i^s < 1 (> 1) \Rightarrow f_i^{(n)} < 1 (> 1)$ очевидны, обратные импликации $f_i^{(n)} < 1 (> 1) \Rightarrow \forall s = 1, \dots, k \quad f_i^s < 1 (> 1)$ выполнены, поскольку априори все $f_i^1, \dots, f_i^k$ для каждого $i = 1, 2, \dots$ либо < 1 , либо > 1 .

$$\begin{aligned}
&\text{либо } \text{pr}_j^s > 0, f_j^s \neq 1, \text{ для всех } s = 1, \dots, k, j = 1, 2, \dots, \\
&\text{либо для некоторого } q \text{ } \text{pr}_1^s \geq \dots \geq \text{pr}_q^s > 0 = \text{pr}_{q+1}^s = \dots, \\
&f_1^s \neq 1, \dots, f_q^s \neq 1 = f_{q+1}^s = \dots, \text{ для всех } s = 1, \dots, k. \quad (3.20)
\end{aligned}$$

Теорема 3.3.2 останется справедливой и в этом случае, поскольку согласно второй группе условий в (3.20) $1 = f_{q+1}^s = f_{q+2}^s = \dots, s = 1, \dots, k, \Rightarrow 1 = f_{q+1}^{(n)} = f_{q+2}^{(n)} = \dots \Rightarrow 1 = \hat{f}_{q+1}^{(n)} = \hat{f}_{q+2}^{(n)} = \dots, n = 1, 2, \dots$. Более того, теорема 3.3.2 будет справедлива, если в 3.20 допустить зависимость q от $s = 1, \dots, k$.

3.3.3. [. Алгоритм эмпирического восстановления возможности, максимально согласованной с вероятностью, не изменяющейся в процессе испытаний] Алгоритм эмпирического восстановления возможности, максимально согласованной с вероятностью, не изменяющейся в процессе испытаний [14]

В этом параграфе рассмотрен алгоритм эмпирического восстановления возможности P , максимально согласованной с вероятностью $\text{Pr} \in \mathbb{P}_r$, основанный на результатах взаимно независимых испытаний, в модели которых

$$(\Omega^n, \mathcal{P}(\Omega^n), \text{Pr}^{(n)}) = (\Omega, \mathcal{P}(\Omega), \text{Pr}) \times \dots \times (\Omega, \mathcal{P}(\Omega), \text{Pr}), \quad n = 1, 2, \dots, \quad (3.21)$$

вероятность $\text{Pr}^{(n)} = \text{Pr} \times \dots \times \text{Pr}$ неизвестна, но включение $\text{Pr} \in \mathbb{P}_r$ означает, что априори $\text{pr}_1 \geq \text{pr}_2 \geq \dots, \text{pr}_i = \text{Pr}(\{\omega_i\}), i = 1, 2, \dots$, см. § 3.7 Приложение.

Речь идет о задаче эмпирического восстановления класса $\mathbb{P}_{r(e)} \subset \mathbb{P}_r$ вероятностей, содержащего вероятность Pr , определяющую модель испытаний (3.21), поскольку класс $\mathbb{P}_{r(e)}$ определяет класс $\mathbb{P}_{(e)}$ взаимно эквивалентных возможностей P и, следовательно, — класс $(\Omega, \mathcal{P}(\Omega), P), P \in \mathbb{P}_{(e)}$ взаимно эквивалентных возможностных моделей испытаний.

Так как для любой возможности $P \in \mathbb{P}_{(e)}$, максимально согласованной с вероятностью ¹⁾ $\text{Pr} \in \mathbb{P}_{r(e)}$, упорядоченность возможностей $p_i = P(\{\omega_i\}), i = 1, 2, \dots$, определяется условиями

$$\text{либо } e_i = 1 \Leftrightarrow p_i > p_{i+1} \Leftrightarrow f_i = \text{pr}_1 + \dots + \text{pr}_{i-1} + 2\text{pr}_i > 1, \quad (3.22)$$

$$\text{либо } e_i = 0 \Leftrightarrow p_i = p_{i+1} \Leftrightarrow f_i < 1, \quad i = 1, 2, \dots, \quad (3.23)$$

то задача эмпирического восстановления возможности согласно определению (3.3.1) сводится к задаче теории статистических решений, в которой на основе значений частот $\nu_i^{(n)}, i = 1, 2, \dots, s$, элементарных событий $\{\omega_i\}, i = 1, 2, \dots, s$, наблюдаемых в последовательности n вза-

¹⁾ Вероятность Pr предполагается регулярной, т.е. $f_i \neq 1, i = 1, 2, \dots$

имно независимых испытаний, $n = 1, 2, \dots$, модель которых определена в (3.21), для всех $i = 1, \dots, s$ требуется принять одну из гипотез (3.22) или (3.23).

В связи с этой задачей рассмотрим следующий *алгоритм проверки гипотез* (3.22), (3.23): для всех $i = 1, \dots, s$ и каждого $n = 1, 2, \dots$

- если $\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}$, то $\square_1$: считать $f_i > 1$;
- если $\hat{f}_i^{(n)} < 1 - \delta^{(n,s)}$, то $\square_2$: считать $f_i < 1$;
- если $|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}$, то $\circ$: продолжить испытания,

где

$$\hat{f}_i^{(n)} \stackrel{\text{def}}{=} \nu_1^{(n)} + \dots + \nu_{i-1}^{(n)} + 2\nu_i^{(n)}, \quad i = 1, 2, \dots \quad (3.25)$$

В алгоритме (3.24) при каждом $n = 1, 2, \dots$ проверяются условия *решений* $\square_1$, $\square_2$ и $\circ$ для всех $i = 1, \dots, s$; если при этом приняты только решения $\square_1$ или $\square_2$, то алгоритм завершен, если же хотя бы для одного i принято решение $\circ$, то после каждого дополнительного испытания для всех $i = 1, \dots, s$ проверяются условия решений $\square_1$, $\square_2$ и $\circ$, ранее принятые решения корректируются, $\dots$, и так — до тех пор, пока, наконец, алгоритм будет завершен.

Значения $\delta^{(n,s)}$, $n = 1, 2, \dots$, в (3.24) определим, задав верхние границы $\alpha^{(s)}$, $s = 2, 3, \dots$, вероятностей ошибочных решений $\square_1$ и $\square_2$ и воспользовавшись *неравенством Хёфдинга* [10, 27]

$$\Pr^{(n)}(\zeta_n - M\zeta_n > n\varepsilon) \leq \exp(-2n^2\varepsilon^2 / \sum_{k=1}^n (b_k - a_k)^2), \quad (3.26)$$

в котором $\zeta_n = \sum_{k=1}^n \xi_k$, случайные величины $\xi_1, \dots, \xi_n$ взаимно независимы, $\Pr^{(n)}(a_k \leq \xi_k \leq b_k) = 1$, $k = 1, \dots, n$, и $\varepsilon > 0$ — любое.

Вероятность ошибочного решения $\square_1$, при котором наблюдается значение $\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}$, но на самом деле $f_i < 1$,

$$\begin{aligned} \Pr^{(n)}(\{\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}\} | f_i < 1) &\leq \Pr^{(n)}(\{\hat{f}_i^{(n)} - f_i > \delta^{(n,s)}\} | f_i < 1) \leq \\ &\leq \exp(-n(\delta^{(n,s)})^2 / 2) \stackrel{\text{def}}{=} \alpha^{(s)}, \quad i = 1, \dots, s. \end{aligned} \quad (3.27)$$

Первое неравенство в (3.27) обусловлено тем, что событие $\{\hat{f}_i^{(n)} - f_i > \delta^{(n,s)}\}$ является следствием события $\{\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}\}$, если $f_i < 1$, второе неравенство является вариантом неравенства Хёфдинга (3.26), поскольку согласно (3.25) в равенстве

$$\hat{f}_i^{(n)} = \frac{1}{n} \sum_{j=1}^n (\chi_1(\omega^{(j)}) + \dots + \chi_{i-1}(\omega^{(j)}) + 2\chi_i(\omega^{(j)}))$$

все слагаемые при $j = 1, 2, \dots, n$ взаимно независимы, принимают значения 0, 1 или 2, и $M^{(n)}\hat{f}_i^{(n)} = f_i$, $i = 1, 2, \dots$. Значение $\alpha^{(s)}$ в (3.27) оценивает сверху вероятность ошибочного решения $\square_1$ и определяет

$$\delta^{(n,s)} = \left(\frac{2}{n} \ln \frac{1}{\alpha^{(s)}} \right)^{\frac{1}{2}}, \quad n = 1, 2, \dots, \quad s = 1, 2, \dots \quad (3.28)$$

Аналогично, вероятность неверного решения $\square_2$

$$\Pr^{(n)}(\{\hat{f}_i^{(n)} < 1 - \delta^{(n,s)}\} | f_i > 1) \leq \Pr^{(n)}(\{\hat{f}_i^{(n)} - f_i < -\delta^{(n,s)}\} | f_i > 1) \leq \exp(-n(\delta^{(n,s)})^2/2) = \alpha^{(s)}.$$

Покажем, что событие $\{1 - \delta^{(n,s)} \leq \hat{f}_i^{(n)} \leq 1 + \delta^{(n,s)}\}$, приводящее в алгоритме (3.24) к решению $\circ$ о продолжении испытаний, при любом условии $f_i > 1$ или $f_i < 1$, $i = 1, \dots, s$, для каждого $s = 1, 2, \dots$ может выполняться лишь для п.н. конечного числа n испытаний.

Действительно, пусть $|f_i - 1| \geq 2\varepsilon_i > 0$, $i = 1, \dots, s$. Тогда для каждого $s = 1, 2, \dots$ и всех $n \geq n_i^{(s)} = \min\{k, \delta^{(k,s)} < \varepsilon_i\}$,

$$\begin{aligned} & \Pr^{(n)}(|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)} | |f_i - 1| \geq 2\varepsilon_i) \leq \\ & \leq \Pr^{(n)}(|f_i - 1| - |\hat{f}_i^{(n)} - 1| \geq 2\varepsilon_i - \delta^{(n,s)} | |f_i - 1| \geq 2\varepsilon_i) \leq \\ & \leq \Pr^{(n)}(|\hat{f}_i^{(n)} - f_i| \geq 2\varepsilon_i - \delta^{(n,s)}) \leq \Pr^{(n)}(|\hat{f}_i^{(n)} - f_i| \geq \varepsilon_i) = \\ & = \Pr^{(\infty)}(|\hat{f}_i^{(n)} - f_i| \geq \varepsilon_i) \leq 2 \exp(-n\varepsilon_i^2/2), \end{aligned}$$

и утверждение, согласно которому для каждого $s = 1, 2, \dots$ происходит п.н. конечное число событий $\{1 - \delta^{(n,s)} \leq \hat{f}_i^{(n)} \leq 1 + \delta^{(n,s)}\} = \{|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}\}$, $n = 1, 2, \dots$, следует теперь из леммы Бореля-Кантелли, так как $2 \sum_{n=n_i^{(s)}}^{\infty} \exp(-n\varepsilon_i^2/2) < \infty$, $i = 1, \dots, s$. Поэтому, для любого

$s = 1, 2, \dots$, и любого $i = 1, \dots, s$ существует $N_i^{(s)}$, такое, что для любого $n \geq N_i^{(s)}$ $\Pr^{(n)}(|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}) = 0$, а поскольку для любого $n = 1, 2, \dots$ $\Pr^{(n)}(\{\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}\}) + \Pr^{(n)}(\{\hat{f}_i^{(n)} < 1 - \delta^{(n,s)}\}) + \Pr^{(n)}(|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}) = 1$, то для любого $n \geq N_i^{(s)}$ $\Pr^{(n)}(\{\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}\}) + \Pr^{(n)}(\{\hat{f}_i^{(n)} < 1 - \delta^{(n,s)}\}) = 1$, и если одно из решений, например, $\square_1$, ошибочно, то $\Pr^{(n)}(\{\hat{f}_i^{(n)} > 1 + \delta^{(n,s)}\}) \leq \alpha^{(s)}$, при этом вероятность верного решения $\square_2$ $\Pr^{(n)}(\{\hat{f}_i^{(n)} < 1 - \delta^{(n,s)}\}) \geq 1 - \alpha^{(s)}$.

Теорема 3.3.3. Для любого $s = 1, 2, \dots$ алгоритм (3.24) согласно условиям (3.22), (3.23) на основе п.н. конечного числа испытаний восстанавливает упорядоченность возможностей элементарных событий $p_1, \dots, p_s$, совпадающую с истинной их упорядоченностью с вероятностью, большей $1 - s\alpha^{(s)} = 1 - \alpha$, если $\alpha^{(s)} = \alpha/s$, где α — априорная оценка вероятности ошибочного упорядочения возможностей s элементарных событий, $s = 1, 2, \dots$

Действительно, вероятность того, что среди s решений $\square_1, \square_2$, полученных по завершении алгоритма, по меньшей мере одно ошибочно, не больше $s\alpha^{(s)}$.

3.3.4. Алгоритм эмпирического восстановления возможности, максимально согласованной с вероятностью, изменяющейся в процессе испытаний. В этом параграфе модель n взаимно независимых испытаний определена как вероятностное пространство

$$(\Omega^{(n)}, \mathcal{P}(\Omega^{(n)}), \Pr_{1,\dots,n}^{(n)}) = (\Omega, \mathcal{P}(\Omega), \Pr_1) \times \dots \times (\Omega, \mathcal{P}(\Omega), \Pr_n), \quad (3.29)$$

в котором $\Pr_{1,\dots,n}^{(n)} = \Pr_1 \times \dots \times \Pr_n$, для всех $j = 1, 2, \dots, n$ вероятности $\Pr_j \in \mathbb{P}_{(e)}$ регулярные, и, следовательно, для каждого $i = 1, 2, \dots$

$$\text{либо } e_i = 1 \Leftrightarrow p_i > p_{i+1} \Leftrightarrow F_i^{(j)} = \Pr_j(\{\omega_1\}) + \dots + \Pr_j(\{\omega_{i-1}\}) + 2\Pr_j(\{\omega_i\}) > 1, \quad (3.30)$$

$$\text{либо } e_i = 0 \Leftrightarrow p_i = p_{i+1} \Leftrightarrow F_i^{(j)} < 1, \quad j = 1, \dots, n, \quad (3.31)$$

для $n = 1, 2, \dots$, ср. с (3.22). Значение $e = 0.e_1e_2\dots$, определяющее упорядоченность возможностей $p_i = P(\{\omega_i\})$, $i = 1, 2, \dots$, разумеется, неизвестно и должно быть определено на основе результатов испытаний.

В рассматриваемом случае проверяемые гипотезы определяются следующими условиями:

$$\text{либо } e_i = 1 \Leftrightarrow p_i > p_{i+1} \Leftrightarrow f_i^{(n)} \stackrel{\text{def}}{=} \text{pr}_1^{(n)} + \dots + \text{pr}_{i-1}^{(n)} + 2\text{pr}_i^{(n)} > 1, \quad (3.32)$$

$$\text{либо } e_i = 0 \Leftrightarrow p_i = p_{i+1} \Leftrightarrow f_i^{(n)} < 1, \quad i = 1, \dots, s, \quad (3.33)$$

где $\text{pr}_i^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n \Pr_j(\{\omega_i\})$, $i = 1, \dots, s$, суть "эмпирические вероятности" элементарных событий.

Так как модель испытаний такова, что условия (3.30), (3.31) $\mathbb{P}_{(e)}$ -измеримости восстанавливаемой возможности выполнены для каждого $i = 1, 2, \dots$ и всех $j = 1, 2, \dots$, то гипотезы (3.32), (3.33) суть следствия условий (3.30), (3.31), но, в отличие от последних, могут быть проверены эмпирически, поскольку при $n \rightarrow \infty$ $\widehat{f}_i^{(n)} - f_i^{(n)} \xrightarrow{\text{п.н.}} 0$, где $\widehat{f}_i^{(n)} \stackrel{\text{def}}{=} \nu_1^{(n)} + \dots + \nu_{i-1}^{(n)} + 2\nu_i^{(n)}$, $\nu_i^{(n)}$ — частота элементарного события $\{\omega_i\}$ в последовательности n испытаний, $i = 1, 2, \dots$.

Рассмотрим алгоритм решения задач проверки гипотез (3.32), (3.33), согласно которому для всех $i = 1, \dots, s$ и каждого $n = 1, 2, \dots$

- если $\widehat{f}_i^{(n)} > 1 + \delta^{(n,s)}$, то $\square_1$: считать $f_i^{(n)} > 1$,
- если $\widehat{f}_i^{(n)} < 1 - \delta^{(n,s)}$, то $\square_2$: считать $f_i^{(n)} < 1$,
- если $|\widehat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}$, то $\circ$: продолжить испытания.

(3.34)

Значения $\delta^{(n,s)}$, $n = 1, 2, \dots$, $s = 1, 2, \dots$ в (3.34) определим, как и в параграфе 3.3.3, задав соответственно верхние границы $\alpha^{(s)}$, $s = 1, 2, \dots$, вероятностей ошибочных решений $\square_1$ и $\square_2$. Так как

$$\widehat{f}_i^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n (\chi_1(\omega^{(j)}) + \dots + \chi_{i-1}(\omega^{(j)}) + 2\chi_i(\omega^{(j)})),$$

где слагаемые при $j = 1, 2, \dots, n$ взаимно независимы, принимают значения 0, 1 или 2 и $M_{1,\dots,n}^{(n)} \widehat{f}_i^{(n)} = f_i^{(n)}$, $i = 1, 2, \dots$, то, как и в параграфе 3.3.3, вероятности ошибочно принять решения $\square_1, \square_2$

$$\Pr_{1,\dots,n}^{(n)}(\{\widehat{f}_i^{(n)} > 1 + \delta^{(n,s)}\} | f_i^{(n)} < 1) \leq \Pr_{1,\dots,n}^{(n)}(\{\widehat{f}_i^{(n)} - f_i^{(n)} > \delta^{(n,s)}\} | f_i^{(n)} < 1) \leq \exp(-n(\delta^{(n,s)})^2/2) \stackrel{\text{def}}{=} \alpha^{(s)}, \quad (3.35)$$

$$\Pr_{1,\dots,n}^{(n)}(\{\widehat{f}_i^{(n)} < 1 - \delta^{(n,s)}\} | f_i^{(n)} > 1) \leq \exp(-n(\delta^{(n,s)})^2/2) = \alpha^{(s)}, \quad i = 1, \dots, s.$$

Поэтому и связь между $\alpha^{(s)}$ и $\delta^{(n,s)}$ определена, как и при неизменяющейся вероятности, равенством (3.28). Если же наблюдаемое значение

$$\widehat{f}_i^{(n)} \in \left[1 - \left(\frac{2}{n} \ln \frac{1}{\alpha^{(s)}} \right)^{\frac{1}{2}}, 1 + \left(\frac{2}{n} \ln \frac{1}{\alpha^{(s)}} \right)^{\frac{1}{2}} \right] \stackrel{\text{def}}{=} I^{(n,s)}, \quad (3.36)$$

то, согласно (3.34), принимается решение $\circ$, и испытания должны быть продолжены, причем при изменяющейся вероятности событие $\{\widehat{f}_i^{(n)} \in I^{(n,s)}\}$, вообще говоря, может выполняться для бесконечно многих $n = 1, 2, \dots$

Сформулируем требования к модели испытаний (3.29), гарантирующие, что произойдет п.н. конечное число событий (3.36).

Лемма 3.3.2. Пусть при всех достаточно больших n и всех $i = 1, \dots, s$ $|f_i^{(n)} - 1| \geq \delta^{(n,s)}(1 + \varepsilon_{n,s})$, где $\delta^{(n,s)}$ определены в (3.28), $\varepsilon_{n,s} > 0$ и удовлетворяют условиям $\sum_{n=1}^{\infty} (\alpha^{(s)})^{\varepsilon_{n,s}^2} < \infty$, в которых $\alpha^{(s)} = \alpha/s$, $s = 1, 2, \dots$. Тогда для каждого $s = 1, 2, \dots$ происходит п.н. конечное число событий (3.36).

Доказательство. Так как при условии $|f_i^{(n)} - 1| \geq \delta^{(n,s)}(1 + \varepsilon_{n,s})$, $\varepsilon_{n,s} > 0$, выполнены следующие включения событий

$$\begin{aligned} \{|\widehat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}\} &\subset \{|f_i^{(n)} - 1| - |\widehat{f}_i^{(n)} - 1| \geq \delta^{(n,s)}(1 + \varepsilon_{n,s}) - \delta^{(n,s)} = \\ &= \varepsilon_{n,s}\delta^{(n,s)}\} \subset \{|\widehat{f}_i^{(n)} - f_i^{(n)}| \geq \varepsilon_{n,s}\delta^{(n,s)}\}, \end{aligned}$$

то

$$\Pr_{1,\dots,n}^{(n)}(\{|\widehat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}\} | |f_i^{(n)} - 1| \geq \delta^{(n,s)}(1 + \varepsilon_{n,s})) \leq \Pr^{(n)}(\{|\widehat{f}_i^{(n)} - f_i^{(n)}| \geq \varepsilon_{n,s}\delta^{(n,s)}\}) \leq 2 \exp(-n(\varepsilon_{n,s}\delta^{(n,s)})^2/2) = 2(\alpha^{(s)})^{\varepsilon_{n,s}^2}$$

и, следовательно,

$$\sum_{n=1}^{\infty} \Pr_{1,2,\dots}^{(\infty)}(\{|\hat{f}_i^{(n)} - 1| \leq \delta^{(n,s)}\}) \leq 2 \sum_{n=1}^{\infty} (\alpha^{(s)})^{\varepsilon_{n,s}^2} < \infty.$$

Последнее неравенство гарантирует, что происходит п.н. конечное число событий (3.36). ■

Если модель испытаний удовлетворяет условиям, сформулированным в лемме¹⁾ 3.3.2, то теорема 3.3.3 верна и в случае вероятности, изменяющейся от испытания к испытанию, если существует $e = 0.e_1e_2\dots$, для которого условия (3.30), (3.31) выполнены для всех $j = 1, 2, \dots$. Условия леммы 3.3.2, очевидно, выполнены, если среди вероятностей $\Pr_1, \Pr_2, \dots$ конечное число различных.

Замечание 3.3.4. Если $(\Omega, \mathcal{P}(\Omega), P, N)$ — нечеткая модель стохастического объекта, то задача эмпирического восстановления N сводится к задаче эмпирического восстановления P , поскольку согласно результатам, полученным в § 3.2 гл.2, $\Pr \approx \triangleright N$, если и только если $\Pr \approx \triangleright P$, и при этом $\exists \vartheta(\cdot) \in \Theta N(A) = \vartheta \circ P(\Omega \setminus A)$, $A \in \mathcal{P}(\Omega)$.

3.4. Экспертное восстановление возможности [19]

Речь пойдет о методах построения коллективной экспертной оценки *распределения возможностей значений* $\xi = x_j$, $j = 1, \dots, n$, нечеткого элемента ξ , неформальная модель которого в той или иной степени известна экспертам, или, что эквивалентно, см. § 3.1.3.2, об *эмпирическом построении пространства с возможностью* $(X, \mathcal{P}(X), P^\xi)$, $X = \{x_1, \dots, x_n\}$, для которого нечеткий элемент ξ канонический.

3.4.1. Восстановление распределения нечеткого элемента путем попарных сравнений возможностей его значений. Эксперты могут оценивать *возможности равенств* $\xi = x_1, \dots, \xi = x_n$ в своих шкалах, но поскольку решение, представляющее мнения всех экспертов, должно быть инвариантным относительно выбора их шкал, для построения коллективной экспертизы каждому эксперту следует определить упорядоченность значений

$$p_j \stackrel{\text{def}}{=} P(\xi = x_j), \quad j = 1, \dots, n, \quad (4.1)$$

где $\max_{1 \leq j \leq n} p_j = 1$, $\min_{1 \leq j \leq n} p_j \geq 0$, или, что то же самое, соответствующую $(n+1) \times (n+1)$ матрицу $m = m(p.)$ попарных сравнений, где $p.$ — кортеж $p_1, \dots, p_n$, $p_{n+1} = 0$, определяющую класс взаимно эквивалентных распределений ξ . Элементы матрицы

¹⁾ Условия, сформулированные в лемме, означают, что при $n \rightarrow \infty$ $\hat{f}_i^{(n)}$, $i = 1, \dots, s$, "могут" стремиться к единице, но не "слишком быстро".

$m(p.)$ кортежа $p. \sim p_1, \dots, p_n, p_{n+1} = 0$ определим равенствами

$$(m(p.))_{kj} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{если } p_k > p_j, \\ 0, & \text{если } p_k = p_j, \\ -1, & \text{если } p_k < p_j, \end{cases} \quad k, j = 1, \dots, n+1.$$

Рассмотрим задачу коллективного экспертного оценивания $m(p.)$. Пусть $p_j^{(i)}$ — возможность равенства $\xi = x_j$, $j = 1, \dots, n$, и $m^{(i)} \stackrel{\text{def}}{=} m(p_j^{(i)})$ — матрица попарных сравнений кортежа $p_1^{(i)}, \dots, p_n^{(i)}$, $p_{n+1} = 0$, определенная i -м экспертом, $i = 1, \dots, N$. Предлагаемое ниже решение $m_* \in M_{n+1} \stackrel{\text{def}}{=} \{m(p.), p. \sim p_1, \dots, p_n, p_{n+1} = 0, \max_{1 \leq j \leq n} p_j = 1, \min_{1 \leq j \leq n} p_j \geq 0\}$ определяет коллективную экспертизу, в известном смысле лучше других согласованную с мнениями всех экспертов и с точностью до эквивалентности определяющую искомое распределение ξ . Речь идет о решении m_* задачи

$$\sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, m_*) = \min_{m \in M_{n+1}} \sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, m), \quad (4.2)$$

где

$$\rho(a, b) \stackrel{\text{def}}{=} \left(\sum_{k,j=1}^{n+1} (a_{kj} - b_{kj})^2 \right)^{1/2}, \quad a = \{a_{kj}\}, b = \{b_{kj}\} \quad (4.3)$$

и w_i^2 — "вес" i -го эксперта, позволяющий учесть его "относительную компетентность", $i = 1, \dots, N$, $\sum_{i=1}^N w_i^2 = 1$.

Поскольку в (4.2), как нетрудно убедиться,

$$\sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, m) = \sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, \bar{m}) + \rho^2(\bar{m}, m), \quad (4.4)$$

где $\bar{m} = \sum_{i=1}^N w_i^2 m^{(i)}$, то задача (4.2) эквивалентна задаче отыскания матрицы $m_* \in M_{n+1}$, ближайшей (в смысле (4.3)) к матрице $\bar{m}$, которая, вообще говоря, не является матрицей попарных сравнений, ибо хотя $\bar{m}_{kj} = -\bar{m}_{jk} \in [-1, 1]$, но, возможно, $\bar{m}_{kj} \notin \{-1, 0, 1\}$, $k, j = 1, \dots, n+1$. В данном случае очевидно, что решением задачи $\rho^2(\bar{m}, m) \sim \min_{m \in M_{n+1}}$ является матрица m_* , матричные элементы которой суть

$$m_{*kj} = \begin{cases} 1, & \text{если } \bar{m}_{kj} > 1/2, \\ 0, & \text{если } |\bar{m}_{kj}| \leq 1/2, \\ -1, & \text{если } \bar{m}_{kj} < -1/2, \end{cases} \quad k, j = 1, \dots, n+1. \quad (4.5)$$

Матрица m_* представляет коллективную экспертизу N экспертов, но насколько этой экспертизе следует доверять позволит

судить значение первого слагаемого в правой части (4.4), если использовать вероятностную модель принятия решений экспертами, в которой $m^{(i)} \in M_{n+1}$, $i = 1, \dots, N$, суть значения *взаимно независимых случайных матриц*. Пусть $\omega = (m^{(1)}, \dots, m^{(N)})$, где на сей раз $m^{(i)}$, $i = 1, \dots, N$, — произвольные матрицы из M_{n+1} , Ω — множество всех $(q(n))^N$ таких ω , где $q(n)$ — число всех $(n+1) \times (n+1)$ матриц $m \in M_{n+1}$, $\Pr(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ — вероятность на $(\Omega, \mathcal{P}(\Omega))$, определенная своими значениями

$$\Pr(\{\omega\}) = (q(n))^{-N}, \quad \omega \in \Omega, \quad (4.6)$$

так, чтобы вероятностное пространство $(\Omega, \mathcal{P}(\Omega), \Pr)$ моделировало заключения N экспертов, принимающих решения наугад и *взаимно независимо*.

В этой модели первое слагаемое в правой части (4.4) следует рассматривать как одно из значений статистики

$$s(\omega) = \sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, \bar{m}), \quad \omega \in \Omega, \quad (4.7)$$

определенной на $(\Omega, \mathcal{P}(\Omega), \Pr)$ и принимающей $L = L(n)$ различных значений в множестве $S = \{s(\omega), \omega \in \Omega\} \stackrel{\text{def}}{=} \{s_1, \dots, s_L\}$ с вероятностями $\text{pr}_l \stackrel{\text{def}}{=} \Pr(\{\omega \in \Omega, s(\omega) = s_l\})$, $l = 1, \dots, L$, которые будем считать упорядоченными по убыванию, $\text{pr}_1 \geq \dots \geq \text{pr}_L > 0$, $\text{pr}_1 + \dots + \text{pr}_L = 1$. Если H — гипотеза, согласно которой эксперты принимают решения наугад и *взаимно независимо*, то критерий, *отвергающий H ошибочно* с вероятностью, не большей ε , где, например, $\varepsilon = 0,01$, определится критическим множеством $S_\varepsilon = \{s_{l(\varepsilon)}, \dots, s_L\}$, где $l(\varepsilon) = \min\{l, \text{pr}_l + \dots + \text{pr}_L \leq \varepsilon\}$. Если при этом в правой части (4.4) $\sum_{i=1}^N w_i^2 \rho^2(m^{(i)}, \bar{m}) \in S \setminus S_\varepsilon$, то коллективной экспертизе, представленной матрицей t_* , *доверять не следует*, ибо если эксперты и в самом деле принимают решения наугад и *взаимно независимо*, то для статистики (4.7) $\Pr(\{\omega \in \Omega, s(\omega) \in S \setminus S_\varepsilon\}) \geq 1 - \varepsilon$.

Суммируем полученные результаты.

Теорема 3.4.1.

1. Решением задачи (4.2) коллективного экспертного оценивания возможностей pr_j , $j = 1, \dots, n$, значений нечеткого элемента $\xi = x_j$, $j = 1, \dots, n$, является матрица t_* попарных сравнений значений $\text{pr}_1, \dots, \text{pr}_n, 0$, определенная равенствами (4.5).
2. В вероятностной модели экспертных решений $m^{(1)}, \dots, m^{(N)}$ как значений N *взаимно независимых случайных матриц* гипотеза о несогласованности частных экспертиз $m^{(1)}, \dots, m^{(N)}$ принимается с вероятностью, не меньшей $1 - \varepsilon$, если в (4.7) $s(\omega) \in S \setminus S_\varepsilon$, $0 \leq \varepsilon \leq 1$, где $\omega = (m^{(1)}, \dots, m^{(N)})$.

Чтобы определить значение $q(n)$ в (4.6), заметим, что любую конкретную упорядоченность в $1 = p_1 \geq \dots \geq p_n \geq 0$ можно задать, выбрав $s \in \{1, \dots, n\}$, указав множества $Q_1, \dots, Q_s$ так, чтобы

$$\begin{aligned} \overbrace{1 = p_1 = \dots = p_{k_1}}^{Q_1} &> \overbrace{p_{k_1+1} = \dots = p_{k_1+k_2}}^{Q_2} > \\ &> \dots > \overbrace{p_{k_1+\dots+k_{s-1}+1} = \dots = p_{k_1+\dots+k_s}}^{Q_s} > 0, \end{aligned} \quad (4.8)$$

и выбрав в (4.8) для Q_s условие > 0 или $= 0$ (при $s = 1$ лишь > 0). Поскольку в (4.8)

► для каждого $s = 1, \dots, n$ множества $Q_1, \dots, Q_s$ можно выбрать одним из $C_{n-1}^{s-1} \stackrel{\text{def}}{=} (n-1)!/((s-1)!(n-s)!)$ способов, распределив $s-1$ неравенств $>$ среди $n-1$ вакантных промежутков между $p_1, \dots, p_n$,

► при каждом выборе $Q_1, \dots, Q_s$ можно $n!/(k_1! \dots k_s!)$ способами перепорядочить значения p_i , $i = 1, \dots, n$,

► при $s > 1$ двумя способами выбрать условие > 0 или $= 0$ для наименьшей возможности среди $p_1, \dots, p_n$ (при $s = 1$ лишь > 0 , ибо $1 = p_1 = \dots = p_n > 0$), и, наконец,

► каждому из указанных способов упорядочения $p_1, \dots, p_n$, $p_{n+1} = 0$ отвечает одна матрица попарных сравнений,

то имеет место

Лемма 3.4.1. Число $q(n)$ всех $(n+1) \times (n+1)$ матриц в M_{n+1} определяется равенством

$$q(n) = 2 \sum_{s=2}^n \sum_{\{k_1, \dots, k_s\}} \frac{n!}{k_1! \dots k_s!} + 1, \quad (4.9)$$

где сумма $\sum_{\{k_1, \dots, k_s\}}$ содержит C_{n-1}^{s-1} слагаемых $n!/(k_1! \dots k_s!)$, $s = 2, \dots, n$, двойка учитывает две возможности выбора в (4.8): > 0 или $= 0$, последнее слагаемое отвечает $s = 1$ ($1 = p_1 = \dots = p_n > 0$, $k_s|_{s=1} = n$, $n!/n! = 1$).

Заметим, что если эксперты непременно должны различать возможности $p_1, \dots, p_n$, то в (4.9) $q(n) = 2n!$, а если к тому же обязательно считать, что $p_i > 0$, $i = 1, \dots, n$, то $q(n) = n!$. Этот случай рассмотрен иначе в следующем разделе.

3.4.2. Восстановление распределения нечеткого элемента путем упорядочения возможностей его значений. Рассмотрим задачу экспертного восстановления распределения нечеткого элемента ξ (4.1), в которой каждый эксперт представляет свое решение, указав перестановку $\pi(\cdot) : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$, упорядочивающую значения $p_1, \dots, p_n$ в (4.1) по убыванию.

Пусть $\pi_i(\cdot) : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ — перестановка, упорядочивающая по мнению i -го эксперта значения возможностей в (4.1)

$$1 = p_{i_1} > p_{i_2} > \dots > p_{i_n} > 0, \quad (4.10)$$

где $p_{i_k} \stackrel{\text{def}}{=} (\pi_i * p)_k = p_{\pi_i^{-1}(k)}$, $k = 1, \dots, n$, $i = 1, \dots, N$, и для простоты считается, что эксперты обязаны использовать только строгие неравенства.

В таком случае расстояние $r(\pi_i, \pi_s)$ между «мнениями» i -го и s -го экспертов можно определить, например, следующим выражением

$$r(\pi_i, \pi_s) = \left(\sum_{k=1}^n (\pi_i^{-1}(k) - \pi_s^{-1}(k))^2 \right)^{1/2} \equiv \left(\sum_{k=1}^n (i_k - s_k)^2 \right)^{1/2}. \quad (4.11)$$

Пусть $\pi_*^{-1}(1), \dots, \pi_*^{-1}(n)$ — номера, упорядочивающие (не обязательно строго) значения возможностей, которые указал бы эксперт, выражая мнения всех N экспертов. Тогда перестановка π_* может быть определена, например, как решение задачи

$$\sum_{i=1}^N w_i^2 r^2(\pi_i, \pi_*) = \min_{\pi} \sum_{i=1}^N w_i^2 r^2(\pi_i, \pi), \quad (4.12)$$

в которой минимум вычисляется на множестве всех перестановок $\pi(\cdot) : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$.

В задаче (4.12), как и в задаче (4.2), можно получить, в известном смысле, полное решение, а именно: найти коллективную экспертизу π_* и оценить, в какой степени ей следует доверять. Так как в (4.12), как и в (4.2),

$$\sum_{i=1}^N w_i^2 r^2(\pi_i, \pi) = \sum_{i=1}^N w_i^2 r^2(\pi_i, \bar{\pi}) + r^2(\bar{\pi}, \pi), \quad (4.13)$$

где $\bar{\pi}^{-1}(k) \stackrel{\text{def}}{=} \sum_{i=1}^N w_i^2 \pi_i^{-1}(k)$, $k = 1, \dots, n$, то задача (4.12) эквивалентна задаче отыскания перестановки π_* , ближайшей к функции $\bar{\pi} : r^2(\bar{\pi}, \pi_*) = \min_{\pi} r^2(\bar{\pi}, \pi)$ ($\bar{\pi}^{-1}(1), \dots, \bar{\pi}^{-1}(n)$), вообще говоря, не образуют перестановку $1, \dots, n$). Если для некоторой перестановки $\hat{\pi}$ $\bar{\pi}^{-1}(\hat{\pi}(1)) \leq \dots \leq \bar{\pi}^{-1}(\hat{\pi}(n))$, то, очевидно, $\pi_* = \hat{\pi}$, ибо $\sum_{k=1}^n (\bar{\pi}^{-1}(k) - \hat{\pi}^{-1}(k))^2 = \sum_{k=1}^n (\bar{\pi}^{-1}(\hat{\pi}(k)) - k)^2$ и для любых целых $s, k = 1, \dots, n$, $1 \leq s \leq s+k \leq n$, $(\bar{\pi}^{-1}(\hat{\pi}(s)) - s)^2 + (\bar{\pi}^{-1}(\hat{\pi}(s+k)) - s+k)^2 - (\bar{\pi}^{-1}(\hat{\pi}(s+k)) - s)^2 - (\bar{\pi}^{-1}(\hat{\pi}(s)) - s+k)^2 = 2k(\bar{\pi}^{-1}(\hat{\pi}(s)) - \bar{\pi}^{-1}(\hat{\pi}(s+k))) \leq 0$.

В вероятностной модели экспертных решений, аналогичной рассмотренной в § 3.4.1, значение первого слагаемого в правой части

(4.13) позволяет судить, насколько следует доверять "коллективной" экспертизе π_* .

Обозначим: $\omega \stackrel{\text{def}}{=} (\pi_1, \dots, \pi_N)$ — кортеж N произвольных перестановок, Π^N — класс всех $(n!)^N$ таких ω , $\Pr(\{\omega\}) = (n!)^{-N}$, $\omega \in \Pi^N$, — значения, определяющие вероятность $\Pr$ на $(\Pi^N, \mathcal{P}(\Pi^N))$. Вероятностное пространство $(\Pi^N, \mathcal{P}(\Pi^N), \Pr)$ моделирует заключения N экспертов, взаимно независимо принимающих решения наугад: все $(n!)^N$ возможных их экспертиз равновероятны.

Статистика $t(\omega) \stackrel{\text{def}}{=} \sum_{i=1}^N w_i^2 r^2(\pi_i, \bar{\pi})$, $\omega \in \Pi^N$, принимает значения в $T = \{t(\omega), \omega \in \Pi^N\} \stackrel{\text{def}}{=} \{t_1, \dots, t_K\}$ с вероятностями $\text{pr}_k \stackrel{\text{def}}{=} \Pr(\{\omega \in \Pi^N, t(\omega) = t_k\})$, $k = 1, \dots, K$, которые определим упорядоченными по убыванию: $\text{pr}_1 \geq \dots \geq \text{pr}_K$. Если H — гипотеза, *согласно которой эксперты* принимают решения наугад и взаимно независимо, то критерий, отвергающий H ошибочно с вероятностью, не большей ε , $\varepsilon > 0$, определится критическим множеством $T_\varepsilon \stackrel{\text{def}}{=} \{t_{k(\varepsilon)}, t_{k(\varepsilon)+1}, \dots, t_K\}$, где $k(\varepsilon) = \min\{k, t_k \in T, \text{pr}_k + \dots + \text{pr}_K \leq \varepsilon\}$.

Включение $\sum_{i=1}^N w_i^2 r^2(\pi_i, \bar{\pi}) \in T \setminus T_\varepsilon$ означает, что «коллективной» экспертизе π_* доверять не следует в силу «патологической» несогласованности частных экспертиз.

3.5. Нечеткие оптимальные решения [13]

3.5.0. Введение. Последние четыре десятилетия внимание исследователей, работающих в различных предметных областях, привлекают проблемы оптимизации решений в нечеткой обстановке, при нечетких целях и предпочтениях, при нечетких ограничениях и множествах альтернатив. Исследованы нечеткие варианты теории игр, математического программирования, теории полезности и т. д., см., например, [5–9, 12, 17, 25, 31–35, 37, 39].

В этом параграфе представлены методы оптимальных нечетких (возможностных) решений на примере задач оптимальной идентификации, или, иначе говоря, — оптимального выбора одной из альтернативных гипотез о принадлежности объекта к одному из конечного числа классов, и методы оптимального оценивания нечетких элементов. Качество решения определено в терминах риска сопутствующих решению потерь и охарактеризовано значениями возможности и (или) необходимости потерь. Изложение следует схеме, принятой в теории статистических решений [3, 20].

Напомним основные понятия теории статистических решений, рассмотренной в § 8 гл. 2, на примере задачи идентификации состояний стохастической системы. Система может находиться в одном из q состояний, значения $x \in X$ случайного элемента ξ определяют результаты

наблюдений за системой, причем если система находится в состоянии "k", то значения ξ контролируются (известным) распределением вероятностей $\text{pr}_k(x)$, $x \in X$, $k = 1, \dots, q$. В задаче идентификации обычно задана величина $l_{k,d}$ потерь, сопутствующих решению в пользу состояния "d", в то время как система находится в состоянии "k", $d, k = 1, \dots, q$, и требуется по наблюдению значения $\xi = x$ принять одно из q решений о состоянии системы.

Предположим, что состояния системы априори случайны и моделируются значениями случайной величины $\varkappa$, распределение $\text{pr}^\varkappa(k) = \text{Pr}(\{\varkappa = k\})$, $k = 1, \dots, q$ которой известно. Поскольку $\text{pr}_k(x) = \text{pr}^{\xi|\varkappa}(x|k)$, $x \in X$, — распределение переходной вероятности, контролирующей результаты наблюдений за системой, находящейся в состоянии "k", $k = 1, \dots, q$, то $\text{pr}^{\xi, \varkappa}(x, k) = \text{pr}^{\xi|\varkappa}(x|k)\text{pr}^\varkappa(k)$, $x \in X$, $k = 1, \dots, q$, — совместное распределение вероятностей "наблюдений, состояний", и, следовательно, речь идет о байесовской задаче идентификации, подробнее см. в [13].

Пусть решение $\delta = d$ о состоянии системы при наблюдении $\xi = x$ принимается случайно и $\text{pr}^{\delta|\xi}(d|x)$ — переходная вероятность решения $\delta = d$, $d = 1, \dots, q$, определяющая рандомизированное правило принятия решения о состоянии системы при наблюдении $\xi = x$, $x \in X$. Тогда $\text{pr}^{\varkappa, \delta, \xi}(k, d, x) = \text{pr}^{\delta|\xi}(d|x)\text{pr}^{\xi, \varkappa}(x, k)$, $k, d = 1, \dots, q$, $x \in X$, — совместное распределение "состояний, решений, наблюдений", $\text{pr}^{\varkappa, \delta}(k, d) = \int_X \text{pr}^{\varkappa, \delta, \xi}(k, d, x)\mu(dx)$, $k, d = 1, \dots, q$, — распределение "состояний,

решений" и $L(\text{pr}^{\delta|\xi}) = \sum_{k,d=1}^q l_{k,d}\text{pr}^{\varkappa, \delta}(k, d)$ — ожидаемые потери, обусловленные правилом $\text{pr}^{\delta|\xi}$ идентификации, оптимальное правило идентификации определяется из условия

$$L(\text{pr}^{\delta|\xi}) \sim \min_{\text{pr}^{\delta|\xi}}. \quad (5.1)$$

Прежде чем приступить к изучению нечетких решений, рассмотрим еще одну точку зрения на критерий качества статистического решения. Согласно этой точке зрения в задаче идентификации состояния стохастической системы значение $l_{k,d}$ следует понимать как определенную субъектом, принимающим решения, *вероятность* потерь, неудач и т.п. неприятностей, обусловленных режимом эксплуатации системы, свойственным ее состоянию «d», в то время как на самом деле система находится в состоянии «k», $k, d = 1, \dots, q$. Формально эту точку зрения удобно представить следующим образом: $l_{k,d}$ — *вероятность покрытия точки (k, d) случайным множеством Λ , определенным на некотором вероятностном пространстве и принимающим значения в $\mathcal{P}(\{1, \dots, q\} \times \{1, \dots, q\})$* , т.е. ¹⁾ $l_{k,d} = \text{Pr}((k, d) \in \Lambda)$, $(k, d) \in \{1, \dots, q\} \times \{1, \dots, q\}$.

¹⁾ Pr и Λ определяются условиями эксплуатации системы.

Если $\text{pr}^{\xi, \varkappa}(x, k)$, $x \in X$, $k \in \{1, \dots, q\}$, — совместное распределение вероятностей наблюдений за системой и ее состояний, $\text{pr}^{\delta|\xi}(d|x)$, $d \in \{1, \dots, q\}$, $x \in X$, — переходная вероятность, определяющая рандомизированное правило идентификации, согласно которому $\text{pr}^{\delta|\xi}(d|x)$ — вероятность решения $\delta = d$ о состоянии системы, принятого на основании наблюдения $\xi = x$, то

$$\text{pr}^{\delta, \varkappa}(d, k) = \int_X \text{pr}^{\delta|\xi}(d|x) \text{pr}^{\xi, \varkappa}(x, k) \mu(dx), \quad d \in \{1, \dots, q\}, \quad k \in \{1, \dots, q\},$$

— совместное распределение вероятностей решений о состоянии системы, принимаемых согласно правилу $\text{pr}^{\delta|\xi}$ на основании наблюдения за системой, и ее действительных состояний. А поскольку случайное множество Λ и пару случайных величин $(\delta, \varkappa)$ естественно считать независимыми, то

$$\text{Pr L}(\text{pr}^{\delta|\xi}) \stackrel{\text{def}}{=} \sum_{k,d=1}^q l_{k,d} \text{pr}^{\delta, \varkappa}(d, k) = \sum_{k,d=1}^q l_{k,d} \int_X \text{pr}^{\delta|\xi}(d|x) \text{pr}^{\xi, \varkappa}(x, k) \mu(dx)$$

— *вероятность потерь*, сопутствующих правилу идентификации $\text{pr}^{\delta|\xi}$, которое естественно определить из условия

$$\text{Pr L}(\text{pr}^{\delta|\xi}) \sim \min_{\text{pr}^{\delta|\xi}}. \quad (5.2)$$

Подобный критерий будет использован при построении оптимальных нечетких решений.

Заметим, что критерии оптимальности статистических решений (5.1), (5.2) определяют качество решений лишь асимптотически при условии, что решения принимаются многократно и взаимно независимо.

3.5.1. Нечеткая модель. Идентификация. В этом параграфе речь идет о *нечеткой системе*, которая может находиться в одном из q состояний, полностью или частично определяющих распределение возможностей значений нечеткого элемента ξ , регистрируемых при наблюдении за системой и составляющих основу для принятия решения о ее состоянии.

3.5.1.1 Модель наблюдения и правило идентификации.

Обозначим: K множество состояний нечеткой системы, D — множество решений о ее состоянии, Λ — нечеткое множество, определенное на пространстве с возможностью $(Y, \mathcal{P}(Y), P_Y)$, со значениями в классе $\mathcal{P}(K \times D)$ всех подмножеств $K \times D$, $l_{\cdot, \cdot} \stackrel{\text{def}}{=} l_{\cdot, \cdot}^{\Lambda} : K \times D \rightarrow [0, 1]$ — индикаторную функцию Λ , см. § 3.1.3.1. Значение $l_{k,d} \in [0, 1]$ есть возможность покрытия точки $(k, d) \in K \times D$ нечетким множеством Λ , которую в связи с проблемой идентификации будем интерпретировать

как *возможность потерь*¹⁾, обусловленных использованием нечеткой системы согласно решению « d » о ее состоянии, в то время как система находится в состоянии « k », $d \in D$, $k \in K$; $\sup_{k,d \in K \times D} l_{k,d} = 1$.

В рассматриваемой далее модели, аналогичной байесовской модели в §8 гл. 2, состояние нечеткой системы $\varkappa$ и наблюдение ξ за ней определены как пара нечетких элементов, и задано их распределение $f^{\xi, \varkappa}(\cdot, \cdot) : X \times K \rightarrow [0, 1]$, значение $f^{\xi, \varkappa}(x, k)$ которого есть возможность того, что система находится в состоянии $\varkappa = k \in K$ и $\xi = x \in X$ — результат наблюдения за системой.

Схему наблюдения за системой определим парой нечетких элементов $(\xi, \varkappa)$ («наблюдение», «состояние»), первый из которых наблюдаем, второй — нет. *Распределение $f^{\xi, \varkappa}(\cdot, \cdot)$, определяющее нечеткую связь между состояниями системы и результатами наблюдений за ней, назовем нечеткой моделью системы и схемы наблюдения за системой.*

Тройку $(f^{\xi, \varkappa}, P_Y, \Lambda)$ назовем *моделью идентификации*, в которой $f^{\xi, \varkappa}$ характеризует свойства системы, а P_Y, Λ определены лицом, принимающим решения о состоянии системы.

Решение о состоянии системы определим как нечеткий элемент δ , принимающий значения в D , $\pi^{\delta|\xi}(\cdot|\cdot) : D \times X \rightarrow [0, 1]$ обозначим функцию, которая при каждом $x \in X$ является распределением переходной возможности $\Pi(\cdot|\cdot) : \mathcal{P}(D) \times X \rightarrow [0, 1]$, определенной формулой:

$$\Pi(A|x) = \sup_{d \in A} \pi^{\delta|\xi}(d|x), \quad A \in \mathcal{P}(D).$$

Распределение $\pi^{\delta|\xi}$ назовем *нечетким правилом решения δ о состоянии системы, основанного на наблюдении ξ , или нечетким правилом идентификации*; значение $\pi^{\delta|\xi}(d|x)$ — возможность решения $\delta = d$, когда $\xi = x$ — результат наблюдения за системой, $d \in D$, $x \in X$.

Пара нечетких элементов $\delta, \varkappa$ и нечеткое множество Λ далее считаются независимыми в смысле одноточечного покрытия.

Напомним, что переходной возможностью на $(X, \mathcal{P}(X))$ и $(D, \mathcal{P}(D))$ называется любое отображение $\Pi(\cdot|\cdot) : \mathcal{P}(D) \times X \rightarrow [0, 1]$, такое, что при каждом $x \in X$ $\Pi(\cdot|x) : \mathcal{P}(D) \rightarrow [0, 1]$ есть возможность на $(D, \mathcal{P}(D))$.

Следующее свойство переходной возможности непосредственно вытекает из ее определения.

Лемма 3.5.1. Пусть $P_X(\cdot) : \mathcal{P}(X) \rightarrow [0, 1]$ — любая возможность на $(X, \mathcal{P}(X))$, $f^{\xi}(\cdot) : X \rightarrow [0, 1]$ — ее распределение и $\pi^{\delta|\xi}(\cdot|x) : D \rightarrow [0, 1]$ — распределение переходной

¹⁾ Значение $l_{k,d}$ можно интерпретировать как значение переходной возможности $P(L|k, d)$ события $L =$ "потери, неприятности и т.п." при $\varkappa = k$, $\delta = d$, где δ — нечеткий элемент со значениями в D , определяющий нечеткое правило решения о состоянии системы, подробнее см. [13].

возможности $\Pi(\cdot|x) : \mathcal{P}(D) \rightarrow [0, 1]$, $x \in X$. Тогда $f^{\xi, \delta}(x, d) = \min(\pi^{\delta|\xi}(d|x), f^{\xi}(x))$, $x \in X$, $d \in D$, — распределение возможности $P(\cdot) : \mathcal{P}(X \times D) \rightarrow [0, 1]$, $P(C) = \sup_{(x,d) \in C} f^{\xi, \delta}(x, d)$, $C \in \mathcal{P}(X \times D)$.

Доказательство. См. доказательство Леммы 3.1.6. ■

Замечание 3.5.1. Модель нечеткой системы и наблюдения за ней естественно определить априорным распределением возможностей $f^{\varkappa}(\cdot) : K \rightarrow [0, 1]$ ее состояний и распределением переходной возможности $f^{\xi|\varkappa}(\cdot|k) : X \rightarrow [0, 1]$ для каждого состояния $k \in K$. В таком случае

$$f^{\xi, \varkappa}(x, k) = \min(f^{\xi|\varkappa}(x|k), f^{\varkappa}(k)), \quad x \in X, k \in K, \quad (5.3)$$

— совместное распределение $\xi, \varkappa$. Далее, как правило, будем использовать именно такую конструкцию модели, накладывающую определенные ограничения на распределение $f^{\xi, \varkappa}(\cdot, \cdot)$, см. Леммы 3.1.6 и 3.5.1, при которых распределение $f^{\xi|\varkappa}(\cdot|k)$ переходной возможности можно интерпретировать как распределение условной возможности.

3.5.2. Критерий минимума возможности (риска) потерь. Для рассмотренной в предыдущем параграфе модели согласно Лемме 3.5.1

$$f^{\delta, \xi, \varkappa}(d, x, k) \stackrel{\text{def}}{=} \min(\pi^{\delta|\xi}(d|x), f^{\xi, \varkappa}(x, k)) \quad (5.4)$$

— возможность того, что система находится в состоянии $\varkappa = k \in K$, $\xi = x \in X$ — результат наблюдения за системой и принято решение $\delta = d \in D$ о ее состоянии; $f^{\delta, \xi, \varkappa}(d, x, k)$ — возможность трех равенств $\delta = d$, $\xi = x$ и $\varkappa = k$. Соответственно

$$f^{\varkappa, \delta}(k, d) = \sup_{x \in X} f^{\delta, \xi, \varkappa}(d, x, k) \quad (5.5)$$

— возможность того, что система находится в состоянии $\varkappa = k$ и при нечетком правиле решения $\pi^{\delta|\xi}$ в результате наблюдения за системой принято решение $\delta = d$ о ее состоянии, а так как нечеткое множество Λ и нечеткий элемент $(\delta, \varkappa)$ независимы, то

$$\min(l_{k,d}, f^{\varkappa, \delta}(k, d)) \quad (5.6)$$

— возможность такой ситуации и свойственных ей потерь. Поэтому *возможность потерь (риск потерь), определяющая качество правила решения $\pi^{\delta|\xi}$,*

$$\text{PL}(\pi^{\delta|\xi}) = \sup_{x \in X, k \in K, d \in D} \min(l_{k,d}, \pi^{\delta|\xi}(d|x), f^{\xi, \varkappa}(x, k)). \quad (5.7)$$

Оптимальным естественно считать такое правило $\pi^{* \delta|\xi}$, которое минимизирует возможность потерь (5.7):

$$\text{PL}(\pi^{* \delta|\xi}) = \min_{\pi^{\delta|\xi}} \text{PL}(\pi^{\delta|\xi}). \quad (5.8)$$

Значение $PL^\Lambda(\pi^{*\delta|\xi})$ определяет минимальный риск потерь, а соответствующее нечеткое правило $\pi^{*\delta|\xi}$ при наблюдении $\xi = x$ определяет возможность $\pi^{*\delta|\xi}(k|x)$ принятия решения в пользу k -го состояния системы для каждого $k \in D = \{1, \dots, q\}$.

Заметим, что для принятия решения по правилу $\pi^{*\delta|\xi}$ можно использовать рандомизацию. Если

$$\begin{array}{ccccccc} 1 & = & \pi^{*\delta|\xi}(k_1|x) & \geq & \pi^{*\delta|\xi}(k_2|x) & \geq & \dots & \geq & \pi^{*\delta|\xi}(k_q|x) & \geq & 0, \\ 0. & & e_1 & & e_2 & & e_{q-1} & & e_q & & \end{array} \quad (*)$$

где двоичное число $e = e(x) = 0.e_1 \dots e_q$ определяет упорядоченность в (*) так, что $e_i = 1$ означает ">", а $e_i = 0$ означает "=", $i = 1, \dots, q$, и

$$1 \geq \text{pr}_{k_1}^{(x)} \geq \dots \geq \text{pr}_{k_q}^{(x)} \geq 0 \quad (**)$$

распределение некоторой вероятности $\text{Pr}^{(x)} \in \mathbb{P}r_{(e(x))}$, см. § 3.2, то $\exists \tilde{\gamma}_{(e(x))}(\cdot) \in \tilde{\Gamma}(\text{Pr}^{(x)}) \forall i = 1, \dots, q \pi^{*\delta|\xi}(k_i|x) = \tilde{\gamma}_{(e(x))}(\text{pr}_{k_i}^{(x)})$. Поэтому для принятия решения по правилу $\pi^{*\delta|\xi}$ при $\xi = x$ можно разыграть случайную величину, распределенную согласно (**), и если выпадет k_i , то принять решение $\delta = k_i$.

Характерно, что в то время как согласно правилу $\pi^{*\delta|\xi}$ оптимальное решение "принимается однократно", рандомизация обеспечивает оптимальность случайных решений лишь в среднем, когда они принимаются многократно и взаимно независимо. Если $\nu_k^{(n,x)}$ — частота выпадения $\xi = k$, где n — число принятий решений, и $\tilde{\gamma}_{(e(x))}(\nu_k^{(n,x)}) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} \tilde{\gamma}_{(e(x))}(\text{pr}_k^{(x)}) = \pi^{*\delta|\xi}(k|x)$ равномерно по $x \in X$, $k = 1, \dots, q$, то $PL(\tilde{\gamma}_{(e(\cdot))}(\nu^{(n,\cdot)})) = \sup_{x \in X, k \in K, d \in D} \min(l_{k,d}, \tilde{\gamma}_{(e(x))}(\nu_k^{(n,x)}), f^{\xi, \kappa}(x, k)) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} PL(\pi^{*\delta|\xi})$.

На самом деле, как выяснится в дальнейшем, в любом случае, в том числе и тогда, когда неизвестно априорное распределение $f^{\kappa}(k)$, $k = 1, \dots, q$, состояний системы, см. замечание 3.5.1, правило $\pi^{*\delta|\xi}$ можно выбрать "четким", согласно которому $\pi^{*\delta|\xi}(k|x) = \begin{cases} 1, & k = k(x), \\ 0, & k \neq k(x), \end{cases}$ $k = 1, \dots, q$, $x \in X$, ср. с результатами, полученными в § 2.8.

В этом случае критерий (5.8) определяет качество каждого решения.

3.5.3. Правило решения, минимизирующее риск потерь. Пусть $K = \{1, \dots, q\} = D$ и

$$P_d(x) \stackrel{\text{def}}{=} \max_{1 \leq k \leq q} \min(l_{k,d}, f^{\xi, \kappa}(x, k)), \quad (5.9)$$

— возможность потерь, сопутствующих решению $\delta = d$ и наблюдению $\xi = x$, $x \in X$, $d \in \{1, \dots, q\}$.

Тогда согласно (5.7), (5.8), (5.9) задача определения оптимального нечеткого правила $\pi^{\delta|\xi}$ формулируется как задача на минимум

$$PL(\pi^{\delta|\xi}) = \sup_{x \in X} \max_{1 \leq d \leq q} \min(\pi^{\delta|\xi}(d|x), P_d(x)) \sim \min_{\pi^{\delta|\xi}(\cdot|\cdot)}. \quad (5.10)$$

Ее решение можно получить, решив для каждого $x \in X$ более простую задачу

$$\max_{1 \leq d \leq q} \min(\pi^{\delta|\xi}(d|x), P_d(x)) \sim \min_{\pi^{\delta|\xi}(\cdot|x)}. \quad (5.11)$$

Это показано ниже в Теореме 3.5.1, которая является следствием более общего утверждения, доказанного в следующей Лемме.

Лемма 3.5.2. Пусть $g_j(\cdot) : X \rightarrow [0, 1]$, $f_j(\cdot) : X \rightarrow [0, 1]$, $j = 1, \dots, q$, — некоторые функции. Для каждого $x \in X$:

► упорядочим значения $g_j(x)$, $j = 1, \dots, q$, так, чтобы ¹⁾

$$g(x) \stackrel{\text{def}}{=} \min_{1 \leq j \leq q} g_j(x) = g_{i_1}(x) = \dots = g_{i_t}(x) < g_{i_{t+1}}(x) \leq \dots \leq g_{i_q}(x), \quad (5.12)$$

► выберем $f_j^*(\cdot) : X \rightarrow [0, 1]$, $j = 1, \dots, q$, так, чтобы

$$\max_{1 \leq s \leq t} f_{i_s}^*(x) = 1, \quad \max_{t < s \leq q} f_{i_s}^*(x) = 0, \quad x \in X. \quad (5.13)$$

Тогда для любых функций $f_j(\cdot) : X \rightarrow [0, 1]$, $j = 1, \dots, q$, удовлетворяющих условию $\max_{1 \leq j \leq q} f_j(x) = 1$, $x \in X$,

$$\begin{aligned} \max_{1 \leq j \leq q} \min(g_j(x), f_j^*(x)) &= \max_{1 \leq s \leq t} \min(g_{i_s}(x), f_{i_s}^*(x)) = g(x) \leq \\ &\leq \max_{1 \leq j \leq q} \min(g_j(x), f_j(x)), \quad x \in X, \end{aligned} \quad (5.14)$$

и соответственно

$$\sup_{x \in X} \max_{1 \leq j \leq q} \min(g_j(x), f_j^*(x)) = \sup_{x \in X} g(x) \leq \sup_{x \in X} \max_{1 \leq j \leq q} \min(g_j(x), f_j(x)). \quad (5.15)$$

Доказательство. Равенства в (5.14) выполнены в силу условия (5.12) и того, что согласно (5.13) $\max_{t < s \leq q} \min(g_{i_s}(x), f_{i_s}^*(x)) = 0$, $x \in X$.

Так как

$$\begin{aligned} \max_{1 \leq j \leq q} \min(g_j(x), f_j(x)) &= \max \left(\max_{1 \leq s \leq t} \min(g_{i_s}(x), f_{i_s}(x)), \right. \\ &\quad \left. \max_{t < s \leq q} \min(g_{i_s}(x), f_{i_s}(x)) \right), \quad \text{то} \end{aligned}$$

¹⁾ "Упорядочивающая" функция $i : \{1, \dots, q\} \rightarrow \{1, \dots, q\}$, и, в частности, ее значения $i_1, \dots, i_q$, равно как и t , зависят от $x \in X$.

— если $\max_{t < s \leq q} f_{i_s}(x) = 1$, то в силу условий (5.12), (5.13)
 $\max_{1 \leq j \leq q} \min(g_j(x), f_j(x)) = \max_{t < s \leq q} \min(g_{i_s}(x), f_{i_s}(x)) \geq g_{i_{t+1}}(x) >$
 $> \max_{1 \leq s \leq t} \min(g_{i_s}(x), f_{i_s}^*(x)) = g(x);$
 — если же $\max_{t < s \leq q} f_{i_s}(x) < 1$, то $\max_{1 \leq j \leq q} \min(g_j(x), f_j(x)) \geq$
 $\geq \max_{1 \leq s \leq t} \min(g_{i_s}(x), f_{i_s}(x)) = \max_{1 \leq s \leq t} \min(g_{i_s}(x), f_{i_s}^*(x)) = g(x)$. Неравен-

ство в (5.15) следует из соотношений (5.14). ■

Применительно к задачам (5.10) и (5.11) как следствие Леммы 3.5.2 получаем следующий результат.

Теорема 3.5.1. Пусть $K = D = \{1, \dots, q\}$. Для каждого $x \in X$ упорядочим значения $P_d(x)$, $d = 1, \dots, q$, в (5.9) согласно условию

$$P(x) \stackrel{\text{def}}{=} \min_{1 \leq d \leq q} P_d(x) = P_{d_1}(x) = \dots = P_{d_t}(x) < P_{d_{t+1}}(x) \leq \dots \leq P_{d_q}(x), \quad (5.16)$$

где функция $d : \{1, \dots, q\} \rightarrow \{1, \dots, q\}$ и t зависят от x . Тогда всякое нечеткое правило $\pi^{* \delta|\xi}$, удовлетворяющее условиям

$$\max_{1 \leq s \leq t} \pi^{* \delta|\xi}(d_s|x) = 1, \quad \max_{t < s \leq q} \pi^{* \delta|\xi}(d_s|x) = 0, \quad x \in X, \quad (5.17)$$

— решение задач (5.10) и (5.11), т.е. оптимально. При любом правиле $\pi^{* \delta|\xi}$, удовлетворяющем условиям (5.16), (5.17),

$$PL(\pi^{* \delta|\xi}) = \sup_{x \in X} P(x) \equiv \sup_{x \in X} \min_{1 \leq d \leq q} P_d(x) \leq PL(\pi^{\delta|\xi}), \quad (5.18)$$

где $\pi^{\delta|\xi}$ — произвольное нечеткое правило решения. ■

Заметим, что так как в (5.9) $P_d(x)$ — возможность потерь, сопутствующих решению $\delta = d$ и наблюдению $\xi = x$, то согласно условиям (5.16), (5.17) оптимальное правило $\pi^{* \delta|\xi}$ предписывает при наблюдении $\xi = x$ принимать решение $\delta = d$, возможность которого тем больше, чем меньше возможность потерь, обусловленных таким решением.

Замечание 3.5.2. Достаточные условия оптимальности правила $\pi^{* \delta|\xi}$, эквивалентные условиям (5.16), (5.17), можно сформулировать в следующей, в ряде случаев более удобной, форме. Пусть

$$D^*(x) \stackrel{\text{def}}{=} \left\{ d \in \{1, \dots, q\}, P_d(x) = P(x) \stackrel{\text{def}}{=} \min_{d'} P_{d'}(x) \right\}, \quad x \in X. \quad (5.19)$$

Тогда любое нечеткое правило $\pi^{* \delta|\xi}$, удовлетворяющее условиям

$$\max_{d \in D^*(x)} \pi^{* \delta|\xi}(d|x) = 1, \quad \pi^{* \delta|\xi}(d|x) = 0, \quad d \in \{1, \dots, q\} \setminus D^*(x), \quad x \in X, \quad (5.20)$$

является решением задач (5.10) и (5.11). Поскольку согласно (5.20)

можно выбрать $\pi^{* \delta|\xi}(d|x) = \begin{cases} 1, & \text{если } d \in D^*(x), \\ 0, & \text{если } d \in \{1, \dots, q\} \setminus D^*(x), \end{cases} \quad x \in X,$
 то при наблюдении $\xi = x$ в качестве решения можно использовать

значение $d^*(x)$ любой «четкой» функции $d^*(\cdot) : X \rightarrow D$, удовлетворяющей условию $d^*(x) \in D^*(x)$, $x \in X$.

Замечание 3.5.3. Пусть $K = D = \{1, \dots, q\}$ и

$$l_{k,d} = \begin{cases} 0, & \text{если } d = k \\ 1, & \text{если } d \neq k \end{cases}, \quad k, d = 1, \dots, q, \quad (5.21)$$

т. е. пусть потери невозможны при правильной идентификации и возможность потерь максимальна при ошибочной идентификации. В таком случае правая часть равенства (5.7) определяет *возможность неверной идентификации* $PE(\pi^{\delta|\xi})$, а в (5.9)

$$P_d(x) = \max_{k \neq d} f^{\xi, \varkappa}(x, k), \quad d = 1, \dots, q, \quad x \in X, \quad (5.22)$$

и, следовательно, в (5.19) минимум $P_d(x)$ при каждом $x \in X$ достигается на тех $d \in \{1, \dots, q\}$, при которых $f^{\xi, \varkappa}(x, d)$ достигает максимума, ибо если $f^{\xi, \varkappa}(x, d_*(x)) \geq f^{\xi, \varkappa}(x, k)$, $k = 1, \dots, q$, то $P_{d_*(x)}(x) \leq P_k(x)$, $k = 1, \dots, q$, $x \in X$.

В данном случае в (5.19) для каждого $x \in X$,

$$D^*(x) \supset D_*(x) \stackrel{\text{def}}{=} \left\{ d \in \{1, \dots, q\}, f^{\xi, \varkappa}(x, d) = f^{\xi}(x) \stackrel{\text{def}}{=} \max_{1 \leq k' \leq q} f^{\xi, \varkappa}(x, k') \right\}, \quad (5.23)$$

где $f^{\xi}(x)$ — возможность наблюдения $\xi = x$, $x \in X$. Условимся считать, что в случае (5.21) при каждом $x \in X$ максимум $\pi^{* \delta|\xi}(\cdot|x) : \{1, \dots, q\} \rightarrow [0, 1]$, равный единице, достигается на множестве $D_*(x)$ тех $d \in D = \{1, \dots, q\}$, на которых достигается максимум $f^{\xi, \varkappa}(x, \cdot) : \{1, \dots, q\} \rightarrow [0, f^{\xi}(x)]$, равный $f^{\xi}(x)$; вне этого множества определим $\pi^{* \delta|\xi}(d|x) = 0$, $d \in \{1, \dots, q\} \setminus D_*(x)$. Каждое такое правило $\pi^{* \delta|\xi}$, минимизирующее *возможность ошибки идентификации* $PE(\pi^{\delta|\xi})$, называется правилом *максимальной возможности* [13]. При этом согласно (5.18) $PE(\pi^{* \delta|\xi}) = \sup_{x \in X} \max_{k \neq d_*(x)} f^{\xi, \varkappa}(x, k)$, где $d_*(x) \in D_*(x)$, $x \in X$.

Иначе определение (5.23) можно записать как

$$D_*(x) = \{d \in \{1, \dots, q\}, f^{\varkappa|\xi}(d|x) = 1\}, \quad x \in X, \quad (5.24)$$

если условное распределение $\varkappa$ при условии $\xi = x$ определить равенствами

$$f^{\varkappa|\xi}(k|x) = \begin{cases} f^{\xi, \varkappa}(x, k), & \text{если } f^{\xi, \varkappa}(x, k) < f^{\xi}(x) \\ 1, & \text{если } f^{\xi, \varkappa}(x, k) = f^{\xi}(x) \end{cases}, \quad (5.25)$$

$$k = 1, \dots, q, \quad x \in X, \quad [13].$$

Поэтому всякое правило $\pi^{* \delta|\xi}$, такое, что

$$\max_{d: f^{\varkappa|\xi}(d|x)=1} \pi^{* \delta|\xi}(d|x) = 1, \quad \max_{d: f^{\varkappa|\xi}(d|x)<1} \pi^{* \delta|\xi}(d|x) = 0, \quad x \in X, \quad (5.26)$$

минимизирует *возможность (риск) ошибочного решения*.

Замечание 3.5.4. Если наблюдения за системой невозможны, то в модели системы (5.3) задано лишь априорное распределение возможностей ее состояний $f^{\mathcal{K}}(k) = \sup_{x \in X} f^{\xi, \mathcal{K}}(x, k)$, $k \in K$.

Теорема 3.5.1 характеризует оптимальное решение и в этом случае, если в ее условиях $f^{\xi, \mathcal{K}}(x, k)$ заменить на $f^{\mathcal{K}}(k)$, $\pi^{\delta|\xi}(d|x)$ — на $\pi^{\delta}(d)$, $P_d(x)$ — на $P_d \stackrel{\text{def}}{=} \max_{1 \leq k \leq q} \min(l_{kd}, f^{\mathcal{K}}(k))$,

$P(x)$ — на $P \stackrel{\text{def}}{=} \min_{d \in D} P_d$, $x \in X$, $k \in K$, $d \in D$. В этом случае речь идет о нечетком правиле идентификации $\pi^{*\delta}$, определенном условиями $\max_{1 \leq s \leq t} \pi^{*\delta}(d_s) = 1$, $\max_{t < s \leq q} \pi^{*\delta}(d_s) = 0$, $0 \leq P = P_{d_1} = \dots = P_{d_t} < P_{d_{t+1}} \leq \dots \leq P_{d_q}$, для которого $P_d = \max_{1 \leq k \leq q} \min(l_{k,d}, f^{\mathcal{K}}(k)) \geq \max_{1 \leq k \leq q} \min(l_{k,d}, f^{\xi, \mathcal{K}}(x, k)) = P_d(x)$, $d \in D$, $x \in X$, и, следовательно $P = \min_{1 \leq d \leq q} P_d \geq \sup_{x \in X} \min_{1 \leq d \leq q} P_d(x) = \sup_{x \in X} P(x)$. *Возможность потерь, сопутствующих решению $\pi^{*\delta}(\cdot)$, не учитывающему результат наблюдения за системой, вообще говоря, больше, чем при решении $\pi^{*\delta|\xi}(\cdot|\cdot)$.*

Если, наконец, недоступны не только наблюдения за системой, но и априорное распределение возможностей ее состояний, то, как будет показано в § 3.5.5, все ее состояния априори следует считать равновероятными, $f^{\mathcal{K}}(k) = 1$, $k = 1, \dots, q$, и, следовательно, $P_d = \max_{1 \leq k \leq q} \min(l_{k,d}, f^{\mathcal{K}}(k)) \leq \max_{1 \leq k \leq q} l_{k,d}$, $d = 1, \dots, q$. В этом

случае качество решения $\pi^{*\delta}(\cdot)$, определяемого из условий $\max_{1 \leq s \leq t} \pi^{*\delta}(d_s) = 1$, $\max_{t < s \leq q} \pi^{*\delta}(d_s) = 0$, $0 \leq P = \max_{1 \leq k \leq q} l_{k,d_1} = \dots = \max_{1 \leq k \leq q} l_{k,d_t} < \max_{1 \leq k \leq q} l_{k,d_{t+1}} \leq \dots \leq \max_{1 \leq k \leq q} l_{k,d_q}$, еще ниже.

3.5.4. Четкие правила идентификации. Среди нечетких правил $\pi^{*\delta|\xi}$, определенных условиями (5.16), (5.17), как было отмечено в замечании 3.5.2, всегда есть четкие, согласно которым для каждого $x \in X$ $\pi^{*\delta|\xi}(d_s|x) = 1$ для некоторого $s \in [1, t(x)]$ и $\pi^{*\delta|\xi}(d|x) = 0$ для остальных $d \neq d_s$, $d \in \{1, \dots, q\}$. Поскольку при этом неравенство в (5.18) выполнено, в рассматриваемой модели «фазификация» правила решения не позволяет улучшить качество идентификации, и имеет смысл рассмотреть построение непосредственно четких правил идентификации.

Любое четкое правило определяется упорядоченным разбиением $X = \bigcup_{d=1}^q X_d$, а именно, решение $\delta = d$ о состоянии системы принимается тогда, когда результат наблюдения $\xi = x \in X_d$, $d = 1, \dots, q$.

Пусть $\chi^{\delta|\xi}(d|x) = \begin{cases} 1, & \text{если } x \in X_d \\ 0, & \text{если } x \in X \setminus X_d \end{cases}$, $x \in X$, — индикаторная функция множества X_d , $d = 1, \dots, q$. Тогда возможность потерь, свойственных правилу, определенному разбиением X_* ,

$$PL(X_*) = \sup_{x \in X} \max_{1 \leq d \leq q} \min(\chi^{\delta|\xi}(d|x), P_d(x)) = \max_{1 \leq d \leq q} \sup_{x \in X_d} P_d(x). \quad (5.27)$$

Следующая теорема, определяющая «четкие» правила идентификации, является частным случаем Теоремы 3.5.1.

Теорема 3.5.1*. *Решением задачи на минимум*

$$PL(X_*) \sim \min_{X_*}, \quad (5.28)$$

для возможности потерь (5.27) является любое разбиение X_* , удовлетворяющее условиям

$$X_d^* \subset \left\{ x \in X, P_d(x) = P(x) \stackrel{\text{def}}{=} \min_{1 \leq d' \leq q} P_{d'}(x) \right\}, \quad d = 1, \dots, q, \quad (5.29)$$

(ср. с (5.19)). При этом для любого разбиения X_*

$$PL(X_*^*) = \sup_{x \in X} P(x) \leq PL(X_*) \quad (5.30)$$

■

Замечание 3.5.2*. Согласно теореме 3.5.1* любая функция $d^*(\cdot) : X \rightarrow D$, удовлетворяющая условию $d^*(x) = d$, если $x \in X_d^*$, $d \in D$, определит оптимальное четкое правило идентификации. Разумеется, его можно найти и непосредственно, решив для каждого $x \in X$ задачу на минимум

$$P_d(x) \stackrel{\text{def}}{=} \max_{k \in K} \min(l_{k,d}, f^{\xi, \kappa}(x, k)) \sim \min_{d \in D}.$$

Всякое ее решение $d = d^*(x) \in D^*(x)$, $x \in X$.

Взаимно однозначное соответствие между функциями $d^*(\cdot) : X \rightarrow D$ и разбиениями X_* определяет равенство $d^*(x) = \max_{1 \leq d \leq q} (d \cdot \chi^{\delta|\xi}(d|x))$, $x \in X$, в котором $\chi^{\delta|\xi}(d|\cdot)$ — индикаторная функция множества X_d^* , $d = 1, \dots, q$.

3.5.5. Минимаксное правило. Фазификация. Как уже было отмечено в замечании 3.5.1, модель наблюдения естественно определять априорным распределением $f^{\kappa}(\cdot)$ возможностей состояний системы и распределением $f^{\xi|\kappa}(\cdot|k)$ переходной возможности для каждого состояния системы $k = 1, \dots, q$. В тех случаях, когда распределение $f^{\kappa}(\cdot)$ неизвестно, вместо выражения (5.7) для возможности потерь можно,

воспользовавшись выражениями для возможностей маргинальных потерь

$$PL(\pi^{\delta|\xi}|k) = \sup_{x \in X} \max_{1 \leq d \leq q} \min(l_{k,d}, \pi^{\delta|\xi}(d|x), f^{\xi|\mathcal{K}}(x|k)), \quad k = 1, \dots, q, \quad (5.31)$$

для каждого состояния системы, определить *оптимальное минимаксное правило* $\pi^{+\delta|\xi}$ из условия минимума максимальной среди возможностей маргинальных потерь $PL(\pi^{\delta|\xi}|k)$, $k = 1, \dots, q$, (5.31):

$$\max_{1 \leq k \leq q} PL(\pi^{+\delta|\xi}|k) = \min_{\pi^{\delta|\xi}} \max_{1 \leq k \leq q} PL(\pi^{\delta|\xi}|k). \quad (5.32)$$

Формально в правой части (5.32) $\max_{1 \leq k \leq q} PL(\pi^{\delta|\xi}|k)$ совпадает с $PL(\pi^{\delta|\xi})$ (5.7), если в выражении (5.3) для совместного распределения ξ и $\mathcal{K}$ $f^{\mathcal{K}}(k) = 1$, $k = 1, \dots, q$, и, следовательно, в (5.7) $f^{\xi, \mathcal{K}}(x, k) = \min(f^{\xi|\mathcal{K}}(x, k), f^{\mathcal{K}}(k)) = f^{\xi|\mathcal{K}}(x|k)$, $x \in X$, $k = 1, \dots, q$.

Это означает, что *минимаксное правило* $\pi^{+\delta|\xi}$ (5.32) *совпадает с правилом* $\pi^{*\delta|\xi}$ (5.18), *если в теореме 3.5.1 априорное распределение* $f^{\mathcal{K}}(\cdot)$ *задать как равновозможное*: $f^{\mathcal{K}}(k) = 1$, $k = 1, \dots, q$, и, следовательно, — как "наименее благоприятное", ибо $PL(\pi^{+\delta|\xi}) \geq PL(\pi^{*\delta|\xi})$, см. § 3.5.3.

Поскольку «априорная равновозможность» инвариантна относительно выбора шкалы значений возможности, ее можно считать моделью «незнания» априорного распределения $f^{\mathcal{K}}(\cdot)$ состояний системы, а минимаксное правило — частным случаем нечеткого правила идентификации, охарактеризованного в Теореме 3.5.1, или — четкого, охарактеризованного в Теореме 3.5.1* и в замечании 3.5.2*.

3.5.6. Критерий минимума необходимости потерь. *Дуальная* (полярная) возможности потерь $PL(\pi^{\delta|\xi})$ (5.7) необходимость, см. § 3.1.4, потерь¹⁾

$$\begin{aligned} NL(\pi^{\delta|\xi}) &= \vartheta \left(\sup_{x \in X} \max_{1 \leq k \leq q} \min(\vartheta^{-1} \circ l_{k,d}, \pi^{\delta|\xi}(d|x), f^{\xi, \mathcal{K}}(x, k)) \right) = \\ &= \inf_{x \in X} \min_{1 \leq d \leq q} \max(\vartheta \circ \pi^{\delta|\xi}(d|x), N_d(x)), \end{aligned} \quad (5.33)$$

где $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1]$ — непрерывная, строго монотонно убывающая функция, удовлетворяющая условиям: $\vartheta(0) = 1$, $\vartheta(1) = 0$, т. е. $\vartheta(\cdot) \in \Theta$, и

$$N_d(x) \stackrel{\text{def}}{=} \min_{1 \leq k \leq q} \max(l_{k,d}, \vartheta \circ f^{\xi, \mathcal{K}}(x, k)), \quad x \in X, \quad d \in \{1, \dots, q\}. \quad (5.34)$$

¹⁾ Для любой функции $s(\cdot) : Z \rightarrow [0, 1]$ функция $(\vartheta \circ s)(\cdot) \equiv \vartheta \circ s(\cdot)$ определена равенством $\vartheta \circ s(z) \stackrel{\text{def}}{=} \vartheta(s(z))$, $z \in Z$.

Задача отыскания оптимального нечеткого правила идентификации $\pi_*^{\delta|\xi}$, минимизирующего необходимость потерь (5.33),

$$NL(\pi^{\delta|\xi}) \sim \min_{\pi^{\delta|\xi}(\cdot|\cdot)} , \quad (5.35)$$

может быть решена, если для каждого $x \in X$ известно решение более простой задачи на минимум

$$\min_{1 \leq d \leq q} \max(\vartheta \circ \pi^{\delta|\xi}(d|x), N_d(x)) \sim \min_{\pi^{\delta|\xi}(\cdot|x)} . \quad (5.36)$$

3.5.7. Правило решения, минимизирующее необходимость потерь, дуальную возможности потерь. В приведенной далее теореме 3.5.2 дано решение задачи (5.36) и показано, что каждое семейство ее решений $\pi_*^{\delta|\xi}(\cdot|x)$, $x \in X$, определяет решение $\pi_*^{\delta|\xi}(\cdot|\cdot) : \{1, \dots, q\} \times X \rightarrow [0, 1]$ задачи (5.35). Теорема 3.5.2 является следствием общего результата, сформулированного в следующей лемме.

Лемма 3.5.3. Пусть $g_j(\cdot) : X \rightarrow [0, 1]$, $j = 1, \dots, q$, и $h_j(\cdot) : X \rightarrow [0, 1]$, $j = 1, \dots, q$, — некоторые функции. Для каждого $x \in X$:
• упорядочим значения $g_1(x), \dots, g_q(x)$ так, чтобы

$$g(x) \stackrel{\text{def}}{=} \min_{1 \leq j \leq q} g_j(x) = g_{i_1}(x) = \dots = g_{i_t}(x) < g_{i_{t+1}} \leq \dots \leq g_{i_q}(x), \text{ и} \quad (5.37)$$

• определим $h_j^*(x)$, $j = 1, \dots, q$, согласно условиям

$$h_j^*(x) \geq 0, \quad j = 1, \dots, q, \quad \min_{1 \leq s \leq t} h_{i_s}^*(x) = 0. \quad (5.38)$$

Тогда для любых функций $h_1(\cdot), \dots, h_q(\cdot)$, таких, что $\min_{1 \leq j \leq q} h_j(x) = 0$, $x \in X$,

$$\begin{aligned} \min_{1 \leq j \leq q} \max(g_j(x), h_j^*(x)) &= \min_{1 \leq s \leq t} \max(g_{i_s}(x), h_{i_s}^*(x)) = \\ &= g(x) \leq \min_{1 \leq j \leq q} \max(g_j(x), h_j(x)), \quad x \in X, \end{aligned} \quad (5.39)$$

и, как следствие,

$$\inf_{x \in X} \min_{1 \leq j \leq q} \max(g_j(x), h_j^*(x)) = \inf_{x \in X} g(x) \leq \inf_{x \in X} \min_{1 \leq j \leq q} \max(g_j(x), h_j(x)). \quad (5.40)$$

Доказательство. Соотношение (5.39) непосредственно следует из условий (5.37), (5.38), неравенство (5.40) следует из (5.39). ■

Теорема 3.5.2. Для каждого $x \in X$ упорядочим значения $N_1(x), \dots, N_q(x)$ (5.34) так, чтобы

$$0 \leq N(x) \stackrel{\text{def}}{=} \min_{1 \leq d \leq q} N_d(x) = N_{i_1}(x) = \dots = N_{i_t}(x) < N_{i_{t+1}} \leq \dots \leq N_{i_q}(x) \quad (5.41)$$

и выберем любое нечеткое правило идентификации $\pi_*^{\delta|\xi}$, удовлетворяющее условию

$$\max_{1 \leq s \leq t} \pi_*^{\delta|\xi}(i_s|x) = 1. \quad (5.42)$$

Тогда для любого нечеткого правила идентификации $\pi_*^{\delta|\xi}$

$$NL(\pi_*^{\delta|\xi}) = \inf_{x \in X} N(x) \leq NL(\pi_*^{\delta|\xi}), \quad (5.43)$$

где функция $N(\cdot)$ определена в (5.41).

Доказательство. Действительно, для $g_j(x) = N_j(x)$, $j = 1, \dots, q$, и $h_j^*(x) = \vartheta \circ \pi_*^{\delta|\xi}(j|x)$, $j = 1, \dots, q$, будут выполнены условия Леммы 3.5.3 для задач (5.36), (5.35). ■

Замечание 3.5.5. В ряде случаев удобна следующая формулировка достаточных условий (5.41), (5.42) оптимальности правила $\pi_*^{\delta|\xi}$. Пусть

$$D_*(x) \stackrel{\text{def}}{=} \{d \in \{1, \dots, q\}, N_d(x) = N(x) \stackrel{\text{def}}{=} \min_{d'} N_{d'}(x)\}, \quad x \in X, \quad (5.44)$$

тогда любое нечеткое правило $\pi_*^{\delta|\xi}$, удовлетворяющее условиям

$$\max_{d \in D_*(x)} \pi_*^{\delta|\xi}(d|x) = 1, \quad (5.45)$$

является решением задач (5.35), (5.34).

Заметим, что если $l_{k,d}|_{k=d} = 0$, $k, d = 1, \dots, q$, то значение $N_d(x)$ (5.34) мало для тех решений $d \in \{1, \dots, q\}$, для которых при данном $x \in X$ мало значение $\vartheta \circ f^{\xi, \mathcal{K}}(x, d)$, или — соответственно велико значение $f^{\xi, \mathcal{K}}(x, d)$. Согласно условиям (5.41), (5.42) для этого $x \in X$ и по крайней мере одного такого решения d $\pi_*^{\delta|\xi}(d|x) = 1$.

Замечание 3.5.6. Пусть в отличие от (5.21) $\tilde{l}_{k,d} = \begin{cases} 0, & \text{если } d = k \\ > 0, & \text{если } d \neq k \end{cases}$, $k, d = 1, \dots, q$. Тогда согласно (5.34)

$$N_d(x) = \vartheta \circ f^{\xi, \mathcal{K}}(x, d), \quad x \in X, \quad d = 1, \dots, q, \quad (5.46)$$

В этом случае в (5.44)

$$D_*(x) = \{d \in \{1, \dots, q\}, f^{\xi, \mathcal{K}}(x, d) = \max_{1 \leq k \leq q} f^{\xi, \mathcal{K}}(x, k) \stackrel{\text{def}}{=} f^{\xi}(x)\}, \quad x \in X, \quad (5.47)$$

и правило $\pi_*^{\delta|\xi}$ в (5.45) минимизирует необходимость ошибки идентификации: $NE(\pi_*^{\delta|\xi}) = \inf_{x \in X} \min_{1 \leq k \leq q} \vartheta \circ f^{\xi, \mathcal{K}}(x, k) = \inf_{x \in X} \vartheta \circ f^{\xi}(x) = 0$.

Замечание 3.5.7. Теорема 3.5.2 характеризует оптимальное правило идентификации и в том случае, когда наблюдения за системой невозможны. Для этого в условиях теоремы $f^{\xi, \mathcal{K}}(x, k)$ следует заменить на $f^{\mathcal{K}}(k) = \sup_{x \in X} f^{\xi, \mathcal{K}}(x, k)$, $N_d(x)$ — на $N_d \stackrel{\text{def}}{=} \min_{1 \leq k \leq n} \max(l_{k,d}, \vartheta \circ f^{\mathcal{K}}(k))$, $\pi_*^{\delta|\xi}(d|x)$ — на $\pi^{\delta}(d)$, $x \in X$, $k \in 1, \dots, q$, $d \in 1, \dots, q$, см. замеча-

ние 3.5.4. В этом случае оптимальное нечеткое правило $\pi_*^\delta(d)$, определяется условиями: $\max_{1 \leq s \leq t} \pi_*^\delta(s) = 1$, $N \stackrel{\text{def}}{=} \min_{1 \leq d \leq n} N_d = N_{d_1} = \dots = N_{d_t} < N_{d_{t+1}} \leq \dots \leq N_{d_q}$, и для него $N_d \leq N_d(x)$, $d = 1, \dots, q$, $x \in X$. Поэтому в случае невозможности наблюдения за системой $NL(\pi_*^\delta) = N \leq \inf_{x \in X} N(x) = NL(\pi_*^{\delta|\xi})$, т. е. *в то время как при невозможности наблюдения за системой возможность потерь, сопутствующих решению $\pi_*^{\delta|\xi}$, возрастает, необходимость потерь, сопутствующих решению π_*^δ убывает*. Дело, грубо говоря, в том, что чем меньше информации о системе используется при принятии решения о ее состоянии, тем более верно, что любое решение $\delta = d$ о ее состоянии нельзя считать непременно ошибочным.

Этот вывод в еще большей степени относится к ситуации, в которой неизвестно априорное распределение состояний системы, поскольку при этом $N_d = \min_{1 \leq k \leq n} \max(l_{k,d}, \vartheta \circ f^x(k)) \geq \min_{1 \leq k \leq n} l_{k,d}$, $d = 1, \dots, q$, и, следовательно, необходимость потерь еще меньше, ибо когда все состояния системы равновозможны, $f^x(k) = 1$, $k = 1, \dots, q$, любое решение о ее состоянии тем более нельзя считать непременно ошибочным.

Замечание 3.5.8. Поскольку правило $\pi_*^{\delta|\xi}$ всегда можно выбрать четким, то есть так, чтобы для любого $x \in X$ $\pi_*^{\delta|\xi}(d(x)|x) = 1$ для некоторого $d = d_*(x) \in D_*(x)$ и $\pi_*^{\delta|\xi}(d|x) = 0$ для всех остальных $d \neq d_*(x)$, то *фазификация правила идентификации не позволяет улучшить качество решения*, и имеет смысл рассмотреть задачу построения непосредственно четкого правила.

Чтобы охарактеризовать четкие правила, минимизирующие необходимость потерь, рассмотрим выражение для необходимости потерь

$$NL(X_*) = \inf_{x \in X} \min_{1 \leq d \leq q} \max(\vartheta \circ \chi^{\delta|\xi}(d|x), N_d(x)), \quad (5.48)$$

свойственных правилу решения, определенному разбиением X_* ; в (5.48) $\chi^{\delta|\xi}(d|x) = \begin{cases} 1, & \text{если } x \in X_d \\ 0, & \text{если } x \in X \setminus X_d \end{cases}$, $x \in X$, — индикаторная функция множества X_d , $d = 1, \dots, q$, и $N_d(x)$ определено в (5.34). Следующая теорема является частным случаем Теоремы 3.5.2.

Теорема 3.5.2*. *Решением задачи на минимум*

$$NL(X_*) \sim \min_{X_*} \quad (5.49)$$

для необходимости потерь (5.48) является любое разбиение X_ , удовлетворяющее условиям*

$$X_{*d} \subset \left\{ x \in X, N_d(x) = N(x) \stackrel{\text{def}}{=} \min_{1 \leq d' \leq q} N_{d'}(x) \right\}, \quad d = 1, \dots, q. \quad (5.50)$$

При этом для любого разбиения X .

$$NL(X_{*}) = \inf_{x \in X} N(x) \leq NL(X).$$

■

Замечание 3.5.8*. Как и при минимизации возможности потерь всякое четкое правило идентификации, минимизирующее необходимость потерь, можно определить как функцию $d(\cdot) : X \rightarrow D$ так, чтобы при наблюдениях $\xi = x$ значение $d(x)$ определяло решение о состоянии системы. Согласно теореме 3.5.2* любая функция $d_*(\cdot) : X \rightarrow D$, удовлетворяющая условию $d_*(x) = d$, если $x \in X_{*d}$, $d \in D$, или, что то же самое, $d_*(x) = \max_{1 \leq d \leq q} (d \cdot \chi_{X_{*d}}(x))$, $x \in X$, определит оптимальное четкое правило идентификации. Разумеется, его можно найти и непосредственно, решив для каждого $x \in X$ задачу на минимум $N_d(x) \stackrel{\text{def}}{=} \min_{1 \leq k \leq q} \max (l_{k,d}, \vartheta \circ f^{\xi, \varkappa}(x, k)) \sim \min_{d \in D}$. Всякое ее решение $d = d_*(x)$ определяется условием $d_*(x) \in D_*(x)$, $x \in X$, и определяет оптимальное четкое правило идентификации, которое в случае (5.21) называется правилом максимальной возможности, поскольку в этом случае согласно (5.46) $d_*(x) = d$, если $f^{\xi, \varkappa}(x, d) = f^{\xi}(x)$, $x \in X$, $d \in D$.

Замечание 3.5.9. В тех нередких случаях, когда априорное распределение $f^{\varkappa}(\cdot)$ возможностей состояний системы неизвестно, необходимость маргинальных потерь для каждого состояния $k = 1, \dots, q$

$$NL(\pi^{\delta|\xi}|k) = \inf_{x \in X} \min_{1 \leq d \leq n} \max (l_{k,d}, \vartheta \circ \pi^{\delta|\xi}(d|x), \vartheta \circ f^{\xi|\varkappa}(x|k)). \quad (5.51)$$

и оптимальное правило $\pi_+^{\delta|\xi}$ следует определять согласно условию

$$\min_{1 \leq k \leq n} NL(\pi_+^{\delta|\xi}|k) = \min_{\pi^{\delta|\xi}} \min_{1 \leq k \leq n} NL(\pi^{\delta|\xi}|k), \quad (5.52)$$

формально совпадающему с условием (5.35), если в последнем априорное распределение $f^{\varkappa}(k) = 1$, $k = 1, \dots, q$.

3.5.8. Нечеткая модель. Оценивание.

3.5.8.1 Модель наблюдения и правило оценивания.

В рассматриваемых далее моделях заданы:

► совместное распределение $f^{\xi, \varkappa}(\cdot, \cdot) : X \times K \rightarrow [0, 1]$ нечетких элементов ξ и $\varkappa$. Значение $f^{\xi, \varkappa}(x, k)$ есть возможность того, что наблюдаемый нечеткий элемент $\xi = x \in X$ и оцениваемый нечеткий элемент $\varkappa = k \in K$; $\varkappa$ моделирует какую-либо характеристику нечеткой системы, ξ — наблюдение за системой. Распределение $f^{\xi, \varkappa}(\cdot, \cdot)$ назовём *возможностной моделью системы и наблюдения за системой*;

► распределение $\pi^{\delta|\xi}(\cdot|\cdot) : D \times X \rightarrow [0, 1]$ переходной возможности $\Pi(\cdot|\cdot) : \mathcal{P}(D) \times X \rightarrow [0, 1]$ для пространств $(X, \mathcal{P}(X))$ и $(D, \mathcal{P}(D))$.

Значение $\pi^{\delta|\xi}(d|x)$ есть возможность оценки $\delta = d$ характеристики $\varkappa$ системы, когда $\xi = x$ — результат наблюдения за системой; функцию $\pi^{\delta|\xi}(\cdot|\cdot)$ назовем *нечётким правилом оценивания*.

► индикаторная функция $l(\cdot, \cdot) = l^\Lambda(\cdot, \cdot) : K \times D \rightarrow [0, 1]$ нечеткого множества Λ , определенного на $(Y, \mathcal{P}(Y), P_Y)$ и принимающего значения в $\mathcal{P}(K \times D)$. Возможность $l(k, d) = P_Y((k, d) \in \Lambda)$ того, что Λ покрывает точку $(k, d) \in K \times D$, интерпретируется как возможность потерь, обусловленных использованием $d \in D$ вместо $k \in K$, в качестве оценки k , см. рис. 3.8; нечеткое множество Λ и нечеткий элемент $(\xi, \varkappa, \delta)$ независимы в смысле одноточечного покрытия, см. § 3.1.6.6.

3.5.9. Правило оценивания, минимизирующее возможность потерь. Возможность потерь при оценивании $\varkappa$ согласно нечёткому правилу $\pi^{\delta|\xi}$

$$\begin{aligned} PL(\pi^{\delta|\xi}) &= \sup_{\substack{x \in X, d \in D \\ k \in K}} \min(l(k, d), \pi^{\delta|\xi}(d|x), f^{\xi, \varkappa}(x, k)) = \\ &= \sup_{x \in X, d \in D} \min(\pi^{\delta|\xi}(d|x), P(d, x)), \end{aligned} \quad (5.53)$$

где

$$P(d, x) \stackrel{\text{def}}{=} \sup_{k \in K} \min(l(k, d), f^{\xi, \varkappa}(x, k)) \quad (5.54)$$

— возможность потерь при $\delta = d$ и $\xi = x$, $x \in X$, $d \in D$.

Рассмотрим задачу оценивания как задачу на минимум для возможности потерь (5.53)

$$PL(\pi^{*\delta|\xi}) = \min_{\pi^{\delta|\xi}} PL(\pi^{\delta|\xi}), \quad (5.55)$$

решение $\pi^{*\delta|\xi}$ которой определит оптимальное нечеткое правило оценивания. В следующей теореме дано решение задачи (5.55).

Теорема 3.5.3. Пусть $P(x) = \min_{d \in D} P(d, x)$ и

$$D^*(x) \stackrel{\text{def}}{=} \{d \in D, P(d, x) = P(x)\}, \quad x \in X. \quad (5.56)$$

Определим нечёткое правило $\pi^{*\delta|\xi}$ оценивания $\varkappa$, удовлетворяющее следующим условиям: для каждого $x \in X$

$$\max_{d \in D^*(x)} \pi^{*\delta|\xi}(d|x) = 1, \quad \max_{d \in D \setminus D^*(x)} \pi^{*\delta|\xi}(d|x) = 0. \quad (5.57)$$

Тогда для любого нечёткого правила $\pi^{\delta|\xi}$ и любого нечёткого правила $\pi^{*\delta|\xi}$, удовлетворяющего условиям (5.57),

$$\begin{aligned} \sup_{d \in D} \min(P(d, x), \pi^{*\delta|\xi}(d|x)) &= \sup_{d \in D^*(x)} \min(P(d, x), \pi^{*\delta|\xi}(d|x)) = \\ &= P(x) \leq \sup_{d \in D} \min(P(d, x), \pi^{\delta|\xi}(d|x)), \quad x \in X, \end{aligned} \quad (5.58)$$

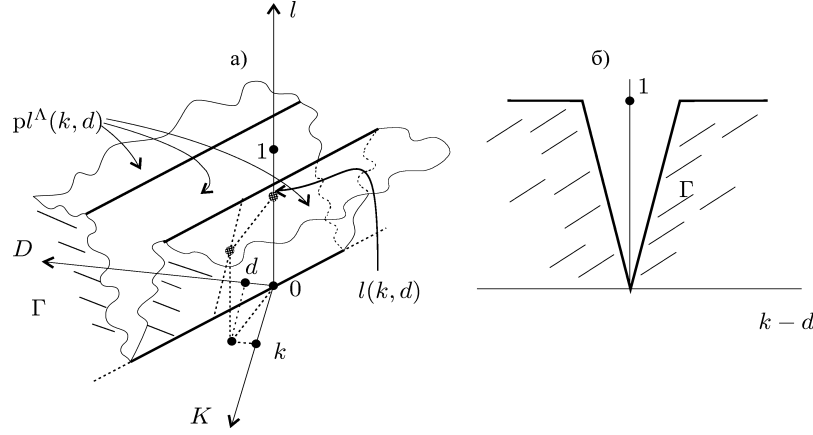


Рис. 3.8. а) График Γ нечеткого множества Λ и график функции $l(\times \times, \cdot) : K \times D \rightarrow [0, 1]$ б) Сечение графика Γ нечеткого множества $\Lambda : Y \rightarrow \mathcal{P}(K \times D)$ плоскостью $\{(k, d, l) \in K \times D \times [0, 1], k + d = 0\}$. В обоих случаях $Y = [0, 1]$, $f(y) = y$, $y \in [0, 1]$, — распределение P_Y , $K = D = \mathcal{R}^1$, $l(k, d) = q|k - d|$, если $q|k - d| \leq 1$, и $l(k, d) = 1$, если $q|k - d| > 1$, $k, d \in \mathcal{R}^1$, $q = \text{const} > 1$.

и соответственно

$$PL(\pi^{*\delta|\xi}) = \sup_{x \in X} P(x) \leq PL(\pi^{\delta|\xi}), \quad (5.59)$$

т. е. любое нечёткое правило $\pi^{*\delta|\xi}$ является оптимальным.

Доказательство. Согласно условиям (5.56) и (5.57) выполнены оба равенства в (5.58). Правая часть в (5.58) равна

$$\max\left(\sup_{d \in D^*(x)} \min(P(x), \pi^{\delta|\xi}(d|x)), \sup_{d \in D \setminus D^*(x)} \min(P(d, x), \pi^{\delta|\xi}(d|x))\right) \quad (5.60)$$

и, если $\max_{d \in D^*(x)} \pi^{\delta|\xi}(d|x) = 1$, то выражение (5.60) равно

$$\max(P(x), \sup_{d \in D \setminus D^*(x)} \min(P(d, x), \pi^{\delta|\xi}(d|x))) \geq P(x),$$

а если $\max_{d \in D \setminus D^*(x)} \pi^{\delta|\xi}(d|x) = 1$, то в силу (5.56) выражение (5.60) не меньше, чем

$$\sup_{d \in D \setminus D^*(x)} \min(P(d, x), \pi^{\delta|\xi}(d|x)) > P(x), \quad x \in X.$$

Соотношения (5.59) следуют из (5.58). ■

Замечание 3.5.10. Поскольку согласно теореме 3.5.3 оптимальное правило $\pi^{*\delta|\xi}$ всегда можно выбрать четким, то есть так, чтобы при любом $x \in X$ для некоторого $d^* \in D^*(x)$ $\pi^{*\delta|\xi}(d^*|x) = 1$, а для остальных

$d \in D$, $d \neq d^*$ $\pi^{*\delta|\xi}(d|x) = 0$, то фазификация правила оценивания не позволяет улучшить его качество.

3.5.10. Четкое правило оценивания. Согласно замечанию 3.5.10 определим четкое правило оценивания $\varkappa$ как функцию $d(\cdot) : X \rightarrow D$, ставящую в соответствие каждому результату наблюдения $\xi = x$ значение $d(x)$ в качестве оценки $\varkappa$, а сопутствующую возможность потерь определим выражением

$$PL(d(\cdot)) = \sup_{x \in X, k \in K} \min(l(k, d(x)), f^{\xi, \varkappa}(x, k)). \quad (5.61)$$

Оптимальное правило оценивания определим как решение $d^*(\cdot)$ задачи на минимум для возможности потерь (5.61)

$$PL(d^*(\cdot)) = \min_{d(\cdot): X \rightarrow D} PL(d(\cdot)), \quad (5.62)$$

которое можно получить, решив для каждого $x \in X$ задачу

$$P(d, x) = \sup_{k \in K} \min(l(k, d), f^{\xi, \varkappa}(x, k)) \sim \min_{d \in D}. \quad (5.63)$$

Всякое семейство решений $d^* = d^*(x)$, $x \in X$, задачи (5.63) является решением задачи (5.62); этот факт следует из теоремы 3.5.3.

Теорема 3.5.3*. 1. Пусть $d^*(x)$, $x \in X$, — некоторое семейство решений задачи (5.63), т. е. пусть $d^*(x) \in D^*(x)$, $x \in X$, см. (5.56). Тогда $d^*(\cdot) : X \rightarrow D$ — четкое правило оценивания, для которого

$$PL(d^*(\cdot)) = \min_{d(\cdot): X \rightarrow D} PL(d(\cdot)) = \sup_{x \in X} P(x).$$

2. Пусть $\hat{d}(x)$, $x \in X$, — любое семейство решений задачи $P(d|x) \stackrel{\text{def}}{=} \sup_{k \in K} \min(l(k, d), f^{\varkappa|\xi}(k|x)) \sim \min_{d \in D}$, $x \in X$, где $f^{\varkappa|\xi}(k|x)$, $k \in K$, — произвольный вариант условного при условии $\xi = x$ распределения $\varkappa$. Тогда $PL(\hat{d}(x), x) = PL(d^*(x), x) \stackrel{\text{def}}{=} P(x)$, $x \in X$, и, следовательно,

$$PL(\hat{d}(\cdot)) = \sup_{x \in X} P(x) = PL(d^*(\cdot)).$$

Доказательство. Первое утверждение следует из теоремы 3.5.3. Для доказательства второго утверждения заметим, что согласно определению функций $d^*(\cdot)$ и $\hat{d}(\cdot)$

$$P(d^*(x), x) = \min_{d \in D} P(d, x) \leq P(\hat{d}(x), x), \quad x \in X, \quad (5.64)$$

и

$$P(\hat{d}(x)|x) = \min_{d \in D} P(d|x) \leq P(d^*(x)|x), \quad x \in X. \quad (5.65)$$

Так как $f^{\xi, \varkappa}(x, k) = \min(f^{\varkappa|\xi}(k|x), f^{\xi}(x))$, $k \in K$, $x \in X$, где $f^{\xi}(x) = \sup_{k \in K} f^{\xi, \varkappa}(x, k)$, $x \in X$, то соответственно неравенству

(5.65) $P(\hat{d}(x), x) = \min (P(\hat{d}(x)|x), f^\xi(x)) \leq \min (P(d^*(x)|x), f^\xi(x)) = P(d^*(x), x)$, $x \in X$. Отсюда и из неравенства (5.65) следуют утверждения в 2. ■

Замечание 3.5.11. Если $K = D$ и

$$l(k, d) = l^\circ(k, d) \stackrel{\text{def}}{=} \begin{cases} 0, & d = k, \\ 1, & d \neq k, \end{cases} \quad k \in K, d \in D, \quad (5.66)$$

то в (5.54)

$$P(d, x) = \sup_{k \in K, k \neq d} f^{\xi, \varkappa}(x, k), \quad d \in D, x \in X, \quad (5.67)$$

а в качестве $d^* = d^*(x)$ можно выбрать любое значение

$$d^* \in \{d \in D, f^{\xi, \varkappa}(x, d) = f^\xi(x) \stackrel{\text{def}}{=} \max_{k' \in K} f^{\xi, \varkappa}(x, k')\} \subset D^*(x), \quad x \in X. \quad (5.68)$$

В этом случае правило $d^*(\cdot)$, минимизирующее *возможность ошибки оценивания* $PE(d(\cdot))$, назовем *правилом максимальной возможности*.

3.5.11. Правило, минимизирующее необходимость потерь, дуальную возможности. Для определенной в § 3.5.8.1 модели наблюдения необходимость потерь, *дуальная возможности* $PL(\pi^{\delta|\xi})$ (5.53), свойственная нечёткому правилу $\pi^{\delta|\xi}(\cdot|\cdot) : D \times X \rightarrow [0, 1]$,

$$\begin{aligned} NL(\pi^{\delta|\xi}) &= \vartheta \left(\sup_{\substack{x \in X, d \in D \\ k \in K}} \min(\vartheta^{-1} \circ l(k, d), \pi^{\delta|\xi}(d|x), f^{\xi, \varkappa}(x, k)) \right) = \\ &= \inf_{x \in X, d \in D} \max(N(d, x), \vartheta \circ \pi^{\delta|\xi}(d|x)), \end{aligned} \quad (5.69)$$

где

$$N(d, x) \stackrel{\text{def}}{=} \inf_{k \in K} \max(l(k, d), \vartheta \circ f^{\xi, \varkappa}(x, k)), \quad d \in D, x \in X, \quad (5.70)$$

и функция $\vartheta(\cdot) \in \Theta$.

Решение задачи определения оптимального нечёткого правила оценивания $\pi_*^{\delta|\xi}$, минимизирующего необходимость потерь,

$$NL(\pi_*^{\delta|\xi}) = \min_{\pi^{\delta|\xi}(\cdot|\cdot)} NL(\pi^{\delta|\xi}), \quad (5.71)$$

может быть получено на основе решения семейства задач на минимум

$$\inf_{d \in D} \max(N(d, x), \vartheta \circ \pi^{\delta|\xi}(d|x)) \sim \min_{\pi^{\delta|\xi}(\cdot|x)} \quad (5.72)$$

для каждого $x \in X$. Это показано в следующей Теореме.

Теорема 3.5.4. Пусть $N(x) = \min_{d \in D} N(d, x)$, $x \in X$, и

$$D_*(x) \stackrel{\text{def}}{=} \{d \in D, N(d, x) = N(x)\}, \quad x \in X. \quad (5.73)$$

Тогда минимум в (5.72) достигается на любом нечётком правиле $\pi_*^{\delta|\xi}$, удовлетворяющем условию: для любого $x \in X$

$$\max_{d \in D_*(x)} \pi_*^{\delta|\xi}(d|x) = 1. \quad (5.74)$$

Для любого такого правила оценивания $\pi_*^{\delta|\xi}$ и любого ¹⁾ правила $\pi^{\delta|\xi}$

$$\begin{aligned} \inf_{d \in D} \max(N(d, x), \vartheta \circ \pi_*^{\delta|\xi}(d|x)) &= \inf_{d \in D_*(x)} \max(N(x), \vartheta \circ \pi_*^{\delta|\xi}(d|x)) = \\ &= N(x) \leq \inf_{d \in D} \max(N(d, x), \vartheta \circ \pi^{\delta|\xi}(d|x)), \quad x \in X, \end{aligned} \quad (5.75)$$

и соответственно

$$NL(\pi_*^{\delta|\xi}) = \inf_{x \in X} N(x) \leq NL(\pi^{\delta|\xi}). \quad (5.76)$$

Доказательство. Действительно, согласно (5.69), (5.70), (5.73) и условию (5.74) $\inf_{d \in D} \max(N(d, x), \vartheta \circ \pi_*^{\delta|\xi}(d|x)) = \min_{d \in D_*(x)} N(d, x) \stackrel{\text{def}}{=} N(x) \leq \inf_{d \in D} \max(N(d, x), \vartheta \circ \pi^{\delta|\xi}(d|x))$, $x \in X$, откуда следует и неравенство (5.76). ■

Замечание 3.5.12. Выберем любую функцию $d_*(x) \in D_*(x)$, $x \in X$, и определим $\pi_*^{\delta|\xi}$ так, чтобы в согласии с (5.74) для каждого $x \in X$ $\pi_*^{\delta|\xi}(d_*(x)|x) = 1$, а для остальных $d \neq d_*(x)$ $\pi_*^{\delta|\xi}(d|x) = 0$. При таком выборе функция $d_*(\cdot) : X \rightarrow D$ определяет четкое правило оценивания, ставящее в соответствие наблюдению $\xi = x$ значение $d_*(x)$ в качестве оценки $\varkappa$, для которого необходимость потерь равна $NL(\pi_*^{\delta|\xi}) = \inf_{x \in X} N(x)$, т. е. *фазификация правила оценивания и в этом случае не позволяет улучшить качество оценивания.*

В связи с замечанием 3.5.12 рассмотрим задачу определения четкого правила оптимального оценивания $d_*(\cdot) : X \rightarrow D$, минимизирующего необходимость потерь

$$NL(d(\cdot)) = \inf_{x \in X, k \in K} \max(l(k, d(x)), \vartheta \circ f^{\xi, \varkappa}(x, k)) \sim \min_{d(\cdot) : X \rightarrow D}, \quad (5.77)$$

дуальную возможности потерь (5.61), ср. с замечанием 3.5.8*. Решение $d_*(\cdot) : X \rightarrow D$ задачи (5.77) можно получить в виде семейства $d_*(x)$, $x \in X$, решений задачи

$$N(d, x) = \inf_{k \in K} \max(l(k, d), \vartheta \circ f^{\xi, \varkappa}(x, k)) \sim \min_{d \in D}, \quad x \in X. \quad (5.78)$$

¹⁾ $\max_{d \in D} \pi^{\delta|\xi}(d|x) = 1$, ибо при любом $x \in X$, $\pi^{\delta|\xi}(\cdot|x) : D \rightarrow [0, 1]$ — распределение переходной возможности.

Теорема 3.5.4*. Пусть $d_*(x)$, $x \in X$, — некоторое семейство решений задач (5.78), т. е. пусть $d_*(x) \in D_*(x)$, $x \in X$, см. (5.70), (5.73). Тогда $d_*(\cdot) : X \rightarrow D$ — четкое правило оценивания, минимизирующее необходимость потерь (5.77).

Доказательство следует теоремы 3.5.4 и замечания 3.5.12. ■

Замечание 3.5.13. Если $K = D$ и $l(k, d)|_{k=d} = 0$, $k \in K$, то $N L(\pi_*^{\delta|\xi}) = 0$. Действительно, в таком случае $N(x) \stackrel{\text{def}}{=} \min_{d \in D} N(d, x) = \inf_{k \in K} \max_{d \in D} (\min_{d \in D} l(k, d), \vartheta \circ f^{\xi, \kappa}(x, k)) = \vartheta \circ \sup_{k \in K} f^{\xi, \kappa}(x, k) = \vartheta \circ f^{\xi}(x)$, $x \in X$, и, следовательно, $N L(\pi_*^{\delta|\xi}) = \inf_{x \in X} N(x) = \vartheta \circ (\sup_{x \in X} f^{\xi}(x)) = 0$.

В связи с замечаниями 3.5.12 и 3.5.13 рассмотрим подробнее четкое правило оптимального оценивания в случае $K = D$ для функции $l^\circ(\times \times, \cdot)$, определенной в (5.66), и для класса функций $\tilde{l}(\cdot, \cdot)$, определенных условиями

$$\tilde{l}(k, d) = \begin{cases} 0, & \text{если } k = d, \\ > 0, & \text{если } k \neq d, \end{cases} \quad k, d \in K = D. \quad (5.79)$$

Согласно замечанию 3.5.13 для любой функции $\tilde{l}(\cdot, \cdot)$ (5.79) (и, в частности, для $l^\circ(\cdot, \cdot)$) в (5.66)

$$N(x) = \vartheta \circ f^{\xi}(x) \stackrel{\text{def}}{=} \vartheta \circ \sup_{k \in K} f^{\xi, \kappa}(x, k), \quad x \in X. \quad (5.80)$$

Поэтому множества $D_*(x)$, $x \in X$, (5.73) для $l^\circ(\cdot, \cdot)$ (5.66) согласно (5.80) определяются равенствами

$$\begin{aligned} D_*^\circ(x) &= \{d \in D, f^{\xi, \kappa}(x, d) = \\ &= \sup_{k \in K} \min(\vartheta \circ l^\circ(k, d), f^{\xi, \kappa}(x, k)) = f^{\xi}(x)\}, \quad x \in X, \end{aligned} \quad (5.81)$$

а для $\tilde{l}(\cdot, \cdot)$ (5.79) — равенствами

$$\tilde{D}_*(x) = \{d \in D, \sup_{k \in K} \min(\vartheta \circ \tilde{l}(k, d), f^{\xi, \kappa}(x, k)) = f^{\xi}(x)\}, \quad x \in X. \quad (5.82)$$

Так как $\tilde{l}(k, d) \leq l^\circ(k, d)$, $k, d \in K = D$, то для любых $d \in D$ и $x \in X$

$$\begin{aligned} f^{\xi, \kappa}(x, d) &= \sup_{k \in K} \min(\vartheta \circ l^\circ(k, d), f^{\xi, \kappa}(x, k)) \leq \\ &\leq \sup_{k \in K} \min(\vartheta \circ \tilde{l}(k, d), f^{\xi, \kappa}(x, k)) \leq f^{\xi}(x). \end{aligned} \quad (5.83)$$

Следовательно, согласно (5.81), (5.82) и (5.83) $D_*^\circ(x) \subset \tilde{D}_*(x)$, $x \in X$, поэтому любое четкое правило оптимального оценивания $d_*^\circ(x) \in D_*^\circ(x)$, $x \in X$, для $l^\circ(\cdot, \cdot)$ (5.66) является таковым и для каждой функции $\tilde{l}(\cdot, \cdot)$ (5.79).

Подведем итоги.

Лемма 3.5.4. Пусть возможность $\tilde{l}(k, d)$ ошибки, обусловленной использованием $d \in D = K$ вместо $k \in K$, равна нулю при $k = d$, как это определено, например, в (5.79). Тогда для любого четкого правила оценивания $d_*^\circ(\cdot) : X \rightarrow D$, удовлетворяющего условию $d_*^\circ(x) \in D_*^\circ(x) \stackrel{\text{def}}{=} \{d \in D, f^{\xi, \nu}(x, d) = f^\xi(x)\}$, $x \in X$, необходимость ошибки оценивания (5.77) $NE(d_*^\circ(\cdot)) = 0$ при любой функции $l(\cdot, \cdot) = \tilde{l}(\cdot, \cdot)$, а возможность ошибки $PE(d(\cdot))$ (5.61) при $d(\cdot) = d_*^\circ(\cdot)$ минимальна, если $l(\cdot, \cdot) = l^\circ(\cdot, \cdot)$ (5.66). Наконец, правилу оценивания $d_*^*(\cdot) : \mathcal{X} \rightarrow D$, $P(d_*^*(x), x) = \min_{d \in \tilde{D}_*(x)} P(d, x)$, $x \in X$, см. (5.60), сопутствует ошибка, возможность которой минимальна на множестве правил, необходимость ошибки которых равна нулю. Доказательство следует из (5.80) – (5.83) и замечания 3.5.11. ■

3.6. Методы анализа и интерпретации данных измерений

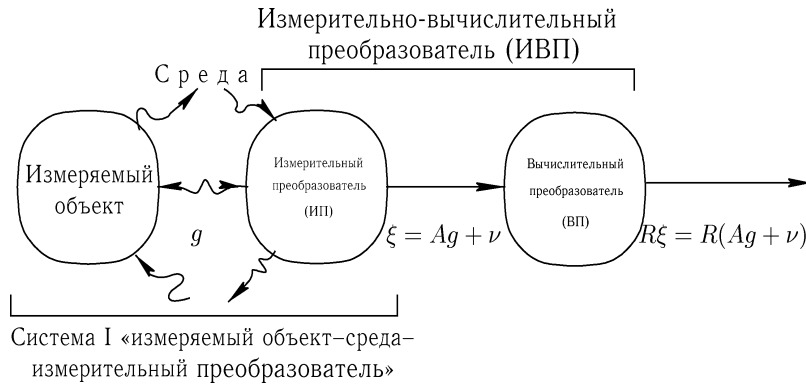


Рис. 3.9. g — входной сигнал ИП, сформированный в системе I при взаимодействии измеряемого объекта, среды и ИП, $\xi = Ag + \nu$ — выходной сигнал ИП, поступивший в ВП, $R\xi$ — выходной сигнал ИВП.

3.6.1. Возможностные модели измерения и его интерпретации [13]. Рассмотрим возможностную модель измерительного эксперимента, заданную совместным распределением возможностей значений следующих нечетких элементов: выходного сигнала ξ измерительного преобразователя (ИП), его входного сигнала γ , сформированного в системе I, см. рис. 3.9, в результате взаимодействия измеряемого объекта, среды и ИП, и характеристик η исследуемого объекта

$$f^{\xi, \gamma, \eta}(x, g, u), (x, g, u) \in \mathcal{X} \times \mathcal{G} \times \mathcal{U}. \quad (6.1)$$

Значение $f^{\xi, \gamma, \eta}(x, g, u)$ равно возможности равенств $\xi = x$, $\gamma = g$, $\eta = u$. Маргинальное распределение

$$f^{\xi, \eta}(x, u) = \sup_{g \in \mathcal{G}} f^{\xi, \gamma, \eta}(x, g, u), \quad (x, u) \in \mathcal{X} \times \mathcal{U}, \quad (6.2)$$

определяет *модель интерпретации измерения*, позволяющую, в частности, получить оценку значения характеристик $\eta = u$, основанную на результате измерения $\xi = x$. При этом, исходя из априорного распределения сигнала ξ

$$f^{\xi}(x) = \sup_{u \in \mathcal{U}} f^{\xi, \eta}(x, u), \quad x \in \mathcal{X}, \quad (6.3)$$

можно оценить и *состоятельность модели эксперимента*. Если, например, $\xi = x$ — результат измерения и $f^{\xi}(x) = 0$, то модель (6.1) следует признать неадекватной.

Задачу интерпретации измерения можно понимать как задачу оптимального оценивания характеристик исследуемого объекта, минимизирующего, например, возможность потерь, обусловленных ошибкой оценивания $PL(d(\cdot)) = \sup_{x \in \mathcal{X}} \min_{u \in \mathcal{U}} (f^{\xi, \eta}(x, u), l(u, d(x))) \sim \min_{d(\cdot): \mathcal{X} \rightarrow \mathcal{U}}$ или

— их необходимость $NL(d(\cdot)) = \inf_{x \in \mathcal{X}} \max_{u \in \mathcal{U}} (\vartheta \circ f^{\xi, \eta}(x, u), l(u, d(x))) \sim \min_{d(\cdot): \mathcal{X} \rightarrow \mathcal{U}}$, см. § 3.3.4. Здесь функция $d(\cdot): \mathcal{X} \rightarrow \mathcal{U}$ определяет

правило оценивания, согласно которому результату измерения $\xi = x$ ставится в соответствие значение $\eta = u = d(x)$ характеристик исследуемого объекта. Оптимальное правило определяется из условия $PL(d^*(\cdot)) = \min_{d(\cdot)} PL(d(\cdot))$, минимизирующего возможность потерь, или

— из условия $NL(d_*(\cdot)) = \min_{d(\cdot)} NL(d(\cdot))$, минимизирующего необходимость потерь, либо, наконец, если $\mathcal{D}_*$ — множество решений последней задачи, содержащее более одного решения $d_*(\cdot)$, то оптимальное правило оценивания $d^*(\cdot)$ может быть найдено из условия $PL(d^*(\cdot)) = \min_{d_*(\cdot) \in \mathcal{D}_*} PL(d_*(\cdot))$.

Для получения распределения (6.1), определяющего модель эксперимента, заметим, что система I (см. рис. 3.9) может быть охарактеризована парой нечетких элементов ξ, γ , и ее модель (модель измерения) естественно задать распределением переходной возможности, см. § 3.1.6.5,

$$f^{\xi|\gamma}(x|g), \quad x \in \mathcal{X}, \quad g \in \mathcal{G}, \quad (6.4)$$

определяющей зависимость распределения нечеткого элемента ξ от значения $g \in \mathcal{G}$ нечеткого элемента γ , и распределением

$$f^{\gamma}(g), \quad g \in \mathcal{G}, \quad (6.5)$$

представляющим априорную информацию о возможных значениях сигнала γ , сформированного в системе I. Равенство

$$f^{\xi, \gamma}(x, g) = \min(f^{\xi|\gamma}(x|g), f^{\gamma}(g)), \quad (x, g) \in \mathcal{X} \times \mathcal{G}, \quad (6.6)$$

определяет распределение пары ξ, γ , т. е. *модель измерения*.

Модель интерпретации входного сигнала измерительного преобразователя (ИП) удобно охарактеризовать распределением переходной возможности

$$f^{\eta|\gamma}(u|g), \quad u \in \mathcal{U}, \quad g \in \mathcal{G}, \quad (6.7)$$

и априорным распределением (6.5). При этом *модель интерпретации входного сигнала* ИП будет задана совместным распределением

$$f^{\gamma, \eta}(g, u) = \min(f^{\eta|\gamma}(u|g), f^{\gamma}(g)), \quad (g, u) \in \mathcal{G} \times \mathcal{U}. \quad (6.8)$$

А поскольку, как нетрудно увидеть, распределение переходной возможности

$$f^{\xi|\gamma, \eta}(x|g, u) = f^{\xi|\gamma}(x|g), \quad x \in \mathcal{X}, \quad g \in \mathcal{G}, \quad u \in \mathcal{U}, \quad (6.9)$$

не зависит от u , то для распределения (6.1) найдем

$$\begin{aligned} f^{\xi, \gamma, \eta}(x, g, u) &= \min(f^{\xi|\gamma}(x|g), f^{\gamma, \eta}(g, u)) = \\ &= \min(f^{\xi|\gamma}(x|g), f^{\eta|\gamma}(u|g), f^{\gamma}(g)), \quad x \in \mathcal{X}, \quad g \in \mathcal{G}, \quad u \in \mathcal{U}. \end{aligned} \quad (6.10)$$

Заметим, что в рассматриваемом случае модель эксперимента (6.1) состоятельна, если и только если состоятельна модель измерения (6.6), а состоятельность последней определяется возможностью $f^{\xi}(x) = \sup_{g \in \mathcal{G}} f^{\xi, \gamma}(x, g)$ результата измерения $\xi = x \in \mathcal{X}$, а именно, модель несостоятельна, если $f^{\xi}(x) = 0$.

Пример 3.6.1. Пусть модель схемы измерения $\xi = A\gamma + \nu$, см. (6.20), задана распределением $f^{\nu, \gamma}(y, g)$, $y \in \mathcal{X}$, $g \in \mathcal{G}$. Тогда возможностная модель измерения (6.6) $f^{\xi, \gamma}(x, g) = \sup\{f^{\nu, \gamma}(y, g) | y \in \mathcal{X}, x = Ag + y\} = f^{\nu, \gamma}(x - Ag, g)$, $x \in \mathcal{X}$, $g \in \mathcal{G}$, причем если как в случае, рассмотренном в § 3.6.3, ν и γ независимы, то есть если $f^{\nu, \gamma}(y, g) = \min(f^{\nu}(y), f^{\gamma}(g))$, $y \in \mathcal{X}$, $g \in \mathcal{G}$, то

$$f^{\xi, \gamma}(x, g) = \min(f^{\nu}(x - Ag), f^{\gamma}(g)), \quad x \in \mathcal{X}, \quad g \in \mathcal{G}, \quad (6.11)$$

— возможностная модель измерения, и в (6.6) $f^{\xi|\gamma}(x|g) = f^{\nu}(x - Ag)$, $x \in \mathcal{X}$, $g \in \mathcal{G}$. Если в (6.7) связь между $g \in \mathcal{G}$ и $u \in \mathcal{U}$ «четкая» и задана равенством

$$f^{\eta|\gamma}(u|g) = \begin{cases} 1, & \text{если } u = Ug, \\ 0, & \text{если } u \neq Ug, \end{cases} \quad g \in \mathcal{G}, \quad u \in \mathcal{U}, \quad (6.12)$$

то $f^{\gamma, \eta}(g, u) = \min(f^{\eta|\gamma}(u|g), f^{\gamma}(g)) = \begin{cases} f^{\gamma}(g), & \text{если } u = Ug, \\ 0, & \text{если } u \neq Ug, \end{cases}$
 $g \in \mathcal{G}$, $u \in \mathcal{U}$, — модель интерпретации входного сигнала (6.8). Согласно (6.11), (6.12) для модели эксперимента

$$(6.10) \quad \text{найдем } f^{\xi, \gamma, \eta}(x, g, u) = \min(f^\nu(x - Ag), f^{\eta|\gamma}(u|g), f^\gamma(g)) = \\ = \begin{cases} \min(f^\nu(x - Ag), f^\gamma(g)), & \text{если } u = Ug, \\ 0, & \text{если } u \neq Ug, \end{cases} \quad x \in \mathcal{X}, \quad g \in \mathcal{G}, \quad u \in \mathcal{U}. \\ \text{Поэтому в (6.2) } f^{\xi, \eta}(x, u) = \sup_{g \in \mathcal{G}} \min_{u=Ug} (f^\nu(x - Ag), f^\gamma(g)), \quad x \in \mathcal{X}, \\ u \in \mathcal{U}, \text{ и в (6.3) } f^\xi(x) = \sup_{g \in \mathcal{G}} \min (f^\nu(x - Ag), f^\gamma(g)), \quad x \in \mathcal{X}.$$

3.6.2. Редукция измерения при априори произвольном измеряемом сигнале. Пусть в измерительном эксперименте регистрируется

$$\xi = Ag + \nu \quad (6.13)$$

— искаженный шумом $\nu \in \mathcal{X}$ выходной сигнал Ag ИП, на вход которого поступил априори произвольный сигнал $g \in \mathcal{G}$ от *измеряемого* объекта и среды, $u = Ug \in \mathcal{U}$ — характеристики *исследуемого* объекта, $f^\nu(\cdot) : \mathcal{X} \rightarrow [0, 1]$ — распределение нечеткого элемента ν , $A : \mathcal{G} \rightarrow \mathcal{X}$, $U : \mathcal{G} \rightarrow \mathcal{U}$ — заданные линейные операторы, первый моделирует ИП, модель которого и модель схемы измерений (6.13) обозначим $[A, f^\nu(\cdot)]$, второй моделирует связь между измеряемым в (6.13) сигналом g и характеристиками u *исследуемого* объекта, не возмущенного измерением, $\mathcal{X}, \mathcal{G}$ и $\mathcal{U}$ — конечномерные евклидовы пространства; в равенстве (6.13) ξ нечеткий элемент $\mathcal{X}$, $f^\xi(x, g) = f^\nu(x - Ag)$, $x \in \mathcal{X}$, — распределение, зависящее от $g \in \mathcal{G}$. В рассматриваемой далее задаче интерпретации измерения (6.13) требуется определить правило оценивания $d(\cdot) : \mathcal{X} \rightarrow \mathcal{U}$ так, чтобы элемент $d(\xi)$ можно было считать в известном смысле оптимальной версией значения Ug характеристик исследуемого объекта.

Оператор U моделирует идеальный ИП, модель которого обозначим $[U, 0]$, на его выходе исследователь получает значения характеристик исследуемого объекта, неискаженные измерением. Рассматриваемая задача называется задачей редукции измерения, выполненного по схеме (6.13), к виду, свойственному измерению на идеальном ИП $[U, 0]$, или — задачей редукции ИП $[A, f^\nu(\cdot)]$ к идеальному ИП $[U, 0]$, $[A, f^\nu(\cdot)] \sim [U, 0]$, см. [13, 18].

Определим индикаторную функцию $l(\cdot, \cdot) = l^\Lambda(\cdot, \cdot) : \mathcal{U} \times \mathcal{U} \rightarrow [0, 1]$ нечеткого множества Λ со значениями в $\mathcal{P}(\mathcal{U} \times \mathcal{U})$, значение $l(Ug, u)$ которой есть возможность потерь, сопутствующих выбору $u \in \mathcal{U}$ в качестве значения характеристик Ug исследуемого объекта.

Охарактеризуем качество правила $d(\cdot) : \mathcal{X} \rightarrow \mathcal{U}$ оценивания значения функции Ug параметра $g \in \mathcal{G}$ распределения $f^\xi(x, g) = f^\nu(x - Ag)$, $x \in \mathcal{X}$, величиной *необходимости потерь, возникающих при использовании $d(\xi)$ вместо Ug* , см. § 3.3.4,

$$\text{NL}(d(\cdot)) \stackrel{\text{def}}{=} \vartheta \left(\sup_{x \in \mathcal{X}} \sup_{g \in \mathcal{G}} \min (f^\nu(x - Ag), \vartheta^{-1} \circ l(Ug, d(x))) \right). \quad (6.14)$$

Оптимальное правило $d_*(\cdot)$ определим из условия $NL(d_*(\cdot)) = \min_{d(\cdot): \mathcal{X} \rightarrow \mathcal{U}} NL(d(\cdot))$. Эта задача рассмотрена в § 3.3.4, согласно полученным там результатам $d_*(x)$, $x \in X$, можно найти, решив для каждого $x \in \mathcal{X}$ задачу

$$\vartheta^{-1}(N(d, x)) = \sup_{g \in \mathcal{G}} \min(f^\nu(x - Ag), \vartheta^{-1} \circ l(Ug, d)) \sim \max_{d \in \mathcal{U}}. \quad (6.15)$$

Рассмотрим важный класс задач (6.15), в которых

$$l(u, v) = \begin{cases} > 0, & \text{если } u \neq v, \\ 0, & \text{если } u = v, \end{cases} \quad u, v \in \mathcal{U}, \quad (6.16)$$

определяет возможность ошибки при оценивании $u \in \mathcal{U}$ значением $v \in \mathcal{U}$. Так как $\max_{d \in \mathcal{U}} \vartheta^{-1} \circ l(Ug, d) = 1$ достигается при единственном $d = Ug$, то $\max_{d \in \mathcal{U}} \sup_{g \in \mathcal{G}} \min(f^\nu(x - Ag), \vartheta^{-1} \circ l(Ug, d)) = \sup_{g \in \mathcal{G}} \min(f^\nu(x - Ag), \max_{d \in \mathcal{U}} \vartheta^{-1} \circ l(Ug, d)) = \sup_{g \in \mathcal{G}} f^\nu(x - Ag)$. Следовательно, $d_*(x) = Ug(x)$, где $g(x)$ — оценка g максимальной возможности: $f^\nu(x - Ag(x)) = \max_{g \in \mathcal{G}} f^\nu(x - Ag)$, $x \in \mathcal{X}$.

Пусть $f^\nu(\cdot) = r(\|\Sigma^{-1/2} \cdot\|)$, где $\Sigma: \mathcal{X} \rightarrow \mathcal{X}$ — положительно определенный оператор (аналог корреляционного оператора¹⁾ ошибки измерения в вероятностной модели $[A, \Sigma_\nu]$, см. § 6 гл. 2), $\|\cdot\|$ — евклидова норма, $r(\cdot): [0, \infty) \rightarrow [0, 1]$ непрерывная строго монотонно убывающая функция, задающая вариант распределения нечеткого вектора ошибки $\nu \in \mathcal{X}$, $r(0) = 1$. При таких предположениях задача (6.15) эквивалентна задаче отыскания оценки максимальной возможности $g \in \mathcal{G}$ при $\xi = x$ как решения задачи на минимум

$$\|\Sigma^{-1/2}(x - Ag)\| \sim \min_{g \in \mathcal{G}}, \quad (6.17)$$

и последующего определения $d_*(x) = Ug(x)$, где $g(x)$ — значение g , на котором минимум в (6.17) достигается; решение $d_*(\cdot)$ не зависит от выбора функции $r(\cdot)$.

Согласно результатам, полученным в § 6 гл. 2, если $A^*\Sigma^{-1}A: \mathcal{G} \rightarrow \mathcal{G}$ — невырожденный оператор, то задача (6.17) имеет единственное решение $g = g(\xi) = (A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}\xi$, при этом $\min_{g \in \mathcal{G}} \|\Sigma^{-1/2}(x - Ag)\| =$

¹⁾ При таком определении распределения $f^\nu(\cdot)$ ошибки измерения можно сравнивать результаты редукции для моделей $[A, \Sigma_\nu]$ и $[A, f^\nu(\cdot)]$, поскольку в обеих моделях распределения как вероятностей так и возможностей ошибки измерения убывают с увеличением $\|\Sigma_\nu^{-1/2}x\|$, $x \in \mathcal{X}$, и сохраняют постоянные значения на эллипсоидах $\{x \in \mathcal{X}, \|\Sigma_\nu^{-1/2}x\| = \text{const}\}$, причем распределение $f^\nu(\cdot) = r(\|\Sigma_\nu^{-1/2} \cdot\|)$ является вариантом продолженного на $\mathcal{P}(\mathcal{X})$ распределения возможностей, максимально согласованного с «гранулированным» нормальным распределением $\mathcal{N}(0, \Sigma_\nu)$, см. [13].

$= \|(I - \Sigma^{-1/2}A(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2})\Sigma^{-1/2}x\|$, и оптимальной оценкой $d_*(\xi)$ значения Ug является нечеткий элемент

$$d_*(\xi) = U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}\xi. \quad (6.18)$$

Заметим, что такая же формула определяет наилучшую в среднем квадратичном линейную минимаксную оценку Ug , если в равенстве $\xi = Ag + \nu$ шум ν является случайным элементом с нулевым математическим ожиданием и корреляционным оператором Σ , см. § 6 гл. 2.

При правиле (6.18) *необходимость ошибки*

$$NE(d_*(\cdot)) = \vartheta(\sup_{x \in \mathcal{X}} r(\|(I - \Sigma^{-1/2}A(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2})\Sigma^{-1/2}x\|)) = 0. \quad (6.19)$$

Точная верхняя грань здесь равна единице и достигается на любом $x \in \mathcal{R}(A)$.

Подведем итоги.

Теорема 3.6.1. Пусть в схеме измерения (6.13) $A: \mathcal{G} \rightarrow \mathcal{X}$ — заданный линейный невырожденный оператор, g — априори произвольный элемент $\mathcal{G}$, ν — нечеткий элемент $\mathcal{X}$, $f^\nu(\cdot): \mathcal{X} \rightarrow [0, 1]$ — распределение ν , т. е. пусть задана модель $[A, f^\nu(\cdot)]$ схемы измерения (6.13). Если $f^\nu(x) = r(\|\Sigma^{-1/2}x\|)$, где $\Sigma: \mathcal{X} \rightarrow \mathcal{X}$ — положительно определенный оператор, $r(\cdot): [0, \infty) \rightarrow [0, 1]$ — непрерывная строго монотонно убывающая функция, $r(0) = 1$, то оптимальной оценкой Ug , где $U: \mathcal{G} \rightarrow \mathcal{U}$ — заданный линейный оператор, является единственный нечеткий элемент (6.18), минимизирующий необходимость ошибки оценивания (6.14), если функция $l(\cdot, \cdot): \mathcal{U} \times \mathcal{U} \rightarrow [0, 1]$ удовлетворяет условию (6.16). Для оценки (6.18) необходимость ошибки оценивания равна нулю.

Замечание 3.6.1. Если $l(u, v) = \begin{cases} 0 & \text{при } u = v, \\ 1 & \text{при } u \neq v, \end{cases} \quad u, v \in \mathcal{U}$, и качество правила $d(\cdot): \mathcal{X} \rightarrow \mathcal{U}$ оценивания охарактеризовать значением возможности $PE(d(\cdot)) = \sup_{x \in \mathcal{X}, g \in \mathcal{G}} \min(f^\nu(x - Ag), l(Ug, d(x)))$ ошибки, а оптимальное правило $d^*(\cdot)$ определить из условия $PE(d(\cdot)) \sim \min_{d(\cdot)}$, то правило $d_*(\cdot)$ (6.18) будет оптимальным, и в этом случае $d^*(\cdot) = d_*(\cdot)$, хотя при этом возможность ошибки $PE(d^*(\cdot)) = \sup_{x \in \mathcal{X}} \sup_{g \in \mathcal{G}} \min(f^\nu(x - Ag), l(Ug, d^*(\cdot))) = 1$.

Замечание 3.6.2. Если $r(\cdot): [0, \infty) \rightarrow [0, 1]$ строго монотонно убывает на $[0, 1]$, причем $r(z) = 0, z \in [\Delta, \infty]$, т. е. если ошибка ν в (6.13) не может быть сколь угодно велика, то кроме фактов, приведенных в теореме 3.6.1, может быть оценена состоятельность модели ИП $[A, f^\nu(\cdot)]$. В частности, если для результата измерения $\xi = x$ в (6.19) $\|(I - \Sigma^{-1/2}A(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2})\Sigma^{-1/2}\xi\| \geq \Delta$, то модель $[A, f^\nu(\cdot)]$ противоречит этому результату измерения.

3.6.3. Редукция измерения при нечеткой априорной информации об измеряемом сигнале. Рассмотрим задачу редукции измерения, в которой не только шум ν , но и измеряемый в (6.13) сигнал g моделируется как нечеткий элемент. Обозначим его γ и запишем схему измерения (6.13) в виде

$$\xi = A\gamma + \nu. \quad (6.20)$$

Пусть в (6.20) $A: \mathcal{G} \rightarrow \mathcal{X}$ — линейный оператор, $\gamma \in \mathcal{G}$ и $\nu \in \mathcal{X}$ суть независимые нечеткие элементы, принимающие значения в евклидовых пространствах $\mathcal{G}$ и $\mathcal{X}$ соответственно, $f^\gamma(\cdot): \mathcal{G} \rightarrow [0, 1]$ и $f^\nu(\cdot): \mathcal{X} \rightarrow [0, 1]$ — их распределения, короче говоря, пусть задана нечеткая модель¹⁾ $[A, f^\gamma(\cdot), f^\nu(\cdot)]$ схемы измерения (6.20). Обозначим $U\gamma$ нечеткий элемент, моделирующий характеристики *исследуемого объекта*, где $U: \mathcal{G} \rightarrow \mathcal{U}$ — линейный оператор, определяющий модель интерпретации сигнала γ .

В задаче редукции измерения (6.20) следует определить правило $d(\cdot): \mathcal{X} \rightarrow \mathcal{U}$ оценивания $U\gamma$ так, чтобы нечеткий элемент $d(\xi)$ можно было считать в известном смысле оптимальной версией (оценкой) нечеткого элемента $U\gamma$.

Качество правила $d(\cdot)$ охарактеризуем величиной *необходимости потерь, сопутствующих ошибочному оцениванию* см. § 3.5.7:

$$\begin{aligned} \text{NL}(d(\cdot)) = & \vartheta \left(\sup_{x \in \mathcal{X}, g \in \mathcal{G}} \min(f^\nu(x - Ag), f^\gamma(g), \vartheta^{-1} \circ l(Ug, d(x))) \right) = \\ & \inf_{x \in \mathcal{X}, g \in \mathcal{G}} \max(\vartheta \circ f^\nu(x - Ag), \vartheta \circ f^\gamma(g), l(Ug, d(x))). \end{aligned} \quad (6.21)$$

Заметим, что если $f^\gamma(g) = 1$, $g \in \mathcal{G}$, т. е. если все значения сигнала γ в (6.20) априори равновозможны, то в (6.21) $\text{NL}(d(\cdot)) = \vartheta \left(\sup_{x \in \mathcal{X}, g \in \mathcal{G}} \min(f^\nu(x - Ag), \vartheta^{-1} \circ l(Ug, d(x))) \right)$ и совпадает с $\text{NL}(d(\cdot))$

(6.21) для априори произвольного сигнала $g \in \mathcal{G}$ в (6.13).

Заметим также, что теперь, в отличие от § 3.6.2, $f^\nu(x - Ag) = f^{\xi|\gamma}(x|g)$, $x \in \mathcal{X}$, $g \in \mathcal{G}$, — распределение переходной возможности, и

$$f^{\xi, \gamma}(x, g) = \min(f^{\xi|\gamma}(x|g), f^\gamma(g)), \quad x \in \mathcal{X}, g \in \mathcal{G}, \quad (6.22)$$

— совместное распределение входного сигнала γ и измерения ξ , а в (6.21) $\text{NL}(d(\cdot)) = \inf_{x \in \mathcal{X}, g \in \mathcal{G}} \max(\vartheta \circ f^{\xi, \gamma}(x, g), l(Ug, d(x)))$, ибо согласно

$$(6.22) \quad \vartheta \circ f^{\xi, \gamma}(x, g) = \max(\vartheta \circ f^{\xi|\gamma}(x|g), \vartheta \circ f^\gamma(g)).$$

Оптимальное правило $d_*(\cdot)$ определим условием

$$\text{NL}(d_*(\cdot)) = \min_{d(\cdot): \mathcal{X} \rightarrow \mathcal{U}} \text{NL}(d(\cdot)), \quad (6.23)$$

¹⁾ Вероятностными аналогами этой модели являются $[A, \tilde{\Sigma}_\gamma, \Sigma_\nu]$ и $[A, \gamma_0, \Sigma_\gamma, \Sigma_\nu]$, см. § 6.5 гл. 2.

согласно которому необходимость потерь, сопутствующих ошибочному оцениванию $U\gamma$ посредством $d_*(\xi)$, минимальна.

Как известно, см. § 3.5.7, для решения задачи (6.23) достаточно при каждом $x \in \mathcal{X}$ решить более простую задачу

$$N(d, x) \stackrel{\text{def}}{=} \inf_{g \in \mathcal{G}} \max(\vartheta \circ f^{\xi, \gamma}(x, g), l(Ug, d)) \sim \min_{d \in \mathcal{U}}, \quad (*)$$

так как семейство решений $d_* = d_*(x)$, $x \in \mathcal{X}$, задачи (*) является решением задачи (6.23). Речь идет о семействе задач на минимум

$$\max(\vartheta \circ f^{\xi, \gamma}(x, g), l(Ug, d)) \sim \min_{g \in \mathcal{G}, d \in \mathcal{U}}, \quad x \in \mathcal{X}. \quad (6.24)$$

Далее будем считать, что $l(\cdot, \cdot)$ в (6.24) удовлетворяет условию (6.16). Это означает, что только при $d = Ug$ потери невозможны, т. е. речь идёт о *необходимости NE($d(\cdot)$) ошибки редукции*. А так как $\min_{d \in \mathcal{U}} \max(\vartheta \circ f^{\xi, \gamma}(x, g), l(Ug, d)) = \max(\vartheta \circ f^{\xi, \gamma}(x, g), \min_{d \in \mathcal{U}} l(Ug, d)) = \vartheta \circ f^{\xi, \gamma}(x, g)$, $x \in \mathcal{X}$, $g \in \mathcal{G}$, где минимум по d достигается при (единственном) $d = Ug$, то задача (6.24) эквивалентна задаче

$$\vartheta \circ f^{\xi, \gamma}(x, g) \sim \min_{g \in \mathcal{G}}, \quad x \in \mathcal{X}, \quad (6.25)$$

ибо $d_* = d_*(x) = Ug(x)$ в точке $g = g(x)$ минимума в (6.25).

Пусть $\vartheta \circ f^{\nu}(x - Ag) = \delta_1(\|\Sigma^{-1/2}(x - Ag)\|^2)$, $\vartheta \circ f^{\gamma}(g) = \delta_2(\|G^{-1/2}(g - g_0)\|^2)$, $g \in \mathcal{G}$, $x \in \mathcal{X}$, где $\delta_1(\cdot)$, $\delta_2(\cdot)$ — строго монотонно возрастающие, непрерывно дифференцируемые на $[0, \infty)$ функции, принимающие значения в $[0, 1]$, $\delta_i(0) = 0$, $i = 1, 2$, $\Sigma: \mathcal{X} \rightarrow \mathcal{X}$ и $G: \mathcal{G} \rightarrow \mathcal{G}$ — положительно определенные операторы, играющие роль, аналогичную роли ковариационных операторов шума и входного сигнала в модели вероятностного аналога схемы измерений (6.20), см. § 6 гл. 2. Рассмотрим задачу (6.25) в этом случае:

$$\max(\delta_1(\|\Sigma^{-1/2}(x - Ag)\|^2), \delta_2(\|G^{-1/2}(g - g_0)\|^2)) \sim \min_{g \in \mathcal{G}}, \quad x \in \mathcal{X}. \quad (6.26)$$

Введем обозначения: $z = G^{-1/2}(g - g_0)$, $B = \Sigma^{-1/2}AG^{1/2}: \mathcal{G} \rightarrow \mathcal{X}$, $y = \Sigma^{-1/2}x - \Sigma^{-1/2}Ag_0$, в которых задача (6.26) имеет вид

$$\max(\delta_1(\|y - Bz\|^2), \delta_2(\|z\|^2)) \sim \min_{z \in \mathcal{G}}. \quad (6.27)$$

Для ее решения вычислим градиенты $\delta_1(\|y - Bz\|^2)$ и $\delta_2(\|z\|^2)$ по $z \in \mathcal{G}$: $\nabla \delta_1(\|y - Bz\|^2) = 2\delta'_1(\|y - Bz\|^2)B^*(Bz - y)$, $\nabla \delta_2(\|z\|^2) = 2\delta'_2(\|z\|^2)z$. Здесь $\delta'_1(\|y - Bz\|^2) = d\delta_1(r)/dr|_{r=\|y - Bz\|^2}$, $\delta'_2(\|z\|^2) = d\delta_2(r)/dr|_{r=\|z\|^2}$. В точке $z = z^*$ минимума (6.27) должны быть выполнены следующие условия [4]:

► либо для некоторого $\alpha \in [0, 1]$

$$\alpha \nabla \delta_1(\|y - Bz^*\|^2) + (1 - \alpha) \nabla \delta_2(\|z^*\|^2) = 0, \quad (6.28)$$

$$\delta_1(\|y - Bz^*\|^2) = \delta_2(\|z^*\|^2);$$

► либо

$$\nabla \delta_1(\|y - Bz^*\|^2) = 0, \delta_1(\|y - Bz^*\|^2) > \delta_2(\|z^*\|^2); \quad (6.29)$$

► либо, наконец,

$$\nabla \delta_2(\|z^*\|^2) = 0, \delta_1(\|y - Bz^*\|^2) < \delta_2(\|z^*\|^2). \quad (6.30)$$

Рассмотрим первый случай. Согласно первому условию (6.28) $\alpha \delta'_1 B^*(Bz^* - y) + (1 - \alpha) \delta'_2 z^* = 0$, откуда следует, что при $\alpha \in [0, 1]$ $z^* = z(\beta) = (B^*B + \beta I)^{-1} B^*y$, где $\beta = (1 - \alpha) \delta'_2 / (\alpha \delta'_1)$ (и соответственно $\alpha \in (0, 1)$) определяется из второго условия (6.28). Пусть для простоты далее $\delta_1(\cdot) = \delta_2(\cdot) = \delta(\cdot)$, где $\delta(\cdot): [0, \infty) \rightarrow [0, 1]$ — строго монотонно возрастающая непрерывно дифференцируемая функция¹⁾. В таком случае второе условие (6.28) эквивалентно равенству

$$\|y - Bz^*\|^2 = \|z^*\|^2. \quad (6.31)$$

Так как $\|y - Bz^*\|^2 = \beta^2 \|(BB^* + \beta I)^{-1} y\|^2 \triangleq r(\beta)$ — монотонно возрастает по $\beta \in (0, \infty)$, причем $\lim_{\beta \rightarrow +0} r(\beta) = \|(I - B(B^*B)^{-1} B^*)y\|^2$,

$\lim_{\beta \rightarrow \infty} r(\beta) = \|y\|^2$, а $\|z^*\|^2 = \|B^*(BB^* + \beta I)^{-1} y\|^2 \stackrel{\text{def}}{=} s(\beta)$ — монотонно убывает по $\beta \in (0, \infty)$, причем $\lim_{\beta \rightarrow +0} s(\beta) = \|(B^*B)^{-1} B^*y\|^2$ и

$\lim_{\beta \rightarrow \infty} s(\beta) = 0$, то условие (6.31) выполняется для некоторого $\beta \in (0, \infty)$, лишь если $\|(I - B(B^*B)^{-1} B^*)y\| < \|(B^*B)^{-1} B^*y\|$. В этом случае существует единственный корень $\beta = \beta(y) \in (0, \infty)$ уравнения (6.31) и

$$z^* = z(\beta(y)) = (B^*B + \beta(y)I)^{-1} B^*y \quad (6.32)$$

определяет стационарную точку (6.27). А так как $\max(\delta(\|y - Bz\|^2), \delta(\|z\|^2))$ — выпуклая функция $z \in \mathcal{X}$, то z^* (6.32) — искомая точка ее минимума.

Если $\|(I - B(B^*B)^{-1} B^*)y\| = \|(B^*B)^{-1} B^*y\|$, то $\beta = \beta(y) = 0$ ($\alpha = 1$) и $z^* = (B^*B)^{-1} B^*y$.

Рассмотрим второй случай. Решение z^* уравнения $\nabla \delta(\|y - Bz\|^2) = 0$ имеет вид [15]

$$z^* = (B^*B)^{-1} B^*y. \quad (6.33)$$

Так как $\|(y - Bz^*)\|^2 = \|(I - B(B^*B)^{-1} B^*)y\|^2$ и $\|z^*\|^2 = \|(B^*B)^{-1} B^*y\|^2$, то z^* (6.33) удовлетворяет условиям (6.29), если и только если

$$0 < \|(I - B(B^*B)^{-1} B^*)y\|^2 - \|(B^*B)^{-1} B^*y\|^2.$$

¹⁾ В общем случае $\delta_1(\cdot) \neq \delta_2(\cdot)$ анализ условий (6.28) – (6.30) принципиально не изменится.

Третий случай, очевидно, невозможен.

Подведем итоги.

Теорема 3.6.2. Пусть в схеме измерения (6.20) $A: \mathcal{G} \rightarrow \mathcal{X}$ — заданный линейный оператор, $\gamma \in \mathcal{G}$ и $\nu \in \mathcal{X}$ — независимые нечеткие элементы, $f^\gamma(\cdot): \mathcal{G} \rightarrow [0, 1]$ и $f^\nu(\cdot): \mathcal{X} \rightarrow [0, 1]$ их распределения, причем $\vartheta \circ f^\gamma(g) = \delta(\|G^{-1/2}(g - g_0)\|^2)$, $g \in \mathcal{G}$, $\vartheta \circ f^\nu(x - Ag) = \delta(\|\Sigma^{-1/2}(x - Ag)\|^2)$, $x \in \mathcal{X}$, где $\delta(\cdot): [0, \infty) \rightarrow [0, 1]$ — строго монотонно возрастающая, непрерывно дифференцируемая на $[0, \infty)$ функция, $\delta(0) = 0$; $\Sigma: \mathcal{X} \rightarrow \mathcal{X}$ и $G: \mathcal{G} \rightarrow \mathcal{G}$ — положительно определенные операторы и выполнено условие (6.16). Обозначим $\Delta(x) = \|(I - \Sigma^{-1/2}A(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2})\Sigma^{-1/2}(x - Ag_0)\|^2 - \|(G^{1/2}A^*\Sigma^{-1}AG^{1/2})^{-1}G^{1/2}A^*\Sigma^{-1/2}\Sigma^{-1/2}(x - Ag_0)\|^2$, $x \in \mathcal{X}$, тогда, если при $\xi = x$

► $\Delta(x) < 0$, то $d_*(x) = U\hat{\gamma}$, где $\hat{\gamma} = g_0 + GA^*(AGA^* + \omega(x)\Sigma)^{-1}(x - Ag_0)$, $\omega = \omega(x)$ — корень уравнения

$$\omega\|(BB^* + \omega I)^{-1}y\| = \|B^*(BB^* + \omega I)^{-1}y\|,$$

в котором $B = \Sigma^{-1/2}AG^{1/2}$, $y = \Sigma^{-1/2}(x - Ag_0)$. В этом случае в задаче (*)

$$\text{NE}(d_*(x), x) =$$

$$= \delta(\|G^{-1/2}(A^*\Sigma^{-1}A + \omega(x)G^{-1})^{-1}A^*\Sigma^{-1}(x - Ag_0)\|^2);$$

► $\Delta(x) = 0$, то $d_*(x) = U\hat{\gamma}$, где $\hat{\gamma} = g_0 + G^{1/2}(G^{1/2}A^*\Sigma^{-1}AG^{1/2})^{-1}G^{1/2}A^*\Sigma^{-1/2}\Sigma^{-1/2}(\xi - Ag_0)$. В этом случае в задаче (*) $\text{NE}(d_*(x), x) = \delta(\|(G^{1/2}A^*\Sigma^{-1}AG^{1/2})^{-1}G^{1/2}A^*\Sigma^{-1/2}\Sigma^{-1/2}(x - Ag_0)\|^2)$.

► $\Delta(x) > 0$, то $d_*(x) = U\hat{\gamma}$, где

$$\hat{\gamma} = g_0 + G^{1/2}(G^{1/2}A^*\Sigma^{-1}AG^{1/2})^{-1}G^{1/2}A^*\Sigma^{-1/2}\Sigma^{-1/2}(x - Ag_0),$$

В этом случае в задаче (*)

$$\text{NE}(d_*(x), x) = \delta(\|(I - \Sigma^{-1/2}A(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2})\Sigma^{-1/2}(x - Ag_0)\|^2).$$

Во всех случаях неизбежность ошибки редукции (6.21)

$$\text{NE}(d_*(\cdot)) = \inf_{x \in \mathcal{X}} \text{NE}(d_*(x), x) = 0. \quad \blacksquare$$

На рис. 3.10 приведены результаты редукции измерения, полученные для описанной в теореме 3.6.2 возможностной модели $[A, f^\gamma(\cdot), f^\nu(\cdot)]$ схемы измерения (6.20), и для вероятностной модели $[A, G, g_0, \Sigma]$ схемы (6.13) [13].

Чтобы пояснить результаты вычислительного эксперимента, напомним, что наилучшей в среднем квадратичном (с. к.) оценкой случайного элемента γ в классе всех функций случайного элемента ξ (6.13), имеющих первые два момента, является условное математическое ожидание $\tilde{\gamma} = M(\gamma|\xi)$, а в классе всех аффинных функций $R\xi + r$ — статистика

$$\hat{\gamma} = \gamma_0 + \Sigma_\gamma A^*(A\Sigma_\gamma A^* + \Sigma_\nu)^{-1}(\xi - A\gamma_0), \quad (6.34)$$

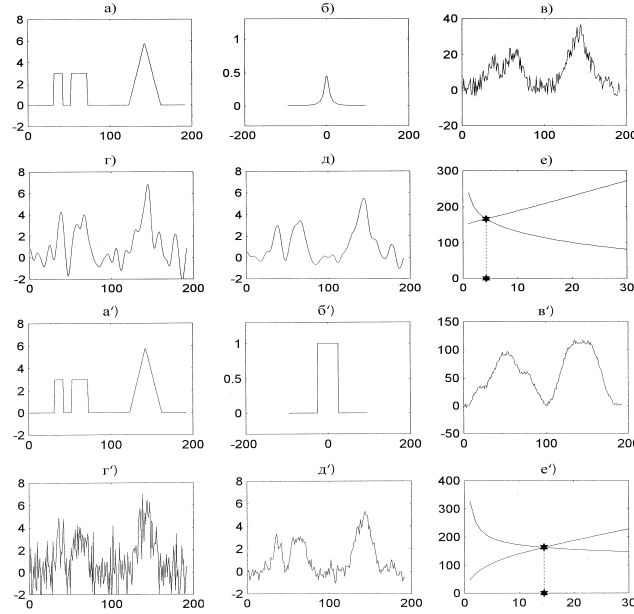


Рис. 3.10. а, а') Измеренные в (6.13) и (6.20) сигналы g и γ , соответственно. б, б') Аппаратные функции $a(i, j)$, $i, j = 1, \dots, 192$, измерительного прибора A как функции разности $i - j$ при $j = 96$; $(A\gamma)(i) = \sum_{j=1}^{192} a(i, j)\gamma(j)$, $i = 1, \dots, 192$; в, в') Результаты измерения в (6.13) (6.20) $\xi(i) = (A\gamma)(i) + \nu(i) \equiv (A\gamma)(i) + \nu(i)$, $i = 1, \dots, 192$, при аппаратных функциях б, б'), соответственно; г) вероятностная редукция измерения (2.1) $\hat{\gamma} = \gamma_0 + \Sigma_\gamma A^* (A \Sigma_\gamma A^* + \Sigma_\nu)^{-1} (\xi - A\gamma_0)$, $\Sigma_\nu = \sigma^2 I$, $\Sigma_\gamma = \varphi^2 I$, минимизирующая средне-квадратичную ошибку оценивания $M\|\hat{\gamma} - \gamma\|^2 = \min_{R, r} M\|R\xi + r - \gamma\|^2$, $\sigma^2 = 2$, $\varphi^2 = 2.62$; γ и ν независимы и гауссовы; д) возможностная редукция измерения (6.20) $\hat{\gamma} = g_0 + GA^* (AGA^* + \omega(\xi)\Sigma)^{-1} (\xi - Ag_0)$, минимизирующая неизбежность ошибки оценивания, $\Sigma = \sigma^2 I$, $G = \varphi^2 I$, $\sigma^2 = 2$, $\varphi^2 = 2.62$; г'), д') — полученные при аппаратной функции б') аналоги результатов редукции г), д), полученных при аппаратной функции б), для независимых γ и ν , равномерно распределенных на $[-\sqrt{3}\sigma, \sqrt{3}\sigma]^{192}$ и соответственно на $[-\sqrt{3}\varphi, \sqrt{3}\varphi]^{192}$ при $\sigma^2 = 2$, $\varphi^2 = 2.62$; е, е') левые и правые части равенства (6.31) для аппаратных функций б, б') как функции $\sigma^2\beta$; на горизонтальной оси отмечены соответствующие значения $\sigma^2\omega(\xi)$.

если в (6.13) γ и ν независимы, $\gamma_0 = M\gamma$, $M\nu = 0$, Σ_γ и Σ_ν — ковариационные операторы γ и ν соответственно [15]. При этом, вообще говоря,

$$M\|\tilde{\gamma} - \gamma\|^2 \leq M\|\hat{\gamma} - \gamma\|^2 = \text{tr}(\Sigma_\gamma - \Sigma_\gamma A^* (A \Sigma_\gamma A^* + \Sigma_\nu)^{-1} A \Sigma_\gamma). \quad (6.35)$$

Если случайные элементы γ и ξ (совместно) нормально распределены, оценки $\tilde{\gamma}$ и $\hat{\gamma}$ совпадают и, следовательно, максимум с.к. погрешности

$M\|\tilde{\gamma} - \gamma\|^2$ по всем распределениям независимых случайных элементов γ и ν , имеющих фиксированные ковариационные операторы Σ_γ и Σ_ν , достигается при нормальном совместном распределении γ и ξ и равен правой части (6.35). Иначе говоря, при фиксированных Σ_γ и Σ_ν , случай нормального распределения выделяется тем, что при нем с.к. ошибка оценивания максимальна, равна правой части (6.35), а наилучшая оценка $M(\gamma|\xi)$ — аффинная функция ξ (6.34).

Отсюда следует, что при нормально распределенном шуме ν оценка, график которой приведен на рис. 3.10, g , неуплучшаема, соответствующая возможностная оценка 3.10, d имеет несколько большую погрешность. С другой стороны, если шум ν имеет равномерное распределение с такими же моментами, то возможностная оценка 3.10, d' имеет существенно меньшую погрешность, чем линейная оценка 3.10, e' (6.34).

Характерно, что *редукция при возможностной модели измерения не зависит от явного вида функции $\delta(\cdot)$, определяющей распределения*, т.е. инвариантна относительно выбора шкалы значений возможности, общей для нечетких элементов γ и ν , а оценка $\hat{\gamma} = \gamma_0 + GA^*(GA^* + \omega(x)\Sigma)^{-1}(x - A\gamma_0)$ в теореме 3.6.2 аналогична оценке Джеймса – Стейна, см., например, [15].

3.7. Приложение

В рассмотренных в § 3.3 математических методах и алгоритмах эмпирического восстановления стохастически измеримой возможности предполагалось, что вероятности элементарных событий априори упорядочены по невозрастанию либо как в (3.1), либо как в (3.8). В этом заключительном параграфе будут рассмотрены алгоритмы эмпирического упорядочения и интервального оценивания вероятностей.

3.7.0. Введение. В задаче эмпирического упорядочения вероятностей элементарных событий моделью последовательности n взаимно независимых испытаний является вероятностное пространство $(\Omega^n, \mathcal{P}(\Omega^n), \text{Pr}_{1,\dots,n}^{(n)}) = (\Omega, \mathcal{P}(\Omega), \text{Pr}_1) \times \dots \times (\Omega, \mathcal{P}(\Omega), \text{Pr}_n)$, в котором $\Omega = \{\omega_1, \omega_2, \dots\}$, $\text{Pr}_{1,\dots,n}^{(n)} = \text{Pr}_1 \times \dots \times \text{Pr}_n$ и $(\Omega, \mathcal{P}(\Omega), \text{Pr}_j)$ — модель j -го испытания. Результатом n испытаний является совокупность частот элементарных событий

$$\nu_i^{(n)} = \frac{1}{n} \sum_{j=1}^n \chi_i(\omega^{(j)}), \quad i = 1, 2, \dots, \quad (7.1)$$

где $\omega^{(j)}$ — исход j -го испытания, $\chi_i(\omega) = \begin{cases} 1, & \omega = \omega_i, \\ 0, & \omega \neq \omega_i, \end{cases}$ — индикаторная функция элементарного события $\{\omega_i\}$, и все слагаемые $\chi_i(\omega^{(i)})$, $i = 1, \dots, n$, взаимно независимы, $i = 1, 2, \dots$. Вероятности элементарных событий $\text{Pr}_j(\{\omega_i\}), i = 1, \dots, n$, попарно различны и одинаково упорядочены при всех испытаниях $j = 1, 2, \dots$, т.е.

для каждой пары $\omega_i, \omega_k, i \neq k$, либо $\text{Pr}_j(\{\omega_i\}) > \text{Pr}_j(\{\omega_k\})$, либо $\text{Pr}_j(\{\omega_i\}) < \text{Pr}_j(\{\omega_k\})$, для всех $j = 1, 2, \dots$. Значения $\text{Pr}_j(\{\omega_i\})$, $i = 1, 2, \dots, j = 1, 2, \dots$, разумеется, неизвестны, представляет интерес лишь их упорядоченность, определяемая знаками всех разностей $\text{Pr}_j(\{\omega_i\}) - \text{Pr}_j(\{\omega_k\})$, $i \neq k, i, k = 1, 2, \dots$, которая и подлежит определению на основе наблюдения значений частот (7.1).

Для построения алгоритмов эмпирического упорядочения вероятностей $\text{Pr}_j(\omega_1), \text{Pr}_j(\omega_2), \dots, j = 1, 2, \dots$, воспользуемся неравенством Хёффдинга (3.26).

Если в неравенстве (3.26) $\xi_k = \chi_i(\omega^{(k)})$, где $\omega^{(k)}$ — исход k -го испытания, модель которого $(\Omega, \mathcal{P}(\Omega), \text{Pr}_k)$, $k = 1, \dots, n$, то

$$\text{Pr}_{1, \dots, n}^{(n)}(\{\nu_i^{(n)} - M_{1, \dots, n}^{(n)} \nu_i^{(n)} > \varepsilon\}) \leq \exp(-2n\varepsilon^2), \quad (7.2)$$

$$\text{Pr}_{1, \dots, n}^{(n)}(\{\nu_i^{(n)} - M_{1, \dots, n}^{(n)} \nu_i^{(n)} < -\varepsilon\}) \leq \exp(-2n\varepsilon^2), \quad (7.3)$$

$$\text{Pr}_{1, \dots, n}^{(n)}(\{|\nu_i^{(n)} - M_{1, \dots, n}^{(n)} \nu_i^{(n)}| > \varepsilon\}) \leq 2 \exp(-2n\varepsilon^2), \quad i = 1, 2, \dots \quad (7.4)$$

Неравенства (7.2 – 7.4) позволяют получить следующие интервальные оценки для "эмпирических вероятностей"¹⁾ $\text{pr}_i^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n \text{Pr}_j(\{\omega_i\}) = M_{1, \dots, n}^{(n)} \nu_i^{(n)}$, $i = 1, 2, \dots$:

$$\text{Pr}_{1, \dots, n}^{(n)} \left(\left\{ \text{pr}_i^{(n)} \geq \nu_i^{(n)} - \left(\frac{1}{2n} \ln \frac{1}{\alpha} \right)^{1/2} \right\} \right) \geq 1 - \alpha, \quad i = 1, 2, \dots, \quad (7.5)$$

если в (7.2) $\exp(-2n\varepsilon^2) = \alpha$,

$$\text{Pr}_{1, \dots, n}^{(n)} \left(\left\{ \text{pr}_i^{(n)} \leq \nu_i^{(n)} + \left(\frac{1}{2n} \ln \frac{1}{\alpha} \right)^{1/2} \right\} \right) \geq 1 - \alpha, \quad i = 1, 2, \dots, \quad (7.6)$$

если в (7.3) $\exp(-2n\varepsilon^2) = \alpha$, и

$$\begin{aligned} \text{Pr}_{1, \dots, n}^{(n)} \left(\left\{ \nu_i^{(n)} - \left(\frac{1}{2n} \ln \frac{2}{\alpha} \right)^{1/2} \leq \text{pr}_i^{(n)} \leq \nu_i^{(n)} + \left(\frac{1}{2n} \ln \frac{2}{\alpha} \right)^{1/2} \right\} \right) \geq \\ \geq 1 - \alpha, \quad i = 1, 2, \dots, \end{aligned} \quad (7.7)$$

если в (7.4) $2 \exp(-2n\varepsilon^2) = \alpha$, характеризующие ошибки приближения "эмпирической вероятности" $\text{pr}_i^{(n)}$ частотой $\nu_i^{(n)}$, $i = 1, 2, \dots, n = 1, 2, \dots$. Заметим, что согласно (7.4) при $n \rightarrow \infty$, $\nu_i^{(n)} - \text{pr}_i^{(n)} \xrightarrow{\text{П.Н.}} 0$, ибо при любом $\varepsilon > 0 \sum_{n=1}^{\infty} \exp(-2n\varepsilon^2) < \infty$, и, следовательно, при любом

¹⁾ При каждом n , $\text{pr}_i^{(n)} \geq 0$, $i = 1, 2, \dots$, и $\sum_{i=1}^{\infty} \text{pr}_i^{(n)} = 1$. Так как при $n \rightarrow \infty$ $\nu_i^{(n)} - \text{pr}_i^{(n)} \xrightarrow{\text{П.Н.}} 0$, то "эмпирические вероятности" $\text{pr}_i^{(n)}$, $i = 1, 2, \dots$, могут быть восстановлены эмпирически

$\varepsilon > 0$ согласно лемме Бореля-Кантелли может произойти лишь п. н. конечное число событий $\{|\nu_i^{(n)} - \text{pr}_i^{(n)}| > \varepsilon\}$, $n = 1, 2, \dots$

3.7.1. Алгоритм эмпирического упорядочения вероятностей элементарных событий, не изменяющихся в процессе испытаний. Моделью последовательности n взаимно независимых испытаний в рассматриваемом случае является вероятностное пространство $(\Omega^n, \mathcal{P}(\Omega^n), \text{Pr}^{(n)}) = (\Omega, \mathcal{P}(\Omega), \text{Pr}) \times \dots \times (\Omega, \mathcal{P}(\Omega), \text{Pr})$, в котором $\text{Pr}^{(n)} = \underbrace{\text{Pr} \times \dots \times \text{Pr}}_{n \text{ раз}}$. Задача эмпирического упорядочения вероятностей $\text{pr}_i = \text{Pr}(\{\omega_i\})$, $i = 1, \dots, s$, элементарных событий состоит в том, чтобы на основе результатов наблюдений частот $\nu_1^{(n)}, \dots, \nu_s^{(n)}$, $n = 1, 2, \dots$, (7.1) элементарных событий определить с гарантированной вероятностью правильные отношения $\text{pr}_i > \text{pr}_k$ или $\text{pr}_i < \text{pr}_k$ для $s(s-1)/2$ пар $(i < k) \in \{1, \dots, s\} \times \{1, \dots, s\}$, $s = 2, 3, \dots$

Обозначим

$$\nu_{ik}^{(n)} \stackrel{\text{def}}{=} \nu_i^{(n)} - \nu_k^{(n)}, \quad \text{pr}_{ik} \stackrel{\text{def}}{=} \text{pr}_i - \text{pr}_k = \mathbf{E}^{(n)} \nu_{ik}^{(n)}, \quad i, k = 1, 2, \dots, i < k, \quad (7.8)$$

и сформулируем задачу эмпирического упорядочения вероятностей $\text{pr}_1, \dots, \text{pr}_s$ как задачу определения знаков pr_{ik} , $i < k$, на основе наблюдаемых значений $\nu_{ik}^{(n)}$, $i < k$, $i, k = 1, \dots, s$, $n = 1, 2, \dots$, $s = 2, 3, \dots$

Так как

$$\nu_{ik}^{(n)} = \frac{1}{n} \sum_{j=1}^n (\chi_i(\omega^{(j)}) - \chi_k(\omega^{(j)})), \quad (7.9)$$

где слагаемые для $j = 1, \dots, n$ взаимно независимы и принимают значения $-1, 0$ и 1 , то согласно (3.26) подобно оценкам (7.2 – 7.4) для любого $\delta > 0$ соответственно

$$\text{Pr}^{(n)}(\{\nu_{ik}^{(n)} - \text{pr}_{ik} > \delta\}) \leq \exp(-n\delta^2/2), \quad (7.10)$$

$$\text{Pr}^{(n)}(\{\nu_{ik}^{(n)} - \text{pr}_{ik} < -\delta\}) \leq \exp(-n\delta^2/2), \quad (7.11)$$

$$\text{Pr}^{(n)}(\{|\nu_{ik}^{(n)} - \text{pr}_{ik}| > \delta\}) \leq 2 \exp(-n\delta^2/2), \quad (7.12)$$

и подобно (7.5 – 7.7) соответственно

$$\text{Pr}^{(n)}\left(\left\{\text{pr}_{ik} \geq \nu_{ik}^{(n)} - \left(\frac{2}{n} \ln \frac{1}{\alpha}\right)^{1/2}\right\}\right) > 1 - \alpha, \quad (7.13)$$

$$\text{Pr}^{(n)}\left(\left\{\text{pr}_{ik} \leq \nu_{ik}^{(n)} + \left(\frac{2}{n} \ln \frac{1}{\alpha}\right)^{1/2}\right\}\right) > 1 - \alpha, \quad (7.14)$$

если в (7.10), (7.11) $\exp(-n\delta^2/2) = \alpha$, и

$$\text{Pr}^{(n)}\left(\left\{\nu_{ik}^{(n)} - \left(\frac{2}{n} \ln \frac{2}{\alpha}\right)^{1/2} \leq \text{pr}_{ik} \leq \nu_{ik}^{(n)} + \left(\frac{2}{n} \ln \frac{2}{\alpha}\right)^{1/2}\right\}\right) > 1 - \alpha, \quad (7.15)$$

если в (7.12) $2 \exp(-n\delta^2/2) = \alpha$.

Условия (7.13 – 7.15) определяют случайные множества

$$\left\{ \text{pr}_{ik} \in [-1, 1], \text{pr}_{ik} \geq \nu_{ik}^{(n)} - \left(\frac{2}{n} \ln \frac{1}{\alpha} \right)^{1/2} \right\}, \quad (7.16)$$

$$\left\{ \text{pr}_{ik} \in [-1, 1], \text{pr}_{ik} \leq \nu_{ik}^{(n)} + \left(\frac{2}{n} \ln \frac{1}{\alpha} \right)^{1/2} \right\}, \quad (7.17)$$

$$\left\{ \text{pr}_{ik} \in [-1, 1], \nu_{ik}^{(n)} - \left(\frac{2}{n} \ln \frac{2}{\alpha} \right)^{1/2} \leq \text{pr}_{ik} \leq \nu_{ik}^{(n)} + \left(\frac{2}{n} \ln \frac{2}{\alpha} \right)^{1/2} \right\}, \quad (7.18)$$

покрывающие истинное значение pr_{ik} с вероятностью, большей $1 - \alpha$, $\alpha \in (0, 1)$, $i < k$, $i, k = 1, 2, \dots$, $n = 1, 2, \dots$.

Поскольку согласно оценкам (7.10 – 7.15) для определения знака pr_{ik} требуется тем больше испытаний, чем меньше $|\text{pr}_{ik}|$, алгоритм эмпирического упорядочения $\text{pr}_1, \dots, \text{pr}_s$ должен "распознавать малость" $|\text{pr}_{ik}|$ и "принимать решение" продолжить испытания, если значение $|\nu_{ik}^{(n)}|$ свидетельствует о малости $|\text{pr}_{ik}|$.

Рассмотрим следующий алгоритм статистических решений об упорядоченности вероятностей $\text{pr}_1, \dots, \text{pr}_s$:

для всех пар $(i < k) \in \{1, \dots, s\} \times \{1, \dots, s\}$ и каждого $n = 1, 2, \dots$

- ▶ если $\nu_{ik}^{(n)} > \delta^{(n,s)}$, то $\square_1$: считать $\text{pr}_{ik} > 0$;
- ▶ если $\nu_{ik}^{(n)} < -\delta^{(n,s)}$, то $\square_2$: считать $\text{pr}_{ik} < 0$;
- ▶ если $|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}$, то $\circ$: продолжить испытания.

(7.19)

В алгоритме (7.19) при каждом $n = 1, 2, \dots$ проверяются условия решений $\square_1, \square_2$ и $\circ$ для всех $s(s-1)/2$ пар $i < k$, $i, k = 1, \dots, s$. Если для всех пар при некотором n приняты только решения $\square_1$ или $\square_2$, то алгоритм (7.19) завершен, если же хотя бы для одной пары принято решение $\circ$, то при каждом следующем испытании проверяются условия решений $\square_1, \square_2$ и $\circ$, и ранее принятые решения корректируются для всех $s(s-1)/2$ пар $i < k$, $i, k = 1, \dots, s$. Подчеркнем, что решения $\square_1$ и $\square_2$ относятся к каждой отдельной паре $i < k$, а решение $\circ$ относится ко всем $s(s-1)/2$ парам.

Задача эмпирического упорядочения вероятностей $\text{pr}_1, \dots, \text{pr}_s$ будет решена, если на основе наблюдаемых значений $\nu_{ik}^{(n)}$, $(i < k) \in \{1, \dots, s\} \times \{1, \dots, s\}$ будут приняты все $s(s-1)/2$ решений $\square_1$ или $\square_2$ и оценены вероятности сопутствующих ошибок, поскольку это позволит определить упорядоченность $\text{pr}_{i_1} > \dots > \text{pr}_{i_s}$ вероятностей $\text{pr}_1, \dots, \text{pr}_s$, с гарантированной вероятностью совпадающую с истинной их упорядоченностью.

Рассмотрим подробнее алгоритм (7.19) для фиксированной пары $i < k$. В (7.19), $\square_{1,2}$ — символы прекращения испытаний и принятия одного из двух решений $\text{pr}_{ik} > 0$ или $\text{pr}_{ik} < 0$, $\circ$ — символ продолжения испытаний как реакции на условие $|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}$, свидетельствующее

о малости $|\text{pr}_{ik}|$ и недостаточности количества n испытаний для определения знака pr_{ik} ; $\delta^{(n,s)} > 0$, $n = 1, 2, \dots$

Таким образом в алгоритме (7.19) число испытаний n увеличивается до тех пор, пока для всех $s(s-1)/2$ пар $i < k$ не завершены циклы $\circ$ и не закончена обусловленная ими ревизия решений, принятых ранее для всех $s(s-1)/2$ пар $i < k$, $i, k = 1, \dots, s$.

Если n_s — число испытаний, при котором первый раз завершены все $s(s-1)/2$ циклов $\circ$, то на n_s -ом испытании алгоритм (7.19) завершен.

Для определения в (7.19) значений $\delta^{(n,s)}$, $n = 1, 2, \dots$, зададим оценку сверху $\alpha^{(s)}$ вероятности ошибочных решений $\square_1$ и $\square_2$.

Пусть наблюдаемое значение $\nu_{ik}^{(n)} > \delta^{(n,s)}$, в то время как на самом деле $\text{pr}_{ik} < 0$. В таком случае решение $\square_1$: "считать $\text{pr}_{ik} > 0$ " ошибочно, его вероятность ¹⁾

$$\Pr^{(n)}(\{\nu_{ik}^{(n)} > \delta^{(n,s)}\} | \text{pr}_{ik} < 0) \leq \Pr^{(n)}(\{\nu_{ik}^{(n)} - \text{pr}_{ik} > \delta^{(n,s)}\} | \text{pr}_{ik} < 0) \leq \exp(-n(\delta^{(n,s)})^2/2) \stackrel{\text{def}}{=} \alpha^{(s)}. \quad (7.20)$$

Первое неравенство в (7.20) является следствием того, что событие $\{\nu_{ik}^{(n)} > \delta^{(n,s)}\}$, в силу условия $\text{pr}_{ik} < 0$, характеризующего вероятность $\Pr^{(n)}$, влечет событие $\{\nu_{ik}^{(n)} - \text{pr}_{ik} > \delta^{(n,s)}\}$, второе неравенство есть следствие (7.10) неравенства Хефдинга, (3.26). Значение $\alpha^{(s)}$ оценивает сверху вероятность ошибочного решения $\square_1$ и определяет значение

$$\delta^{(n,s)} = \left(\frac{2}{n} \ln \frac{1}{\alpha^{(s)}} \right)^{1/2}, \quad n = 1, 2, \dots, \quad s = 2, 3, \dots \quad (7.21)$$

Вероятность ошибочного решения $\square_2$ оценивается сверху тем же значением $\alpha^{(s)}$, ибо

$$\Pr^{(n)}(\{\nu_{ik}^{(n)} < -\delta^{(n,s)}\} | \text{pr}_{ik} > 0) \leq \Pr^{(n)}(\{\nu_{ik}^{(n)} - \text{pr}_{ik} < -\delta^{(n,s)}\} | \text{pr}_{ik} > 0) \leq \exp(-n(\delta^{(n,s)})^2/2) = \alpha^{(s)}. \quad (7.22)$$

Покажем, что событие $\{-\delta^{(n,s)} \leq \nu_{ik}^{(n)} \leq \delta^{(n,s)}\}$, как в случае $\text{pr}_{ik} > 0$, так и в случае $\text{pr}_{ik} < 0$, для каждого $s = 2, 3, \dots$ может выполняться лишь для п.н. конечного числа испытаний $n = 1, 2, \dots$

Пусть $|\text{pr}_{ik}| \geq 2\varepsilon_{ik} > 0$. Тогда при достаточно больших n

$$\begin{aligned} \Pr^{(n)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}| \geq 2\varepsilon_{ik}) &\leq \\ &\leq \Pr^{(n)}(\{|\text{pr}_{ik}| - |\nu_{ik}^{(n)}| \geq 2\varepsilon_{ik} - \delta^{(n,s)}\} | |\text{pr}_{ik}| \geq 2\varepsilon_{ik}) \leq \\ &\leq \Pr^{(n)}(\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq 2\varepsilon_{ik} - \delta^{(n,s)}\} | |\text{pr}_{ik}| \geq 2\varepsilon_{ik}) \leq \end{aligned}$$

¹⁾ Условие $|\text{pr}_{ik} < 0$ в $\Pr^{(n)}(\cdot | \text{pr}_{ik} < 0)$ записано как характеристика вероятности $\Pr^{(n)}$. Если после вертикальной черты записано условие в фигурных скобках, то, например, $\Pr^{(n)}(\cdot | \{A\})$ — условная при условии $\{A\}$ вероятность.

$$\leq \Pr^{(n)}(\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq \varepsilon_{ik}\} \leq 2 \exp(-n\varepsilon_{ik}^2/2). \quad (7.23)$$

Дело в том, что, если $|\text{pr}_{ik}| \geq 2\varepsilon_{ik}$, то событие $\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\}$ влечет событие $\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq 2\varepsilon_{ik} - \delta^{(n,s)}\}$, которое в свою очередь влечет событие $\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq 2\varepsilon_{ik} - \delta^{(n,s)}\}$, откуда следуют первое и второе неравенства в (7.23). Далее, так как согласно (7.21) для любого $s = 2, 3, \dots$ при $n \rightarrow \infty$ $\delta^{(n,s)} \rightarrow 0$, то для всех $n > n(s)$ $\varepsilon_{ik} - \delta^{(n,s)} > 0$, и для таких n событие $\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq 2\varepsilon_{ik} - \delta^{(n,s)}\}$ влечет событие $\{|\text{pr}_{ik} - \nu_{ik}^{(n)}| \geq \varepsilon_{ik}\}$, откуда следует предпоследнее неравенство в (7.23). Последнее неравенство в (7.23) является вариантом неравенства (7.12). Поскольку согласно неравенствам в (7.23)

$$\begin{aligned} \sum_{n=1}^{\infty} \Pr^{(n)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}| \geq 2\varepsilon_{ik} > 0) = \\ = \sum_{n=1}^{\infty} \Pr^{(\infty)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}| \geq 2\varepsilon_{ik}) \leq 2 \sum_{n=1}^{\infty} \exp(-n\varepsilon_{ik}^2/2) < \infty, \end{aligned}$$

то согласно лемме Бореля-Кантелли для любого $s = 2, 3, \dots$ может произойти лишь п.н. конечное число событий $\{-\delta^{(n,s)} \leq \nu_{ik}^{(n)} \leq \delta^{(n,s)}\}$, $n = 1, 2, \dots$, после которых будет принято одно из решений $\square_1$ или $\square_2$.

Чтобы оценить фактическую вероятность хотя бы одной ошибки в $s(s-1)/2$ решениях $\square_1$ и $\square_2$, принятых по завершении алгоритма (7.19), представим себе, что алгоритм исчерпан на $n = n_s$ -ом испытании,

$$\nu_{i_1}^{(n)} > \nu_{i_2}^{(n)} > \dots > \nu_{i_s}^{(n)}, \quad n = n_s, \quad (7.24)$$

— полученная в результате n_s испытаний упорядоченность по убыванию частот i_1 -го, i_2 -го, $\dots$, i_s -го элементарных событий¹⁾ и

$$\nu_{i_1, i_2}^{(n)} > \delta^{(n,s)}, \dots, \nu_{i_{s-1}, i_s}^{(n)} > \delta^{(n,s)}, \quad n = n_s, \quad (7.25)$$

— $(s-1)$ условие, обеспечившее принятие $(s-1)$ решений $\square_1$ или $\square_2$, согласно которым $\text{pr}_{i_1, i_2} > 0, \dots, \text{pr}_{i_{s-1}, i_s} > 0$; при этом вероятность хотя бы одной ошибки среди этих $(s-1)$ решений не превосходит $(s-1)\alpha^{(s)}$.

Последнее замечание требует разъяснений, поскольку $(i_1, \dots, i_s)$ — реализация случайной последовательности $(\iota_1, \dots, \iota_s)$, упорядочивающей частоты $\nu_{\iota_1}^{(n)} \geq \nu_{\iota_2}^{(n)} \geq \dots \geq \nu_{\iota_s}^{(n)}$ для каждого $n = 1, 2, \dots$, а n_s — реализация случайного числа испытаний, при котором первый раз приняты все $s(s-1)/2$ решений $\square_1$ и $\square_2$. Тем не менее

¹⁾ Строгие неравенства в (7.24) следуют из условий (7.25) принятия решений $\square_1$ или $\square_2$, согласно которым $\nu_{i_t}^{(n)} - \nu_{i_{t+1}}^{(n)} = \nu_{i_t, i_{t+1}}^{(n)} > \delta^{(n,s)}$, $t = 1, \dots, s-1$.

²⁾ Дело в том, что, возможно, как $i_t < i_{t+1}$, так и $i_t > i_{t+1}$, $t = 1, \dots, s-1$.

$$\begin{aligned}
\Pr^{(n_s)}(\{\nu_{\iota_t, \iota_{t+1}}^{(n_s)} > \delta^{(n_s, s)}\} | \{\text{pr}_{\iota_t, \iota_{t+1}} < 0\}) &\leq \\
&\leq \sum_{(i_t, i_{t+1})} \Pr^{(n_s)}(\{\nu_{\iota_t, \iota_{t+1}}^{(n_s)} - \text{pr}_{\iota_t, \iota_{t+1}} > \delta^{(n_s, s)}\} | \{\text{pr}_{\iota_t, \iota_{t+1}} < 0\}) \leq \\
&\leq (-n(\delta^{(n_s, s)})^2/2) = \alpha^{(s)}, \quad (7.26)
\end{aligned}$$

поскольку неравенства (7.26) выполняются для любой реализации $\iota_t = i_t$, $t = 1, \dots, s$, удовлетворяющей условиям (7.24) и (7.25).

Что касается остальных $s(s-1)/2 - (s-1) = (s-1)(s-2)/2$ решений, то, поскольку в силу неравенств (7.24), (7.25) для $q > t$ выполнено неравенство $\nu_{i_t, i_q}^{(n_s)} = \nu_{i_t, i_{t+1}}^{(n_s)} + \nu_{i_{t+1}, i_{t+2}}^{(n_s)} + \dots + \nu_{i_{q-1}, i_q}^{(n_s)} > (q-t)\delta^{(n_s, s)}$, то вероятность ошибочности фактического решения в этом случае

$$\begin{aligned}
\Pr^{(n_s)}(\{\nu_{i_t, i_q}^{(n_s)} > (q-t)\delta^{(n_s, s)}\} | \{\text{pr}_{i_t, i_q} < 0\}) &\leq \\
&\leq \exp(-n(q-t)^2(\delta^{(n_s, s)})^2/2) = (\alpha^{(s)})^{(q-t)^2}.
\end{aligned}$$

Поэтому вероятность хотя бы одной ошибки во всех $s(s-1)/2$ решениях не больше, чем ¹⁾ вероятность хотя бы одной ошибки в фактически принятых: $(s-1)$ решениях $\nu_{i_t, i_{t+1}} > \delta^{(n_s, s)}$, $t = 1, \dots, s-1$, $(s-2)$ решениях $\nu_{i_t, i_{t+2}} > 2\delta^{(n_s, s)}$, $t = 1, \dots, s-2, \dots$, $(s-(s-1))$ решениях $\nu_{i_t, i_s} > (\delta^{(n_s, s)})^{s-1}$, т.е. не больше, чем $\frac{(s-1)\alpha}{1-\alpha} - \frac{\alpha^2(1-\alpha^{s-1})}{(1-\alpha)^2} |_{\alpha=\alpha^{(s)}} = (s-1)\alpha^{(s)} + (s-2)(\alpha^{(s)})^2 + \dots + 2(\alpha^{(s)})^{s-2} + (\alpha^{(s)})^{s-1} = (s-1)\alpha^{(s)} + o(\alpha^{(s)})$.

Теорема 3.7.1. Пусть в алгоритме (7.19) эмпирического упорядочения вероятностей элементарных событий $\text{pr}_1, \dots, \text{pr}_s$, $\alpha^{(s)} \in (0, 1)$ оценивает сверху вероятность ошибочного решения $\square_1$ или $\square_2$. Тогда все $s(s-1)/2$ решений будут приняты на основе п.н. конечного числа испытаний, и если n_s — число испытаний, при котором первый раз приняты все $s(s-1)/2$ решений, и $\nu_{i_1}^{(n_s)} > \nu_{i_2}^{(n_s)} > \dots > \nu_{i_s}^{(n_s)}$ — полученная упорядоченность частот, то искомая упорядоченность вероятностей элементарных событий есть $\text{pr}_{i_1} > \text{pr}_{i_2} > \dots > \text{pr}_{i_s}$, и вероятность того, что найденная упорядоченность совпадает с их истинной упорядоченностью, больше $\frac{1-s\alpha}{1-\alpha} + \frac{\alpha^2(1-\alpha^{s-1})}{(1-\alpha)^2} |_{\alpha=\alpha^{(s)}} = 1 - (s-1)\alpha^{(s)} + o(\alpha^{(s)})$.

¹⁾ $\sum_{j=1}^{s-1} (s-j)\alpha^j = -\alpha^{s+1} \frac{d}{d\alpha} \sum_{j=1}^{s-1} \alpha^{j-s} = (s-1) \frac{\alpha}{1-\alpha} - \frac{\alpha^2(1-\alpha^{s-1})}{(1-\alpha)^2}$

Замечание 3.7.1. Если число s упорядочиваемых в алгоритме (7.19) вероятностей элементарных событий априори любое, то оценки вероятностей ошибочных решений целесообразно выбрать согласно условию $\alpha^{(s)} = \alpha/(s-1)$, $s = 2, 3, \dots$, где α — априорная оценка вероятности ошибочного упорядочения вероятностей элементарных событий. При таком определении $\alpha^{(s)}$, $s = 2, 3, \dots$, вероятность ошибочного упорядочения вероятностей любого конечного числа элементарных событий оценивается сверху значением $\alpha + o(\alpha)$.

3.7.2. Алгоритм эмпирического интервального оценивания вероятностей элементарных событий. Алгоритм (7.19) эмпирического упорядочения вероятностей элементарных событий позволит оценить и их значения, если упорядоченность частот (7.24) получить как результат $(s-1)$ решений, обусловленных событием $Q_s \triangleq \bigcap_{k=1}^{s-1} \left\{ \nu_{i_k, i_{k+1}}^{(n_s)} > 2\hat{\delta}^{(n_s, s)} \right\}$, для которого согласно теореме 3.7.1 $\Pr^{(n_s)}(Q_s) > 1 - (s-1)\hat{\alpha}^{(s)} + o(\hat{\alpha}^{(s)})$, где $\hat{\alpha}^{(s)} = \exp\left(-n(2\hat{\delta}^{(n_s, s)})^2/2\right) = \exp\left(-2n(\hat{\delta}^{(n_s, s)})^2\right)$, и учесть, что при этом вероятность события $R_s \stackrel{\text{def}}{=} \bigcap_{k=1}^s \left\{ \nu_{i_k}^{(n_s)} - \hat{\delta}^{(n_s, s)} \leq \text{pr}_{i_k} \leq \nu_{i_k}^{(n_s)} + \hat{\delta}^{(n_s, s)} \right\}$ больше¹⁾, чем $1 - 2s\hat{\alpha}^{(s)}$. Дело в том, что событие $\left\{ \nu_{i_1}^{(n_s)} + \hat{\delta}^{(n_s, s)} \geq \text{pr}_{i_1} \geq \nu_{i_1}^{(n_s)} - \hat{\delta}^{(n_s, s)} > \nu_{i_2}^{(n_s)} + \hat{\delta}^{(n_s, s)} \geq \text{pr}_{i_2} \geq \nu_{i_2}^{(n_s)} - \hat{\delta}^{(n_s, s)} > \dots > \nu_{i_s}^{(n_s)} + \hat{\delta}^{(n_s, s)} \geq \text{pr}_{i_s} \geq \nu_{i_s}^{(n_s)} - \hat{\delta}^{(n_s, s)} \right\}$, интервально оценивающее вероятности $\text{pr}_{i_1} > \text{pr}_{i_2} > \dots > \text{pr}_{i_s}$, равно событию $Q_s \cap R_s$, и, следовательно, его вероятность больше, чем $1 - (3s-1)\hat{\alpha}^{(s)} + o(\hat{\alpha}^{(s)})$.

3.7.3. Алгоритм эмпирического упорядочения вероятностей элементарных событий, изменяющихся в процессе испытаний. В этом параграфе моделью последовательности n взаимно независимых испытаний является вероятностное пространство $(\Omega^n, \mathcal{P}(\Omega^n), \Pr_{1, \dots, n}^{(n)}) = (\Omega, \mathcal{P}(\Omega), \Pr_1) \times \dots \times (\Omega, \mathcal{P}(\Omega), \Pr_n)$, охарактеризованное во введении.

Определим "эмпирические вероятности" элементарных событий и оценивающие их частоты равенствами

$$\text{pr}_i^{(n)} = \frac{1}{n} \sum_{j=1}^n \Pr_j(\omega_i), \quad \nu_i^{(n)} = \frac{1}{n} \sum_{j=1}^n \chi_i(\omega^{(j)}), \quad i = 1, 2, \dots, \quad n = 1, 2, \dots,$$

в которых $\chi_i(\cdot)$ — индикаторная функция $\{\omega_i\}$, $i = 1, 2, \dots$, и, разумеется, $\mathbf{E}_{(1, \dots, n)}^{(n)} \nu_i^{(n)} = \text{pr}_i^{(n)}$, причем при $n \rightarrow \infty$ $\nu_i^{(n)} - \text{pr}_i^{(n)} \xrightarrow{\text{п.н.}} 0$. Поскольку упорядоченности как вероятностей $\Pr_j(\{\omega_1\})$, $\Pr_j(\{\omega_2\})$, ...

¹⁾ Если $\Pr(A_i) \geq 1 - \hat{\alpha}_i$, то $\Pr\left(\bigcap_{i=1}^s A_i\right) \geq 1 - \sum_{i=1}^s \hat{\alpha}_i$.

для всех $j = 1, 2, \dots$, так и "эмпирических вероятностей" $\text{pr}_1^{(n)}, \text{pr}_2^{(n)}, \dots$ для любого $n = 1, 2, \dots$ одинаковы, а упорядоченность последних может быть определена эмпирически, то эмпирически может быть определена и упорядоченность $\text{Pr}_j(\{\omega_1\}), \text{Pr}_j(\{\omega_2\}), \dots$

Рассмотрим аналогичный алгоритму (7.19) алгоритм эмпирического упорядочения "эмпирических вероятностей" $\text{pr}_1^{(n)}, \text{pr}_2^{(n)}, \dots$: для всех $i < k$, $i, k \in \{1, \dots, s\}$, и каждого $n = 1, 2, \dots$:

- ▶ если $\nu_{ik}^{(n)} > \delta^{(n,s)}$, то $\square_1$: считать $\text{pr}_{ik}^{(n)} > 0$;
 - ▶ если $\nu_{ik}^{(n)} < -\delta^{(n,s)}$, то $\square_2$: считать $\text{pr}_{ik}^{(n)} < 0$;
 - ▶ если $|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}$, то $\circ$: продолжить испытания.
- (7.27)

В (7.27)

$$\text{pr}_{ik}^{(n)} = \text{pr}_i^{(n)} - \text{pr}_k^{(n)}, \quad \nu_{ik}^{(n)} = \nu_i^{(n)} - \nu_k^{(n)}, \quad 1 \leq i < k \leq s, \quad n = 1, 2, \dots$$

Свойства алгоритма (7.27) подобны свойствам алгоритма (7.19). Для определения значений $\delta^{(n,s)}$ зададим оценки сверху $\alpha^{(s)}$, $s = 1, 2, \dots$, вероятностей ошибочных решений $\square_1$ и $\square_2$. Если выполнено событие $\{\nu_{ik}^{(n)} > \delta^{(n,s)}\}$ и принято решение $\text{pr}_{ik}^{(n)} > 0$, в то время как на самом деле $\text{pr}_{ik}^{(n)} < 0$, то вероятность ошибочного решения $\square_1$

$$\begin{aligned} \text{Pr}_{1, \dots, n}^{(n)}(\{\nu_{ik}^{(n)} > \delta^{(n,s)}\} | \text{pr}_{ik}^{(n)} < 0) &\leq \text{Pr}_{1, \dots, n}^{(n)}(\{\nu_{ik}^{(n)} - \text{pr}_{ik}^{(n)} > \delta^{(n,s)}\} | \text{pr}_{ik}^{(n)} < \\ &< 0) \leq \exp(-n(\delta^{(n,s)})^2/2) \stackrel{\text{def}}{=} \alpha^{(s)}. \end{aligned} \quad (7.28)$$

В (7.28) заданное априори значение $\alpha^{(s)}$ оценивает сверху вероятность ошибочного решения $\square_1$ и определяет значение

$$\delta^{(n,s)} = \left(\frac{2}{n} \ln \frac{1}{\alpha^{(s)}} \right)^{1/2}, \quad n = 1, 2, \dots, \quad s = 2, 3, \dots, \quad (7.29)$$

такое же, как и в (7.21).

Аналогично вероятность ошибочного решения $\square_2$

$$\text{Pr}_{1, \dots, n}^{(n)}(\{\nu_{ik}^{(n)} < -\delta^{(n,s)}\} | \text{pr}_{ik}^{(n)} > 0) \leq \exp(-n(\delta^{(n,s)})^2/2) = \alpha^{(s)}.$$

Наконец, что касается событий $\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\}$, $n = 1, 2, \dots$, приводящих к решению $\circ$ продолжить испытания, то, в отличие от рассмотренного в параграфе 3.7.1 случая неизменной вероятности, контролирующей исходы испытаний, при изменяющейся вероятности условие, при котором будет выполняться п.н. конечное число таких событий, требует дополнительных предположений о модели испытаний.

Речь, разумеется, идет о моделях, в которых при $n \rightarrow \infty$ для некоторых ¹⁾ $i \neq k$ $\text{pr}_{ik}^{(n)} \rightarrow 0$. В следующей лемме показано, что если при $n \rightarrow \infty$ $\text{pr}_{ik}^{(n)} \rightarrow 0$, но "не слишком быстро", то решение $\odot$ в (7.27) всякий раз будет принято лишь для п.н. конечного числа испытаний.

Лемма 3.7.1. *Если при всех достаточно больших n и всех $i, k = 1, \dots, s$ $|\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}$, где $\delta^{(n,s)}$ определены в (7.29), $\varepsilon_{n,s} > 0$ и удовлетворяют условиям $\sum_{n=1}^{\infty} (\alpha^{(s)})^{\varepsilon_{n,s}^2} < \infty$, в которых $\alpha^{(s)} = \alpha/(s-1)$, $s = 2, 3, \dots$, то с вероятностью единица для каждого $s = 2, 3, \dots$ происходит лишь конечное число событий ²⁾ $\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\}$, $n = 1, 2, \dots$*

Доказательство. Если $|\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}$, то событие $\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\}$ влечет событие $\{|\nu_{ik}^{(n)} - \text{pr}_{ik}^{(n)}| \geq \varepsilon_{n,s}\delta^{(n,s)}\}$, ибо $|\nu_{ik}^{(n)} - \text{pr}_{ik}^{(n)}| \geq |\text{pr}_{ik}^{(n)}| - |\nu_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)} - \delta^{(n,s)} = \varepsilon_{n,s}\delta^{(n,s)}$. Следовательно, $\text{Pr}_{1,\dots,n}^{(n)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}) \leq \text{Pr}_{1,\dots,n}^{(n)}(\{|\nu_{ik}^{(n)} - \text{pr}_{ik}^{(n)}| \geq \varepsilon_{n,s}\delta^{(n,s)}\} | |\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}) \leq 2 \exp(-n(\varepsilon_{n,s}\delta^{(n,s)})^2/2) = 2(\alpha^{(s)})^{\varepsilon_{n,s}^2}$, и, в силу условий леммы, $\sum_{n=1}^{\infty} \text{Pr}_{1,\dots,n}^{(n)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}) = \sum_{n=1}^{\infty} \text{Pr}_{1,2,\dots}^{(\infty)}(\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\} | |\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}) \leq 2 \sum_{n=1}^{\infty} (\alpha^{(s)})^{\varepsilon_{n,s}^2} < \infty$, $s = 2, 3, \dots$

Согласно лемме Бореля-Кантелли это означает, что происходит п.н. конечное число событий в последовательности $\{|\nu_{ik}^{(n)}| \leq \delta^{(n,s)}\}$, $i, k = 1, \dots, s$, $n = 1, 2, \dots$, $s = 2, 3, \dots$ ■

Заметим, что при условиях, сформулированных в лемме 3.7.1, теорема 3.7.1 верна и в случае изменяющейся вероятности, а условия леммы 3.7.1 выполнены, если в последовательности $\text{Pr}_1, \text{Pr}_2, \dots$ вероятностей, контролирующих 1-ое, 2-ое, ... испытания, конечное число различных.

¹⁾ Если для всех $(i < k) \in \{1, \dots, s\}^2$ $|\text{pr}_{ik}^{(n)}| > \text{pr}_{ik} > 0$ при любом $n = 1, 2, \dots$, в частности, если среди вероятностей $\text{Pr}_1, \text{Pr}_2, \dots$ конечное число различных, то ситуация не отличается от рассмотренной в параграфе 3.7.1.

²⁾ Условие леммы можно сформулировать как требование к модели испытаний, а именно: для всех $j = 1, \dots, n$ при всех достаточно больших n $|\text{Pr}_j(\{\omega_i\}) - \text{Pr}_j(\{\omega_k\})| > (1 + \varepsilon_{n,s})\delta^{(n,s)} \Rightarrow |\text{pr}_{ik}^{(n)}| \geq (1 + \varepsilon_{n,s})\delta^{(n,s)}$, $i, k = 1, \dots, s$, $n = 1, 2, \dots$, $s = 2, 3, \dots$

3.8. Список обозначений Главы 3

- — конец доказательства;
- $\stackrel{\text{def}}{=}$ — равенство по определению;
- $\mathcal{A}, \mathcal{B}$ — σ -алгебры; $\mathcal{P}(\Omega)$ — класс всех подмножеств в Ω ;
- $\text{Pr}, \text{Pr}(\cdot)$ — вероятность;
- $\text{Pr}(\cdot|C)$ — условная при условии C вероятность;
- $\text{pr}_i \stackrel{\text{def}}{=} \text{Pr}(\{\omega_i\}), i = 1, 2, \dots$;
- $\mathbb{P}\text{r}$ — класс вероятностей $\text{Pr}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1], \Omega = \{\omega_1, \omega_2, \dots\}$,
удовлетворяющих условию $\text{pr}_1 \geq \text{pr}_2 \geq \dots, \text{pr}_1 + \text{pr}_2 + \dots = 1$;
- $\text{P}, \text{P}(\cdot)$ — возможность (мера возможности);
- $\text{P}(\cdot|C)$ — условная при условии C возможность;
- $\text{p}_i \stackrel{\text{def}}{=} \text{P}(\{w_i\}), i = 1, 2, \dots$;
- $\mathbb{P}$ — класс возможностей $\text{P}(\cdot) : \mathcal{P}(\Omega) \rightarrow [0, 1], \Omega = \{\omega_1, \omega_2, \dots\}$,
удовлетворяющих условию $1 = \text{p}_1 \geq \text{p}_2 \geq \dots$;
- $\mathcal{P}\text{r} \stackrel{\text{def}}{=} \{(\Omega, \mathcal{P}(\Omega), \text{Pr}), \text{Pr} \in \mathbb{P}\text{r}\}$;
- $\mathcal{P} \stackrel{\text{def}}{=} \{(\Omega, \mathcal{P}(\Omega), \text{Pr}), \text{Pr} \in \mathbb{P}\}$;
- $(\Omega, \mathcal{P}(\Omega), \text{Pr}), (X, \mathcal{A}, \text{Pr})$ — вероятностные пространства;
- $(\Omega, \mathcal{P}(\Omega), \text{P}), (X, \mathcal{A}, \text{P}), (X, \mathcal{P}(X), \text{P})$ — пространства с возможностью;
- $\mathcal{L} \stackrel{\text{def}}{=} ([0, 1], \leq, +, \bullet)$ — шкала значений возможности,
 $+ \sim \max, \bullet \sim \min$;
- $\text{N}, \text{N}(\cdot)$ — необходимость, мера необходимости;
- $\text{N}(\cdot|C)$ — условная при условии C необходимость;
- $\tilde{\mathcal{L}} \stackrel{\text{def}}{=} ([0, 1], \lesssim, \tilde{+}, \tilde{\bullet})$ — шкала значений необходимости,
 $\tilde{+} \sim \min, \tilde{\bullet} \sim \max$;
- Γ — класс непрерывных строго монотонно возрастающих
функций $\gamma(\cdot) : [0, 1] \rightarrow [0, 1], \gamma(0) = 0, \gamma(1) = 1$,
группа автоморфизмов шкалы $\mathcal{L}$;
- $\tilde{\Gamma}$ — класс монотонно возрастающих функций
 $\tilde{\gamma}(\cdot) : [0, 1] \rightarrow [0, 1], \tilde{\gamma}(0) = 0, \tilde{\gamma}(1) = 1$;
- $\tilde{\Gamma}(\text{Pr})$ — класс функций $\tilde{\gamma}(\cdot) \in \tilde{\Gamma}$,
определяющих возможность P , максимально
согласованную с вероятностью Pr ;
- $\text{Pr} \approx_{>}^{\mathcal{A}} \text{P} : \text{P}(A) = \tilde{\gamma}(\text{Pr}(A)), A \in \mathcal{P}(\Omega), \Omega = \{\omega_1, \omega_2, \dots\}$;
- $\text{Pr} \approx_{>}^{\mathcal{A}} \text{P}$: возможность P ,
максимально согласованная с вероятностью Pr на σ -алгебре $\mathcal{A}$;
- Θ — класс непрерывных строго монотонно убывающих
функций $\vartheta(\cdot) : [0, 1] \rightarrow [0, 1], \vartheta(0) = 1, \vartheta(1) = 0$;

$\mathcal{L}(X), \tilde{\mathcal{L}}(X)$ – классы функций $f(\cdot) : X \rightarrow [0, 1]$;
 $\mathcal{L}(\mathcal{L}(X))$ – класс интегралов $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$;
 $p(\cdot) : \mathcal{L}(X) \rightarrow [0, 1]$ – интеграл;
 $p_g(f(\cdot)) \stackrel{\text{def}}{=} \sup_{x \in X} \min(f(x), g(x))$;
 $n(\cdot) : \tilde{\mathcal{L}}(X) \rightarrow [0, 1]$ – интеграл;
 $n_h(f(\cdot)) \stackrel{\text{def}}{=} \inf_{x \in X} \max(f(x), h(x))$;
 $\Gamma \circ \stackrel{\text{def}}{=} \{\gamma \circ : \mathcal{L}(X) \rightarrow \mathcal{L}\} \quad \gamma \circ f(\omega) = \gamma(f(\omega)), \omega \in \mathcal{X}, f(\cdot) \in \mathcal{L}(X)$;
 $\Gamma * \stackrel{\text{def}}{=} \{\gamma * : \mathcal{L}(\mathcal{L}(X)) \rightarrow \mathcal{L}\} \quad \gamma * p(f(\cdot)) = \gamma(p(\gamma^{-1} \circ f(\cdot)))$;
 $\gamma \circ P(A) = \gamma(P(A)), A \in \mathcal{P}(\Omega)$;
 $\text{Pr} \sim > P$ ($\text{Pr} \approx > P$) – возможность P согласована
 (максимально согласована) с вероятностью Pr ;
 $\mathbb{P}_{(e)}$ – класс возможностей, удовлетворяющих условиям
 $1 = p_1 \stackrel{0.}{\geq} p_2 \stackrel{e_1}{\geq} p_3 \stackrel{e_2}{\geq} \dots$,
 где $e = 0.e_1e_2\dots$ – двоичная запись числа из $[0, 1]$,
 $e_i = 1 \sim \llcorner \ggcorner, e_i = 0 \sim \llcorner = \ggcorner, i = 1, 2, \dots$;
 $\mathbb{Pr}_{(e)}$ – класс вероятностей, с каждой из которых максимально согласована
 каждая возможность $P \in \mathbb{P}_{(e)}$, $\forall \text{Pr} \in \mathbb{Pr}_{(e)}, \forall P \in \mathbb{P}_{(e)} \text{Pr} \approx > P$;
 $\mathbb{Pr}_{(e)} = \{\text{Pr} \in \mathbb{Pr}, \mathbb{P}(\text{Pr}) = \mathbb{P}_{(e)}\}$;
 $\mathbb{P}_{(e)} = \{P \in \mathbb{P}, \mathbb{P}(P) = \mathbb{P}_{(e)}\}, e \in [0, 1]$;
 $\mathbb{P} = \bigcup_{e \in [0, 1]} \mathbb{P}_{(e)}, \mathbb{Pr} = \bigcup_{e \in [0, 1]} \mathbb{Pr}_{(e)}$;

Список литературы

1. Биркгоф Г. Теория решеток. М.: Наука, 1984.
2. Волков Б. И., Пытьев Ю. П. Измерительно-вычислительные преобразователи на основе датчиков с сосредоточенными параметрами. // Журнал вычислительной математики и математической физики, 2003, т. 43, № 8, с. 1265–1280.
3. Гроот М. Де. Оптимальные статистические решения. М.: Мир, 1974.
4. Демьянов В. Ф., Малоземов В. Н. Введение в минимакс. М.: Наука, 1972.
5. Дюбуа Д., Прад А. Теория возможностей. — М.: Радио и связь, 1990
6. Жучко О. В., Пытьев Ю. П. Восстановление функциональной зависимости теоретико-возможностными методами. // ЖВМ и МФ, 2003, т. 43, № 5, с. 765-781.
7. Нечеткие множества в моделях управления и искусственного интеллекта // Под ред. Д. А. Поспелова. — М.: Наука, 1986.
8. Нечеткие множества и теория возможностей, ред. Р. Ягер, М.: Радио и связь, 1986.
9. Орловский С. А. Проблемы принятия решения при нечеткой исходной информации. М., Наука, 1981.
10. Петров В. В. Предельные теоремы для сумм независимых случайных величин. М.: Наука, 1987.
11. Поппер К. Логика и рост научного знания. М., 1983.
12. Пытьев Ю. П. Возможность. Элементы теории и применения. — М.: Эдиториал УРСС, 2000 г.
13. Пытьев Ю. П. Возможность как альтернатива вероятности. Математические и эмпирические основы, применение. М.: Физматлит, 2007.
14. Пытьев Ю. П. Математические методы и алгоритмы эмпирического восстановления стохастических и нечетких моделей. // Интеллектуальные системы, т. 11, вып. 1÷4, с. 277–327, М. 2007.
15. Пытьев Ю. П. Математические методы интерпретации эксперимента. М.: Высшая школа, 351 с., 1989.
16. Пытьев Ю. П. Методы математического моделирования измерительно-вычислительных систем. М.: Физматлит, 2004.
17. Пытьев Ю. П. Неопределенные нечеткие модели и их применения. // Интеллектуальные системы. — 2004. — т. 8 вып. 1÷4. — с. 147–310.
18. Пытьев Ю. П. Измерительно-вычислительные преобразователи как средство измерения //Автоматика и телемеханика №2, 2010, с. 141-158.
19. Пытьев Ю. П. Эмпирическое восстановление мер возможности и правдоподобия возможности в моделях экспертных решений. // Автоматика и телемеханика, № 3, 2010, с. 131–146.
20. Рао С. Р. Линейные статистические методы и их применения. М.: Наука 1968.

21. *Теребиж В. Ю.* Восстановление изображений при минимальной априорной информации // Успехи физических наук. — Т. 165, № 2.
22. *Тихонов А. Н., Арсенин В. Я.* Методы решения некорректных задач. — М.: Наука, 1979, — 286 с.
23. *Худсон Д.* Статистика для физиков. М.: Мир, 1970, — 296 с.
24. *Шилов Г. Е., Гуревич Б. Л.* Интеграл, мера и производная. М.: Наука, 1967.
25. *Bellman K., Kalaba R. and Zadeh L.* Abstraction and Pattern Classification. // Journal of Mathematical Analysis and Applications. — 1966. — v. 13. p. 1–7.
26. *Cootan G. de.* Possibility theory I, II, III. // International Journal of General Systems. — 1997. — 25. — P. 291–371.
27. *Hoeffding W.* Probability of sums of bounded random variables // J. Amer. Statist. Assoc. — 1963. - V. 58, № 301, pp. 213–226.
28. *Koopman B. O.* The axioms and algebra of intuitive probability. // Ann. of Math., 41 (1940), 269–292.
29. *Kyburg Jr. , Smokler H. E. (ed).* Studies in Subjective probability, John Wiley and Sons, Inc., N.Y., 1964.
30. *Marquardt D. W.* Generalized Inverse, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation // Technometrics. — 1970. — V. 12, № 3. — P. 591–612.
31. *Matveeva T. V., Pyt'ev Yu. P.* On Possibility–Theoretic Methods for Measurement Interpretation. // Pattern Recognition and Image Analysis. — 2002. — v. 12, № 3, p.316–325.
32. *Pyt'ev Yu. P.* Methods of the Theory of Possibilities in the Problems of Optimal Estimation and Decision Making: III, IV. // Pattern Recognition and Image Analysis. — 1999. — v. 9, № 3. — p. 416–426; — 2000. — v. 10, № 1. — p. 43–52.
33. *Savage L. J.* The Foundations of Statistics, Dover. — New-York, 1972.
34. *Shackle G. L. S.* Decision, order and time in human affairs. 2nd edition. — Cambridge University Press, 1961.
35. *Shafer G.* A mathematical theory of evidence. Princeton N. J.: Princeton University Press, 1976.
36. *Sugeno M.* Fuzzy Measure and Fuzzy Intergral. // Trans. SICE, 1972, v.8, N 2, pp.95–102.
37. *Wolkenhauer O.* Possibility Theory with Applications to Data Analysis. — Research Studies Press, 1998.
38. *Zadeh L. A.* Fuzzy Sets. // Information and Control, 1965, v. 8, pp. 235–350.
39. *Zadeh L. A.* Fuzzy sets as a basis for a theory of possibility. // Fuzzy Sets and Systems, 1978, N 1, pp. 3–28.

Рекомендованные учебные пособия

40. Кибзун А. И., Горяинова Е. Р., Наумов А. В. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. М.: Физматлит, 2005.
41. Феллер В. Введение в теорию вероятностей и ее приложения. М.: Мир, 1984.
42. Ширяев А. Н. Вероятность. М.: МЦНМО, 2004.

