Методы математической статистики. Лекции.

Пособие предназначено для освоения материала по математической статистике для студентов 4 курса физического факультета МГУ. Материал данного пособия несколько шире, чем был изложен на лекциях. Доказательства теорем и некоторые необязательные сведения выделены мелким шрифтом. Список использованной литературы и краткое оглавление приведены на последних страницах. (1-й вариант, 20 декабря 2015 г.)

1 Введение

В математической статистике на основе исходных экспериментальных данных (как правило, это наблюдения над случайными величинами) требуется вынести то или иное суждение или решение о природе рассматриваемого явления, точнее, о параметрах математической (вероятностной) модели, описывающей это явление. Речь идет об интерпретации (анализе) данных эксперимента с ощутимой (не пренебрежимой) изменчивостью.

Среди содержательных статистических задач можно выделить следующие типы:

1) компактное описание информации, полученной в ходе исследования (построение вариационных рядов, расчет неизвестных параметров распределений, статистические группировочные таблицы и пр.);

2) нахождение и измерение взаимосвязи между признаками явления (корреляционный, дисперсионный анализ и пр.);

3) описание связей и взаимосвязей между признаками изучаемых явлений (регрессионный анализ, математическое моделирование);

4) выявление латентных (скрытых) факторов, детерминирующих связи изучаемого набора параметров данного явления (факторный, латентный анализ);

5) классификация признаков объектов, в том числе построение типологий (дискриминантный, кластерный анализ);

6) проверка содержательных гипотез статистическими методами, оценка значимости расчетных значений статистических показателей и параметров распределений;

7) прогнозирование путем выявления основных тенденций развития определенного процесса (временные, динамические ряды).

Как правило, исследования ведутся в рамках некоторого семейства вероятностных моделей, зачастую полностью определенных с точностью до ограниченного числа неизвестных параметров.

Кроме известных из теории вероятностей величин, таких, как функция распределения $F_\vartheta(x)$, характеристическая функция $f(t)$, моменты $\mathsf{E}\xi^k$, в статистике используются специфические понятия и обозначения[1]:

мода — точка максимума плотности вероятности,

квантиль (p-квантиль) $x_p = \mathrm{res}(F(x) = p)$, $0 < p < 1$,

в частности,

$x_{0.5}$ —медиана,

$x_{0.75}$ — верхняя квартиль,

$x_{0.25}$ — нижняя квартиль,

[1]Для обозначения математического ожидания в математической статистике чаще используется символ E — от иностранных слов: **expectation** (англ.) или **Erwartung** (нем.)

$x_{0.1}$ $x_{0.2}$,...,$x_{0.9}$ — децили.

Кроме моментов в математической статистике используются следующие параметры распределений :

асимметрия: $\gamma_1 = \dfrac{\mathsf{E}\Big(\xi - \mathsf{E}\xi\Big)^3}{\Big(\mathsf{D}\xi\Big)^{3/2}}$

эксцесс: $\gamma_2 = \dfrac{\mathsf{E}\Big(\xi - \mathsf{E}\xi\Big)^4}{\Big(\mathsf{D}\xi\Big)^2}$ или $\widetilde{\gamma}_2 = \gamma_2 - 3$.

Примеры (проверьте самостоятельно).

Нормальное распределение $\mathcal{N}(0,1)$: $\gamma_1 = 0$, $\widetilde{\gamma}_2 = 0$.

Биномиальное распределение: $\gamma_1 = \frac{q-p}{\sqrt{npq}}$, $\widetilde{\gamma}_2 = \frac{1}{npq} - \frac{6}{n}$.

Равномерное распределение $U(-1,1)$: $\gamma_1 = 0$, $\widetilde{\gamma}_2 = -\frac{6}{5}$.

Центральным понятием математической статистики является выборка.

1.1 Определения

1. Однородной выборкой (выборкой) объема n при $n \geqslant 1$ называется случайный вектор $\vec{\xi} = (\xi_i, ..., \xi_n)^{\mathsf{T}}$, координаты которого ξ_i, $i = 1, \ldots, n$, называемые элементами выборки, являются независимыми случайными величинами с одной и той же функцией распределения $F(x)$. Будем говорить, что выборка $\vec{\xi}$ соответствует функции распределения $F(x)$.

2. Реализацией выборки называется неслучайный вектор $\vec{x} = (x_i, ..., x_n)^{\mathsf{T}}$, координатами которого являются реализации соответствующих элементов выборки $\xi_i, i = 1, \ldots, n$. Из определений 1 и 2 вытекает, что реализацию выборки $\vec{\xi}$ можно также рассматривать как последовательность $x_1..., x_n$ из n реализаций одной и той же случайной величины ξ, полученных в серии из n независимых одинаковых опытов, проводимых в одинаковых условиях. Поэтому можно говорить, что выборка $\vec{\xi}$ порождена наблюдаемой случайной величиной ξ, имеющей распределение $F_\xi(x) = F(x)$.

3. Если координаты вектора ξ_n независимы, но их распределения $F_1(x_1), ..., F_n(x_n)$ различны, то такую выборку называют неоднородной.

4. Множество X всех реализаций выборки ξ_n называется **выборочным пространством** или *генеральной совокупностью*. Выборочное пространство может быть всем n-мерным евклидовым пространством R^n или его частью, если случайная величина ξ непрерывна, а также может состоять из конечного или счетного числа точек из R^n, если случайная величина ξ дискретна. На практике при исследовании конкретного эксперимента распределения $F_1(x_1), ..., F_n(x_n)$ случайных величин $\xi_i, ..., \xi_n$ редко бывают известны полностью. Часто априори (до опыта) можно лишь утверждать, что распределение $F_{z_n}(z_n) = F_1(x_1) \cdot ... \cdot F_n(x_n)$ случайного вектора ξ_n принадлежит некоторому классу (семейству) $\mathcal{P}$.

Пара $(X, \mathcal{A})$, где $(\mathcal{A})$ — σ-алгебра подмножеств X является измеримым пространством, а тройка $(X, \mathcal{A}, \mathcal{P})$ — статистической структурой.

5. Если распределения $F(x, \vartheta)$ из класса $\mathcal{P}$ определены с точностью до некоторого векторного параметра $\vartheta \in \Theta \subset \mathcal{R}^s$, то такая статистическая модель называется параметрической. В некоторых случаях выборочное пространство может не зависеть от неизвестного параметра распределения (ϑ). В зависимости от вида статистической модели в математической статистике формулируются соответствующие задачи по обработке информации, содержащейся в выборке.

6. Случайная величина $\eta = \varphi(\vec{\xi})$, где $\varphi(\cdot)$ — произвольная измеримая функция от выборки, определенная на выборочном пространстве X и не зависящая от распределения F, называется статистикой.

7. **Вариационный ряд**. Упорядочим элементы реализации выборки $x_i, ..., x_n$ по возрастанию: $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$, где индекс в скобках соответствует номеру элемента в упорядоченной последовательности. Обозначим через $\xi_{(k)}$, $k = 1, \ldots, n$, случайные величины,

которые при каждой реализации $\vec{x}$ выборки $\vec{\xi}$ принимают k-ое значения $x_{(k)}$. Упорядоченную последовательность случайных величин $\xi_{(1)} \leqslant \xi_{(2)} \leqslant \ldots \leqslant \xi_{(n)}$ называют вариационным рядом выборки.

8. Элементы $\xi_{(k)}$ вариационного ряда называются порядковыми статистиками, а крайние члены вариационного ряда $\xi_{(1)}$, $\xi_{(n)}$ — экстремальными порядковыми статистиками.

Например, для $k = 1$ функция $\varphi(\vec{\xi})$ для статистики $\xi_{(1)} = \varphi(\vec{\xi})$ определяется следующим образом:

$$\varphi(\vec{\xi}) = \min\{\xi_{(k)} : k = 1, \ldots, n.\}$$

Если однородная выборка ξ_n соответствует распределению $F(x)$, то k-ая порядковая статистика $\xi_{(k)}$ имеет следующую функцию распределения:

$$F_{(k)}(x) = P\{\xi_{(k)} < x\} = \sum_{i=k}^{n} C_n^i [F(x)]^i [1 - F(x)]^{n-i}.$$

В частности, для $k = 1$ и $k = n$ имеем

$$F_{(1)}(x) = 1 - [1 - F(x)]^n, \quad F_{(n)}(x) = [F(x)]^n.$$

Ранги и ранжирование. Рангом называется номер в упорядоченной (обычно по возрастанию) совокупности. При совпадении значений обычно берут средний номер (средние ранги).

Ранжирование — переход к последовательности рангов.

Ранжировка — результат перехода.

Вариационныый ряд — совокупность, упорядоченная в соответствии с возрастанием рангов (обозначение $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$).

Выборочная (эмпирическая) функция распределения

$$F^*(x) = F_n^*(x) = \frac{k}{n}, \quad x \in (x_{(k)}, x_{(k+1)}], \quad k = 1, \ldots, n-1,$$
$$F^*(x) = 0, \quad x \leqslant x_{(1)}, \qquad F^*(x) = 1, \quad x > x_{(n)}.$$

Выборочное математическое ожидание — $\overline{x} = \frac{1}{n}\sum\limits_{1}^{n} x_i$

Выборочная дисперсия — $\frac{1}{n}\sum\limits_{1}^{n}(x_i - \overline{x})^2$ или (несмещенная) $\frac{1}{n-1}\sum\limits_{1}^{n}(x_i - \overline{x})^2$

Выборочная ковариация — $\frac{1}{n}\sum\limits_{1}^{n}(x_i - \overline{x})(y_i - \overline{y})$.

Соответствующим образом определяются выборочные асимметрия и эксцесс.

Вернемся к задачам, стоящим перед математической (теоретической) статистикой.

Исследуются вопросы:

1). Согласуются ли данные с выбранным семейством вероятностных моделей.

2). Какие заключения можно сделать о значениях неизвестных параметров и функций от них (эти проблемы связаны, но методически их чаще всего различают).

В рамках параметрической статистики можно представить себе эксперимент как вероятностный автомат:

$$\vartheta \longrightarrow \boxed{P_\vartheta} \longrightarrow x$$

Всякий статистический анализ должен по наблюдению x вынести решение $\delta(x)$, $\delta(x) = d \in D$ относительно истинного значения параметра ϑ. $\delta(\cdot) : \ x \mapsto D$ — решающая функция (правило) или статистическая стратегия (критерий).

$D = \{d\}$ — множество возможных решений относительно истинного значения параметра ϑ (пространство решений).

Как правило, ограничиваются априори некоторым множеством Δ допустимых стратегий.

Случай 1. Решение — точка ϑ_0 из Θ, $D = \Theta$, — теория точечных оценок. Δ — например, несмещенные оценки.

Случай 2. Решение — некоторое подмножество Θ и $D \subset \mathcal{M}(\Theta)$, где $\mathcal{M}(\Theta)$ — множество всех подмножеств множества Θ. Это теория доверительных множеств.

Случай 3. Пусть $\Theta = \Theta_1 + \Theta_2 + \dots + \Theta_s$, $D = (d_1, \dots, d_s)$, где d_i — решения вида $\vartheta \in \Theta_i$. Это теория проверки сложных гипотез. Задача решается до конца в случае одноточечных множеств ϑ_i (простые гипотезы).

Конечная цель статистического исследования — выбор стратегии. При этом желательно определить критерии предпочтительности.

2 Точечные оценки.

Пусть $\{\xi_i\}$, $i = 1, 2, \dots, n$ — независимая выборка из распределения $P(x, \vartheta)$, где ϑ — неизвестный параметр. Нас интересует оценка величины $\tau(\vartheta)$ (здесь $\tau(\cdot)$ — известная функция), причем роль оценки играет некоторая *статистика* $t(\xi)$.

Терминология: X — выборочное пространство, n — объем выборки, всякая измеримая функция от выборки называется статистикой, следовательно по определению любая точечная оценка — статистика.

Желательные свойства оценок:

1. Несмещенность $\mathsf{E}t(\xi) = \tau(\vartheta)$. (Гарантирует от накопления систематических ошибок).

2. Состоятельность $T_n(\xi) \xrightarrow[n\to\infty]{P} \tau(\vartheta)$ (фактически рассматривается последовательность оценок).

3. Минимальность дисперсии (если оценка несмещенная) — качество оценки при фиксированном объеме выборки.

Примеры.

Несмещенность $\widehat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ очевидна. Состоятельность $\widehat{\mu}$ — утверждение З.Б.Ч.

Если $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\widehat{\sigma}^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(\xi_k - \widehat{\mu})^2$ — несмещенная и состоятельная оценка. Действительно, при этом $\widehat{\sigma}^2 = \frac{1}{n-1}\sigma^2\chi^2_{n-1}$, $\mathsf{E}\widehat{\sigma}^2 = \sigma^2$, а $\mathsf{D}\widehat{\sigma}^2 = \frac{\sigma^4}{(n-1)^2}\mathsf{D}\chi^2_{n-1} = \frac{\sigma^4}{(n-1)^2}2(n-1)$ и по неравенству Чебышёва $P\{|\widehat{\sigma}^2 - \sigma^2| > \varepsilon\} < \frac{2\sigma^4}{\varepsilon^2(n-1)} \xrightarrow[n\to\infty]{} 0$. Если ξ не является нормальной, то несмещенность оценки сохраняется:

$$\mathsf{E}\sum(\xi_i - \widehat{\mu})^2 = \mathsf{E}\sum[(\xi_i - \mu)^2 - 2(\xi_i - \mu)(\widehat{\mu} - \mu) + (\widehat{\mu} - \mu)^2] =$$

$$= \mathsf{E}[\sum(\xi_i - \mu)^2 - 2(\widehat{\mu} - \mu)\sum(\xi_i - \mu) + \sum(\widehat{\mu} - \mu)^2] =$$

$$= \mathsf{E}[\sum(\xi_i - \mu)^2 - 2\frac{1}{n}\sum(\xi_i - \mu)\sum(\xi_i - \mu) + \frac{n}{n^2}\sum(\xi_i - \mu)\sum(\xi_i - \mu)] =$$

$$= n\sigma^2 - 2\sigma^2 + \sigma^2 = (n-1)\sigma^2,$$

а состоятельность — нет.

Минимальность дисперсии — желательное свойство, однако заметим, что смещение может уменьшить ср. кв. уклонение. Например, задача

$$\mathsf{E}(k\sum(\xi_i - \widehat{\mu})^2 - \sigma^2)^2 \sim \min_k$$

имеет решение для $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$ при $k = \frac{1}{n+1}$. (Доказать).

Рассмотрим специально несмещенные оценки минимальной дисперсии (НОМД).

Лемма. Если существует НОМД, то она единственна (с вероятностью 1).

Доказательство. Пусть $t_1(\xi)$ и $t_2(\xi)$ — НОМД, т.е. $\mathsf{E}t_i(\xi) = \tau(\vartheta)$, $\mathsf{D}t_i(\xi) = \delta$, $i = 1, 2$.

Рассмотрим $t_3 = \frac{1}{2}(t_1 + t_2)$. Тогда $\mathsf{D}t_3 =$

$$= \frac{1}{4}(\mathsf{D}t_1 + 2\mathrm{cov}t_1t_2 + \mathsf{D}t_2) \leqslant \frac{1}{4}(\mathsf{D}t_1 + 2\sqrt{\mathsf{D}t_1\mathsf{D}t_2} + \mathsf{D}t_2) = \frac{1}{4}(\sqrt{\delta} + \sqrt{\delta})^2 = \delta,$$

но $\mathsf{D}t_3 \geqslant \delta$, т.е. возможно лишь равенство, откуда следует, что $t_1(\xi) - \tau(\vartheta) = k(\vartheta)(t_1(\xi) - \tau(\vartheta))$ с verояьностью 1 ($k^2 = 1$), и далее, из $\mathrm{cov}t_1t_2 = \delta$ получаем $k = 1$. □

Иногда качество оценки можно определить, зная минимально возможное значение ее дисперсии (неравенство Рао-Крамера).

Определение. Функцией правдоподобия для некоторого распределения $P(x, \vartheta)$ называется $L(x, \vartheta) = p(x_1, \vartheta)p(x_2, \vartheta) \ldots p(x_n, \vartheta)$, где $p(x_i, \vartheta)$ — либо плотность распределения $p_\xi(x, \vartheta)$ случайной величины ξ, либо $P_\vartheta\{\xi = x\}$[1].

2.1 Теорема Рао-Крамера.

(Одномерный скалярный случай.)

Пусть $L(x|\vartheta) = L(x_1, ..., x_n|\vartheta)$ — функция правдоподобия и пусть выполняются следующие условия:

1. $\vartheta \in \Theta \subset \mathcal{R}_1$, $\mathsf{E}t(\xi) = \tau(\vartheta)$.
2. L и τ - дифференцируемы по ϑ.
3. Множество $\{x \in \mathcal{R}_n,\ L(x|\vartheta) \neq 0\}$ не зависит от ϑ

$$\frac{d}{d\vartheta}\int L(x, \vartheta)dx = \int \frac{d}{d\vartheta}L(x, \vartheta)dx$$

и

$$\frac{d}{d\vartheta}\int t(x)L(x, \vartheta)dx = \int t(x)\frac{d}{d\vartheta}L(x, \vartheta)dx.$$

Тогда

$$\mathsf{D}_\vartheta t(\xi) \geqslant \frac{[\tau'(\vartheta)]^2}{\mathsf{E}(\frac{\partial \log L(\xi,\vartheta)}{\partial\vartheta})^2}. \tag{2.1}$$

Если в (2.1) выполняется равенство, то

$$\frac{\partial \log L(\xi, \vartheta)}{\partial\vartheta} = a(\vartheta)[t(\xi) - \tau(\vartheta)] \tag{2.2}$$

с вероятностью единица для некоторого $a(\vartheta)$.

Доказательство. Дифференцируя тождества $\int L(x, \vartheta)dx = 1$ и $\int t(x)L(x, \vartheta)dx = \tau(\vartheta)$, получим

$$\int \frac{d\ln L(x, \vartheta)}{d\vartheta}L(x, \vartheta)dx = 0, \int t(x)\frac{d\ln L(x, \vartheta)}{d\vartheta}L(x, \vartheta)dx = \tau'(\vartheta)$$

или

$$\int [t(x) - \tau(\vartheta)]\frac{d\ln L(x, \vartheta)}{d\vartheta}L(x, \vartheta)dx = \tau'(\vartheta)$$

и отсюда по неравенству Коши-Буняковского получаем (2.1). □

Если в условиях теоремы Рао-Крамера имеет место равенство, то справедливо (2.2) с вероятностью единица. В этом случае оценка называется эффективной и ее дисперсия равна $\mathsf{D}t(\xi) = \frac{|\tau'(\vartheta)|}{|a(\vartheta)|}$.

К таким оценкам, например, приводят распределения, плотности которых можно представить в виде

$$p(x, \vartheta) = \exp\{a(\vartheta)b(x) + c(\vartheta) + d(x)\},\ x \in R_1,\ \vartheta \in R_1.$$

[1] Обычно функция правдоподобия рассматривается как функция от ϑ, а значения $x_1, x_2, \ldots, x_n$ (выборка) — параметры. Если же, наоборот, ϑ считать параметром, то $L(x, \vartheta)$ можно считать плотностью вероятности вектора ξ с независимыми координатами.

(Они называются экспоненциальными семействами.) Тогда

$$L(x,\vartheta) = \exp\{a(\vartheta)\sum b(x_i) + nc(\vartheta) + \sum d(x_i)\}$$

и

$$\frac{\partial \ln L(\xi,\vartheta)}{\partial\vartheta} = a'(\vartheta)n\left\{\frac{1}{n}\sum b(x_i) + \frac{c'(\vartheta)}{a'(\vartheta)}\right\}.$$

Пусть условия теоремы Рао-Крамера выполнены, тогда $t(\xi) = \frac{1}{n}\sum b(x_i)$ есть эффективная оценка $\tau(\vartheta) = -\frac{c'(\vartheta)}{a'(\vartheta)}$ с дисперсией

$$\left|\frac{\tau'(\vartheta)}{na'(\vartheta)}\right|.$$

Экспоненциальному семейству принадлежат многие важные для практики распределения: нормальное, Пуассона, Бернулли (биномиальное), гамма-распределение и другие.

Примеры.

1. $\mathcal{N}(\vartheta,\sigma^2)$: $p(x|\vartheta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2\sigma^2}(x-\vartheta)^2\}$. $\frac{\partial\ln L}{\partial\vartheta} = \frac{n}{\sigma^2}[\overline{x}-\vartheta]$, т.е. $\overline{x}$ — эффективная оценка ϑ с дисперсией $\frac{\sigma^2}{n}$.

2. $\mathcal{N}(\mu,\vartheta^2)$. $\frac{\partial\ln L}{\partial\vartheta} = \frac{n}{\vartheta^3}[\frac{1}{n}\sum(x_k-\mu)^2 - \vartheta^2]$, поэтому для $\tau = \vartheta^2$ $\frac{1}{n}\sum(x_k-\mu)^2$ — эффективная оценка с дисперсией $\frac{2\vartheta^4}{n}$.

3. $f(x|\vartheta) = e^{-\vartheta}\frac{\vartheta^x}{x!}$, $x = 0,1,2,\ldots$. $\frac{\partial\ln L}{\partial\vartheta} = \frac{n}{\vartheta}[\overline{x}-\vartheta]$. $\overline{x}$ — эффективная оценка ϑ с дисперсией $\frac{\vartheta}{n}$

4. Распределение Коши $f(x|\vartheta) = \frac{1}{\pi}\frac{1}{1+(x-\vartheta)^2}$. $\frac{\partial\ln L}{\partial\vartheta} = 2\sum\frac{x_i-\vartheta}{1+(x_i-\vartheta)^2}$
— эффективной оценки ϑ не существует.

Замечание. Из (2.2) видно, что если н.о.м.д. существует, то она существует только лишь для специальной функции $\tau(\vartheta)$ параметра ϑ и не существует ни для какой функции, отличной от $c_1\tau(\vartheta)+c_2$, где c_1, c_2 — постоянные.

Многомерный аналог неравенства Рао-Крамера.

Теорема. Пусть $L_\xi(x,\vartheta)$, $x\in\mathcal{R}_n$, $\vartheta\in\mathcal{R}_m$, — семейство плотностей, отвечающих семейству распределений вероятностей $P(\cdot,\vartheta)$, $\vartheta\in\mathcal{R}_m$ случайного вектора $\xi\in\mathcal{R}_n$ и

$$\mathsf{E}_\vartheta t(\xi) = \int\limits_{\mathcal{R}_n} t(x)L_\xi(x,\vartheta)dx = \tau(\vartheta), \tag{2.3}$$

где[1] $\tau = (\tau_1,\ldots,\tau_k)^\mathsf{T}$, $t = (t_1,\ldots,t_k)^\mathsf{T}$ — известные функции, определенные на $\mathcal{R}_m$ и $\mathcal{R}_n$ соответственно, и принимающие значения в $\mathcal{R}_k$.

Обозначим $S = \{S_{ij}\}$, $M = \{M_{ij}\}$ и $D = \{D_{ij}\}$ матрицы, матричные элементы которых даются равенствами

$$\begin{aligned}
S_{ij} = \frac{\partial\tau_i(\vartheta)}{\partial\vartheta_j} = \mathsf{E}_\vartheta t_i\frac{\partial\ln L_\xi(\xi,\vartheta)}{\partial\vartheta_j}, &\quad i=1,\ldots,k,\ j=1,\ldots,m\\
0 = \mathsf{E}_\vartheta\frac{\partial\ln L_\xi(\xi,\vartheta)}{\partial\vartheta_j}, &\qquad\qquad j=1,\ldots,m\\
M_{ij} = \mathsf{E}_\vartheta\frac{\partial\ln L_\xi(\xi,\vartheta)}{\partial\vartheta_i}\frac{\partial\ln L_\xi(\xi,\vartheta)}{\partial\vartheta_j}, &\quad i,j=1,\ldots,m,\\
D_{ij} = \mathsf{E}_\vartheta(t_i(\xi)-\tau_i(\vartheta))(t_j(\xi)-\tau_j(\vartheta)), &\quad i,j=1,\ldots,k.
\end{aligned} \tag{2.4}$$

Предполагается, как и в одномерном случае, что дифференцирование может быть выполнено под знаком интеграла. Тогда справедливо (матричное) неравенство Крамера-Рао

$$D(\vartheta) \geqslant S(\vartheta)M^{-1}(\vartheta)S^\mathsf{T}(\vartheta), \quad \vartheta\in\mathcal{R}_m, \tag{2.5}$$

[1] Здесь T — знак транспонирования матриц или векторов.

причем равенство в (3.3) имеет место тогда и только тогда, когда существует $k \times m$ матрица $C = C(\vartheta)$, не зависящая от ξ, такая, что с вероятностью единица

$$t(\xi) - \tau(\vartheta) = C(\vartheta)\ \alpha(\xi, \vartheta), \tag{2.6}$$

где

$$\alpha(\xi, \vartheta) = \left(\frac{\partial \ln L_\xi(\xi, \vartheta)}{\partial \vartheta_1}, \dots, \frac{\partial \ln L_\xi(\xi, \vartheta)}{\partial \vartheta_m} \right)^\mathsf{T}. \tag{2.7}$$

Доказательство. Согласно обозначениям (3.1), (3.2) и (3.4)

$$D = \mathsf{E}_\vartheta (t-\tau)(t-\tau)^\mathsf{T}, \quad M = \mathsf{E}_\vartheta \alpha\alpha^\mathsf{T} = \Sigma_\alpha, \tag{2.8}$$

$$\mathsf{E}_\vartheta \alpha = 0, \quad S = \mathsf{E}_\vartheta t\alpha^\mathsf{T} = \mathsf{E}_\vartheta (t-\tau)\alpha^\mathsf{T} = \frac{\partial(\tau_i(\vartheta))}{\partial \vartheta_j} = \Sigma_{t\alpha}. \tag{2.9}$$

Для любой $k \times m$ матрицы C получаем

$$\begin{aligned} &0 \leqslant \mathsf{E}_\vartheta (t-\tau-C\alpha)(t-\tau-C\alpha)^\mathsf{T} = \\ &= \mathsf{E}_\vartheta (t-\tau)(t-\tau)^\mathsf{T} - CS^\mathsf{T} - SC^\mathsf{T} + CMC^\mathsf{T} = \\ &= \mathsf{E}_\vartheta (t-\tau)(t-\tau)^\mathsf{T} - SM^{-}S^\mathsf{T} + (C - SM^{-1})M(C-SM^{-1})^\mathsf{T}, \end{aligned} \tag{2.10}$$

поскольку
$(C - SM^{-1})M(C - SM^{-1})^\mathsf{T} = CMC^\mathsf{T} - CS^\mathsf{T} - SC^\mathsf{T} + SM^{-1}S^\mathsf{T}$, то есть $-CS^\mathsf{T} - SC^\mathsf{T} + CMC^\mathsf{T} = -SM^{-1}S^\mathsf{T} + (C - SM^{-1})M(C - SM^{-1})^\mathsf{T}$.

Наконец, положив в (2.10) $C = SM^{-1}$, получим неравенство (3.3).

Пусть существует матрица $C = C(\vartheta)$, такая, что выполняется равенство (3.4). Тогда $\mathsf{E}_\vartheta t\alpha^\mathsf{T} = C\mathsf{E}_\vartheta \alpha\alpha^\mathsf{T}$, или, иначе, $S = CM$. Следовательно, $(C - SM^{-1})M(C - SM^{-1})^\mathsf{T} = 0$ и, согласно равенству (2.10),

$$\mathsf{E}_\vartheta (t-\tau)(t-\tau)^\mathsf{T} = D = SM^{-1}S^\mathsf{T}. \tag{2.11}$$

Наоборот, пусть выполнено равенство (2.11). Тогда, согласно соотношениям (2.10) для всякой матрицы C

$$\mathsf{E}_\vartheta (t-\tau-C\alpha)(t-\tau-C\alpha)^\mathsf{T} = (C - SM^{-1})M(C - SM^{-1})^\mathsf{T}.$$

Выбрав $C = SM^{-1}$, найдем, что с вероятностью единица выполняется равенство (3.4). □

Оценка $t(\xi)$ для $\tau(\vartheta) = \tau(\vartheta_1, \dots, \vartheta_m)$ называется R-эффективной, если нижняя граница дисперсии этой оценки, задаваемая многомерным аналогом неравенства Рао-Крамера, достигается для всех $\vartheta = (\vartheta_1, \dots, \vartheta_m)$.

Пример. Чибисов, Пагурова, Задача 3.6 стр. 46. Пример 4. Пусть $\xi_1, ..., \xi_n$ — независимая выборка из $\mathcal{N}(\mu, \sigma^2)$. Проверить, является ли (*несмещенная!*) оценка $(\overline{\xi}, \frac{1}{n-1}\sum(\xi_i - \overline{\xi})^2)$ параметра (μ, σ^2) эффективной.

Решение.

$$\ln(L(x, \mu, \sigma^2)) = -\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2}\ln(\sigma^2),$$

$$\frac{\partial \ln(L(x, \mu, \sigma^2))}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu),$$

$$\frac{\partial \ln(L(x, \mu, \sigma^2))}{\partial \sigma^2} = \frac{1}{2\sigma^4}\left\{ \sum_{i=1}^n (x_i - \mu)^2 - n\sigma^2 \right\}.$$

Учитывая, что $\sum_{i=1}^{n}(\xi_i-\mu)^2=\sigma^2\chi_n^2$, $\mathsf{E}\chi_n^2=n$, $\mathsf{D}\chi_n^2=2n$, получаем

$$M_{1,1}=\frac{1}{\sigma^4}\mathsf{E}\left\{\sum_{i=1}^{n}(\xi_i-\mu)\right\}^2=\frac{1}{\sigma^4}\mathsf{E}\sum_{i=1}^{n}(\xi_i-\mu)^2=\frac{n}{\sigma^2},$$

$$M_{2,2}=\frac{1}{4\sigma^8}\mathsf{E}\left\{\sum_{i=1}^{n}(x_i-\mu)^2-n\sigma^2\right\}^2=\frac{1}{4\sigma^8}\mathsf{D}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{n}{2\sigma^4},$$

$$M_{1,2}=M_{2,1}=0.$$

Учитывая, что S — единичная матрица и вычисляя матрицу D для независимых $\overline{\xi}$ и $\frac{1}{n-1}\sum(\xi_i-\overline{\xi})$, получаем неравенство (3.3) в виде

$$D=\begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n-1}\end{pmatrix}>\begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n}\end{pmatrix}=M^{-1},$$

что означает, что оценка не является эффективной.

3 Распределение ортогональных проекций нормального вектора.

Рассмотрим n-мерное евклидово пространство $\mathcal{R}_n$, $x=(x_1,\ldots,x_n)\in\mathcal{R}_n$, $(x,y)=\sum_{k=1}^{n}x_ky_k$. Пусть $\xi=(\xi_1,\ldots,\xi_n)$ — случайный вектор в этом пространстве, причем $\xi\sim\mathcal{N}(0,\sigma^2I)$.

Напомним, что в общем случае, когда $\xi\sim\mathcal{N}(0,\Sigma)$,

$$p_\xi(x)=\frac{(\det\Sigma^{-1})^{1/2}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(x,\Sigma^{-1}x)\right\}.$$

Если $\Sigma=\sigma^2I$, где I — единичный оператор, то координаты ξ_i независимы и

$$p_{\xi_i}(x_i)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}x_i^2\right\}.$$

3.1 Ортогональный проектор. Ортогональное преобразование.

Пусть L — линейное подпространство $\mathcal{R}_n$ и $L^\perp$ — ортогональное дополнение L в $\mathcal{R}_n$, т. е. множество векторов $x\in\mathcal{R}_n$, ортогональных всем векторам из L:

$$L^\perp=\{x\in\mathcal{R}_n,\ (x,y)=0,\ y\in L\}.$$

Очевидно, $\mathcal{L}^\perp$ — также линейное пространство $\mathcal{R}_n$. Как известно, всякий вектор $x\in\mathcal{R}_n$ может быть представлен в виде суммы

$$x=x_1+x_2,\quad x_1\in\mathcal{L},x_2\in\mathcal{L}^\perp. \tag{3.1}$$

Разложение (3.1) единственно. Действительно, равенство $x=x_1+x_2=x_1'+x_2'$, $x_1'\in L$, $x_2'\in L^\perp$, совместно с (3.1) влечет $(x_1-x_1')^2+(x_2-x_2')^2=0$. Слагаемые в последнем равенстве ортогональны, так как $x_1-x_1'\in L$, $x_2-x_2'\in L^\perp$, 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 (3.1) ставится в соответствие единственный вектор $x_1\in L$:

$$x_1=\Pi x. \tag{3.2}$$

Π называется оператором ортогонального проецирования на L, или ортогональным проектором на L. Отметим следующие свойства оператора Π.

1) Π — линейный оператор. Действительно, пусть

$$x = x_1 + x_2,\ y = y_1 + y_2,\ x_1, y_1 \in L,\ x_2, y_2 \in L^{\perp}. \tag{3.3}$$

Тогда

$$\alpha x + \beta y = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2),\ \alpha x_1 + \beta y_1 \in L,\ \alpha x_2 + \beta y_2 \in L^{\perp}.$$

Следовательно, согласно определению (3.2) $\Pi \alpha x + \beta y = \alpha x_1 + \beta y_1 = \alpha \Pi x + \beta \Pi y$.

2) Π — самосопряженный оператор, т.е. для любых $x, y \in \mathcal{R}_n$ $(\Pi x, y) = (x, \Pi y)$. Действительно, воспользовавшись разложением (3.3), найдем

$$(\Pi x, y) = (x_1, y) = (x_1, y_1) = (x, y_1) = (x, \Pi y).$$

3) Оператор Π удовлетворяет уравнению $\Pi^2 = \Pi$ («идемпотентность»). Действительно, для всякого $x \in \mathcal{R}_n$: $\Pi x = x_1 = \Pi x_1 = \Pi(\Pi x)$, поскольку для $\in L$ разложение (3.1) имеет вид $x_1 = x_1 = \Pi x_1$.

На самом деле свойства 1) — 3) не только необходимы, но и достаточны для того, чтобы оператор Π был ортогональным проектором. Для доказательства предположим, согласно свойству 1, что Π — линейный оператор. Обозначим через L множество решений уравнения $\Pi x = x$. через N — множество решений уравнения $\Pi x = 0$. Легко убедиться, что L и N — линейные подпространства $\mathcal{R}_n$, причем ортогональные, если Π удовлетворяет условию 2). В самом деле, если $x \in L$, $y \in N$. то $(x, y) = (\Pi x, y) = (x, \Pi y) = (x, 0) = 0$. Для всякого вектора $x \in \mathcal{R}_n$ можно записать тождество

$$x = \Pi x + (I - \Pi)x. \tag{3.4}$$

Если Π удовлетворяет условию 3), то $\Pi(\Pi x) = \Pi x$, т. е. $\Pi x \in L$ и $\Pi(I - \Pi)x = (\Pi - \Pi^2)x = 0$, т. е. $(I - \Pi)x \in N$.

Следовательно, Π — оператор ортогонального проектирования на $L = \{x \in \mathcal{R}_n,\ \Pi x = x\}$. Из разложения (3.4) следует также, что оператор $I - \Pi$ ортогонально проецирует на $N = \{x \in \mathcal{R}_n\ (I - \Pi)x = x\} = L^{\perp}$.

Отметим следующее важное свойство ортогонального проектора. Пусть Π — оператор ортогонального проецирования на линейное подпространство L и

$$\rho(x, L) = \inf\{||x - y||\ |\ y \in L\} \tag{3.5}$$

— расстояние от x до L. Тогда

$$\rho(x, L) = ||x - \Pi x||. \tag{3.6}$$

Действительно, пусть $y \in L$. Тогда

$$\Pi x - y \in L,\ x - \Pi x = (I - \Pi)x \in L^{\perp}$$

и, следовательно,

$$||x - y||^2 = ||x - \Pi x + \Pi x - y||^2 = ||x - \Pi x||^2 + ||\Pi x - y||^2 \geqslant ||x - \Pi x||^2,$$

причем равенство здесь выполняется лишь в случае $\Pi x = y$.

Ортогональное преобразование.

Обозначим через U оператор ортогонального преобразования (оператор перехода от одного ортонормированного базиса к другому), $\widetilde{e}_i = Ue_i$. Он обладает свойствами: $U^{\mathsf{T}} = U^{-1}$ и $||Ux|| = ||x||$ $(||U^{\mathsf{T}}x|| = ||x||)$, $\det \widehat{U} = \pm 1$. (Здесь $\check{U}$ — матрица оператора U).

Теорема. Распределение вектора $\eta = U\xi$ совпадает с распределением вектора ξ (т.е. $\eta_i \sim \mathcal{N}(0, \sigma^2)$ и независимы).

Доказательство.

$$p_\eta(y) = p_\xi(U^{-1}y)|\det \widehat{U}^{-1}| = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}||U^{-1}y||^2\right\} =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{||y||^2}{2\sigma^2}\right\}. \quad \square$$

3.2 Распределения, связанные с нормальным.

Определение 1. Распределением χ_m^2 или Пирсона (E.S.Person) с m степенями свободы называется распределение сл. величины, равной сумме квадратов $\sum_{k=1}^{m} \xi_k^2$ независимых сл. величин $\xi_k \sim \mathcal{N}(0,1)$.

Свойства[1]. Плотность:

$$p_{\chi_m^2}(x) = \frac{1}{2^{m/2}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}.$$

Моменты: $\mathcal{M}\chi_m^2 = m$, $\mathsf{D}\chi_m^2 = 2m$.

Следствие 1. $||\Pi_m\xi||^2 = \sigma^2\chi_m^2$.

Доказательство. Пусть $\Pi_m x \in L_m \subset \mathcal{R}_n$. Рассмотрим новый базис $\{\widetilde{e}_i\}$, $i = 1, \dots, n$, удовлетворяющий условиям $\widetilde{e}_i \in L_m$, $i = 1, \dots, m$, $\widetilde{e}_i \in L_m^\perp$, $i = m+1, \dots, n$ и пусть U — ортогональный оператор перехода от старого базиса $\{e_i\}$ к новому $\{\widetilde{e}_i\}$, $\widetilde{e}_i = Ue_i$, $i = 1, \dots, n$. Тогда $||\Pi_m\xi||^2 = \sum_{i=1}^{m}(\xi, \widetilde{e}_i)^2 = \sigma^2\chi_m^2$, т.к. по теореме $(\xi, \widetilde{e}_i) \sim \mathcal{N}(0, \sigma^2)$. □

Определение 2. Пусть χ_m^2 и χ_k^2 и независимы. Тогда сл. величина $F_{m,k} = \frac{\frac{1}{m}\chi_m^2}{\frac{1}{k}\chi_k^2}$ контролируется распределением Фишера (R.A.Fischer).

Свойства. Плотность:

$$p_{F_{m,k}}(x) = \frac{k^{\frac{k}{2}} m^{\frac{k}{2}} x^{\frac{m}{2}-1}(kx+m)^{-\frac{k+m}{2}}}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})} \Gamma(\frac{k+m}{2}), \quad x > 0.$$

Моменты: $\mathcal{M}F_{k,m} = \frac{m}{m-2}$, $m > 2$ $\mathsf{D}F_{k,m} = \frac{2m^2}{(m-2)^2(m-4)}(1 + \frac{m-2}{k})$, $m > 4$.

Следствие 2. Пусть $\mathcal{R}_n = L_k \oplus L_m \oplus L_s$, $k + m + s = n$, Π_k — ортогональный проектор на L_k, Π_m — ортогональный проектор на L_m. Тогда $F_{k,m} = \frac{\frac{1}{k}||\Pi_k\xi||^2}{\frac{1}{m}||\Pi_m\xi||^2}$.

Доказательство. Пусть новый базис $\{\widetilde{e}_i\}$, $i = 1, \dots, n$ таков, что $\widetilde{e}_i \in L_k$, если $i = 1, \dots, k$, $\widetilde{e}_i \in L_m$, если $i = k+1, \dots, k+m$, $\widetilde{e}_i \in L_s$, если $i = k+m+1, \dots, n$. Тогда $||\Pi_k\xi||^2 = \sigma^2\chi_k^2$, $||\Pi_m\xi||^2 = \sigma^2\chi_m^2$ и независимы. □

Определение 3. Пусть $\xi \sim \mathcal{N}(0,1)$ и χ_m^2 независимы. Тогда сл. величина $t_m = \frac{\xi}{\sqrt{\frac{\chi_m^2}{m}}}$ контролируется распределением Стьюдента (В.С.Госсета) с m степенями свободы.

Свойства. Плотность:

$$p_{t_m}(x) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \frac{1}{\sqrt{\pi m}} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \ x > 0.$$

Моменты: $\mathcal{M}t_m = 0$, $\mathsf{D}t_m = \frac{m}{m-2}$.

Следствие 3. Обозначим $e = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$, L_1 — одномерное подпространство, определенное вектором e, Π_1 — ортогональный проектор на L_1, $\Pi_1\xi = (e,\xi)e = (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i)e = (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i, \dots, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i)$.

Пусть $\xi \sim \mathcal{N}(0, \sigma^2 I)$, тогда

$$\frac{(\xi, e)}{[\frac{1}{n-1}\|(I - \Pi_1)\xi\|^2]^{1/2}} = t_{n-1} \tag{3.7}$$

— распределение Стьюдента с $n-1$ степенью свободы.

Доказательство. Перейдем от базиса $\{e_i\}$ к базису $\{e, \widetilde{e}_1, \dots, \widetilde{e}_{n-1}\}$, так что $L = \mathcal{L}(e)$, $L^\perp = \mathcal{L}(\widetilde{e}_1, \dots, \widetilde{e}_{n-1})$. Тогда координаты в этом новом базисе независимы и распределены как $\mathcal{N}(0, \sigma^2)$, а $\|(I - \Pi_1)\xi\|^2 = \sigma^2\chi_{n-1}^2$ и по определению 3 получаем (3.7).

Перепишем (3.7) еще раз

[1]Вывод соответствующих формул см. в учебнике Пытьева, Шишмарева, (Изд. 2010 г.), стр. 276 – 280.

$$\frac{(\xi, e)}{[\frac{1}{n-1}\|(I-\Pi_1)\xi\|^2]^{1/2}} = \frac{\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\xi_i}{[\frac{1}{n-1}\sum\limits_{k=1}^{n}(\xi_k - \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i)^2]^{1/2}} =$$

$$= \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i}{[\frac{1}{n(n-1)}\sum\limits_{k=1}^{n}(\xi_k - \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i)^2]^{1/2}}. \quad \square$$

Если $\xi \sim \mathcal{N}(\overline{\mu}, \sigma^2 I)$, где $\overline{\mu} = (\mu, \ldots, \mu)$, то последнюю формулу можно заменить на

$$t_{n-1} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \mu)}{[\frac{1}{n(n-1)}\sum\limits_{k=1}^{n}(\xi_k - \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i)^2]^{1/2}} = \frac{\widehat{\mu} - \mu}{\sqrt{\frac{1}{n}\widehat{\sigma}^2}}, \tag{3.8}$$

где $\widehat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i, \quad \widehat{\sigma}^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(\xi_k - \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i)^2 = \frac{1}{n-1}\|(I-\Pi_1)\xi\|^2$.

3.3 Интервальные оценки нормального распределения.

Пусть $\{\xi_i\}$, $i = 1.2.\ldots$ — последовательность независимых нормально распределенных $\mathcal{N}(\mu, \sigma^2)$ сл. величин (независимых измерений). Требуется оценить значения неизвестных параметров μ и σ^2. Рассмотрим четыре случая.

1. Оценивание μ при известном σ^2.

Очевидно, $\frac{\sum_{i=1}^{n}(\xi_i-\mu)}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0,1)$.

Тогда

$$P\left(\left|\frac{\sum\limits_{i=1}^{n}(\xi_i - \mu)}{\sqrt{n\sigma^2}}\right| < \varepsilon\right) = \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon),$$

или, преобразуя неравенство, получаем

$$P\left(\widehat{\mu} - \varepsilon\sqrt{\frac{\sigma^2}{n}} < \mu < \widehat{\mu} + \varepsilon\sqrt{\frac{\sigma^2}{n}}\right) = \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon),$$

где $\widehat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ — середина интервала, шириной $2\varepsilon\sqrt{\frac{\sigma^2}{n}}$, которому с вероятностью $1 - \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon)$ принадлежит неизвестный параметр μ.

2. Оценивание σ^2 при известном μ. Из определения 1 следует, что $\frac{\sum_{i=1}^{n}(\xi_i-\mu)^2}{\sigma^2} \sim \chi^2_n$, поэтому

$$P\left(\varepsilon_1 < \frac{\sum\limits_{i=1}^{n}(\xi_i - \mu)^2}{\sigma^2} < \varepsilon_2\right) = 1 - \alpha(\varepsilon_1, \varepsilon_2),$$

или

$$P\left(\frac{\sum\limits_{i=1}^{n}(\xi_i - \mu)^2}{\varepsilon_2} < \sigma^2 < \frac{\sum\limits_{i=1}^{n}(\xi_i - \mu)^2}{\varepsilon_1}\right) = 1 - \alpha(\varepsilon_1, \varepsilon_2),$$

причем обычно ε_1 и ε_2 выбирают так, чтобы $P(\chi_n^2 < \varepsilon_1) = P(\chi_n^2 > \varepsilon_2)$. Интервал, которому удовлетворяет σ^2 с вероятностью $1-\alpha$, называется интервальной оценкой σ^2.

3. Оценивание μ при неизвестном σ^2.

Воспользуемся формулой (3.8)

$$P(|t_{n-1}| < \varepsilon) = P\left(\left|\frac{\widehat{\mu}-\mu}{\sqrt{\frac{1}{n}\widehat{\sigma}^2}}\right| < \varepsilon\right) = 1-\alpha(\varepsilon)$$

и получаем выражение, аналогичное пункту 1:

$$P\left(\widehat{\mu} - \varepsilon\sqrt{\frac{\widehat{\sigma}^2}{n}} < \mu < \widehat{\mu} + \varepsilon\sqrt{\frac{\widehat{\sigma}^2}{n}}\right) = 1-\alpha_{n-1}(\varepsilon),$$

но с тем отличием, что вместо σ^2 стоит $\widehat{\sigma}^2$ и $1-\alpha_{n-1}(\varepsilon)$ соответствует распределению Стьюдента с $n-1$ степенями свободы.

4. Оценивание σ^2 при неизвестном μ.

Здесь по аналогии с пунктом 2 получаем

$$P\left(\frac{\sum\limits_{i=1}^{n}(\xi_i-\widehat{\mu})^2}{\varepsilon_2} < \sigma^2 < \frac{\sum\limits_{i=1}^{n}(\xi_i-\widehat{\mu})^2}{\varepsilon_1}\right) = 1-\alpha_{n-1}(\varepsilon_1,\varepsilon_2),$$

где $\alpha_{n-1}(\varepsilon_1,\varepsilon_2)$ вычисляется по распределению χ^2_{n-1} с $n-1$ степенями свободы.

Построение доверительных интервалов в случае других распределений.

Пусть $F_\vartheta, \vartheta \in \Theta \subset \mathrm{R}^1$ — некоторое параметрическое семейство распределений, и $\xi_1,\ldots,\xi_n$ — выборка из распределения F_ϑ.

Пусть $\vartheta_n^- = \vartheta_n^-(\xi_1,\ldots,\xi_n)$ и $\vartheta_n^+ = \vartheta_n^+(\xi_1,\ldots,\xi_n)$ — некоторые статистики. Случайный интервал $(\vartheta_n^-, \vartheta_n^+)$ называется доверительным интервалом уровня $1-\varepsilon$, если

$$P_\vartheta\{\vartheta \in (\vartheta_n^-, \vartheta_n^+)\} \geqslant 1-\alpha.$$

Случайный интервал $(\vartheta_n^-, \vartheta_n^+)$ называется точным доверительным интервалом уровня $1-\varepsilon$, если при всех ϑ

$$P_\vartheta\{\vartheta \in (\vartheta_n^-, \vartheta_n^+)\} = 1-\alpha.$$

Для построения точного доверительного интервала обычно используется следующий подход. Выбирается функция $G(\xi_1,\ldots,\xi_n,\vartheta)$ такая, что распределение $P_\vartheta\{G(\xi_1,\ldots,\xi_n,\vartheta) \in (y^-,y^+)\}$ не зависит от параметра ϑ (распределение свободно от параметра ϑ). Функция G должна быть монотонной и обратимой функцией аргумента ϑ при любых фиксированных значениях выборки $\xi_1,\ldots,\xi_n$. Пусть, для определённости, функция G возрастает. Обозначим через $t(\xi_1,\ldots,\xi_n,y)$ функцию, обратную к функции $G(\xi_1,\ldots,\xi_n,\vartheta)$ по параметру ϑ. Тогда доверительный интервал уровня $1-\alpha$ имеет вид

$$(t(\xi_1,\ldots,\xi_n,y^-), t(\xi_1,\ldots,\xi_n,y^+)),$$

где числа y^- и y^+ находятся (вообще говоря, неоднозначно) из уравнения

$$P_\vartheta\{y^- < G(\xi_1,\ldots,\xi_n,\vartheta) < y^+)\} = 1-\alpha.$$

Пример.

Пусть $\xi_1,\ldots,\xi_n$ — выборка из равномерного распределения на отрезке $[0,\vartheta]$. С помощью статистики $\xi_{(n)}$ построить точный доверительный интервал уровня $1-\alpha$ для параметра ϑ.

Решение.

Пусть $\eta_i = \xi_i/\vartheta$, $i=1,...,n,$ — элементы выборки объёма n из равномерного распределения на отрезке $[0,1]$. Распределение случайной величины $\eta_{(n)} = \xi_{(n)}/\vartheta$ не зависит от ϑ. Найдем

$\varepsilon \in (0,1)$ такое, что $P\{\varepsilon < \eta_{(n)} < 1\} = 1-\alpha$. Функция распределения максимальной порядковой статистики $\eta_{(n)}$ равна $F(y) = y^n$ для $0 < y < 1$. Поэтому $1-\varepsilon^n = 1-\alpha$ и, соответственно, $\varepsilon = \sqrt[n]{\alpha}$.

Доверительный интервал для ϑ получим из соотношений

$$1-\alpha = P\{\varepsilon < \xi_{(n)}/\vartheta < 1\} = P\{\xi_{(n)} < \vartheta < \xi_{(n)}/\varepsilon\}.$$

Искомый доверительный интервал равен $(\xi_{(n)}, \xi_{(n)}/\sqrt[n]{\alpha})$.

4 Оценки максимального правдоподобия

Среди всех методов нахождения оценок выделяется своей простотой метод максимального правдоподобия, определяющий оценку $\widehat{\vartheta}(x)$ из условия

$$\sup_{\vartheta} L(x,\vartheta) = L(x,\widehat{\vartheta}) \text{ или } \frac{\partial \ln L(x,\vartheta)}{\partial\vartheta}|_{\vartheta=\widehat{\vartheta}} = 0.$$

Задача. Пусть дана выборка размера n из распределения $\mathcal{N}(\mu,\sigma^2)$. Найти оценку максимального правдоподобия $\widehat{\vartheta} = (\widehat{\mu}, \widehat{\sigma^2})$ и показать, что она не является несмещенной и эффективной.

Несмотря на отсутствие в этом принципе какой-либо оптимальности, а также на то, что такая оценка вовсе не обязана быть несмещенной и эффективной, она обладает хорошими асимптотическими свойствами.

Состоятельность оценки максимального правдоподобия.

Теорема.[1]

Пусть выполнены следующие условия регулярности (RR):

1) При каждом $\vartheta \in A$, где A — некоторый невырожденный интервал[2], для почти всех x существуют производные $\frac{\partial \ln f}{\partial\vartheta}$, $\frac{\partial^2 \ln f}{\partial\vartheta^2}$ и $\frac{\partial^3 \ln f}{\partial\vartheta^3}$.

2) При каждом $\vartheta \in A$ имеем $|\frac{\partial f}{\partial\vartheta}| < F_1(x)$, $|\frac{\partial^2 f}{\partial\vartheta^2}| < F_2(x)$ и $|\frac{\partial^3 \ln f}{\partial\vartheta^3}| < H(x)$,

где функции F_1 и F_2 интегрируемы на $(-\infty,\infty)$ и $\int\limits_{-\infty}^{\infty} H(x)f(x,\vartheta)dx < 2M$, причем M не зависит от ϑ.

3) При каждом $\vartheta \in A$ интеграл $\int\limits_{-\infty}^{\infty} \left(\frac{\partial \ln f}{\partial\vartheta}\right)^2 f(x,\vartheta)dx$ конечен и положителен.

Тогда уравнение правдоподобия

$$\frac{\partial \ln L}{\partial\vartheta} = 0 \tag{4.1}$$

имеет решение $\vartheta = \vartheta_n^*(x)$, сходящееся по вероятности при $n \to \infty$ к истинному значению параметра ϑ. Это решение является асимптотически нормальной и асимптотически R-эффективной оценкой для ϑ:

$$\mathcal{N}(\vartheta_0, \tfrac{1}{nk^2}), \text{ где } k^2 = \mathsf{E}\left(\tfrac{\partial \ln f}{\partial\vartheta}\right)^2_{\vartheta=\vartheta_0}.$$

Доказательство

Ограничимся случаем непрерывного распределения с плотностью вероятности $f(x,\vartheta)$.

1° Обозначим через ϑ_0 неизвестное истинное значение параметра ϑ в распределении, из которого производится выбор, и предположим, что ϑ_0 есть внутренняя точка интервала A. Покажем сначала, что уравнение правдоподобия имеет решение, сходящееся по вероятности к ϑ_0. Указывая индексом 0, что следует положить $\vartheta = \vartheta_0$ для каждого ϑ из A получаем

$$0 = \frac{1}{n}\frac{\partial \ln L}{\partial\vartheta} = B_0 + B_1(\vartheta-\vartheta_0) + \frac{1}{2}\alpha B_2(\vartheta-\vartheta_0)^2, \quad |\alpha| < 1, \tag{4.2}$$

где

$$B_0 = \frac{1}{n}\sum_i^n \left(\frac{\partial \ln f_i}{\partial\vartheta}\right)_0, \quad B_1 = \frac{1}{n}\sum_i^n \left(\frac{\partial^2 \ln f_i}{\partial\vartheta^2}\right)_0, \quad B_2 = \frac{1}{n}\sum_i^n H(x_i). \tag{4.3}$$

[1]См. Г.Крамер, Математические методы статистики, Мир, 1975. (Стр. 544). Доказательство принадлежит Дюге (D.Dugué).

[2]Строго говоря, множество A должно быть компактным.

Здесь f_i означает $f(x_i, \vartheta)$.

Величины B_m суть функции от случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, и остается показать, что с вероятностью, стремящейся к 1 при $n \to \infty$, уравнение (4.2) имеет корень, заключенный в пределах $\vartheta_0 \pm \delta$ где δ — сколь угодно малое положительное число.

Рассмотрим поведение величин B_m при больших значениях n. Из условий 1) и 2) следует что для каждого ϑ из A

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial \vartheta} dx = \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial \vartheta^2} dx = 0,$$

и поэтому

$$\mathsf{E}\left(\frac{\partial \ln f}{\partial \vartheta}\right)_0 = \int_{-\infty}^{\infty} \left(\frac{1}{f}\frac{\partial f}{\partial \vartheta}\right)_0 f(x, \vartheta_0) dx = 0,$$

$$\mathsf{E}\left(\frac{\partial^2 \ln f}{\partial \vartheta^2}\right)_0 = \int_{-\infty}^{\infty} \left[\frac{1}{f}\frac{\partial^2 f}{\partial \vartheta^2} - \left(\frac{1}{f}\frac{\partial f}{\partial \vartheta}\right)^2\right]_0 f(x, \vartheta_0) dx = \tag{4.4}$$

$$= -\mathsf{E}\left(\frac{\partial \ln f}{\partial \vartheta}\right)_0^2 = -k^2,$$

где, согласно условию 3), $k > 0$. Таким образом, в силу формулы (4.3), величина B_0 есть среднее арифметическое n независимых случайных величин, имеющих одно и то же распределение с нулевым средним значением. Из теоремы Хинчина[1]следует, что B_0 сходится по вероятности к нулю. Таким же образом убеждаемся, что B_1 $-k^2$, а B_2 сходится по вероятности к неотрицательному значению $\mathsf{E}H(\xi) < M$.

2° Пусть теперь δ и ε — фиксированные произвольно малые положительные числа, а $P(S)$ — совместная вероятностная, функция для случайных величин $x_1, x_2, \ldots, x_n$. При достаточно больших n, скажем, для всех $n > n_0 = n_0(\delta, \varepsilon)$, имеем

$$P_1 = P(|B_0| \geqslant \delta^2) < \frac{1}{3}\,\varepsilon,$$
$$P_2 = P(B_1 \geqslant -\frac{1}{2}k^2) < \frac{1}{3}\,\varepsilon,$$
$$P_3 = P(|B_2| \geqslant 2M) < \frac{1}{3}\,\varepsilon.$$

Пусть S — множество всех точек $x = (x_1, x_2, \ldots, x_n)$, для которых удовлетворяются все три неравенства

$$|B_0| < \delta^2, \quad B_1 < -\frac{1}{2}k^2, \quad |B_2| < 2M.$$

Дополнительное множество $\overline{S}$ состоит из всех точек x, для которых не выполняется хотя бы одно из этих трех неравенств, так что $P(\overline{S}) \leqslant P_1 + P_2 + P_3 < \varepsilon$ и поэтому $P(S) > 1 - \varepsilon$. Таким образом, вероятность попадания точки x в множество S, совпадающая с P-мерой множества S, превышает $1 - \varepsilon$, если $n > n_0(\delta, \varepsilon)$.

При $\vartheta = \vartheta_0 \pm \delta$ правая часть равенства (4.2) принимает значение $B_0 \pm B_1\delta + \frac{1}{2}B_2\delta^2$. В каждой точке x, принадлежащей S, сумма первого и третьего слагаемых в этом выражении по абсолютной величине меньше, чем $(M+1)\delta^2$, а $B_1\delta < -\frac{1}{2}k^2\delta$. Если $\delta < \frac{\frac{1}{2}k^2}{M+1}$, то знак всего выражения при $\vartheta = \vartheta_0 \pm \delta$ определяется вторым слагаемым, так что $\frac{\partial \ln L}{\partial \vartheta} > 0$ при $\vartheta = \vartheta_0 - \delta$ и $\frac{\partial \ln L}{\partial \vartheta} < 0$ при $\vartheta = \vartheta_0 + \delta$.

Далее, по условию 1), функция $\frac{\partial \ln L}{\partial \vartheta}$ для почти всех x есть непрерывная функция от ϑ из A. Таким образом, при произвольно малых δ и ε уравнение правдоподобия имеет с вероятностью, превышающей $1 - \varepsilon$, корень, заключенный в пределах $\vartheta \pm \delta$, если только $n > n_0(\delta, \varepsilon)$. Следовательно, первая часть доказательства закончена.

3° Пусть далее, $\vartheta^* = \vartheta^*(x_1, \ldots, x_n)$ есть решение уравнения правдоподобия, существование которого уже установлено. Учитывая (4.2) и (4.3), получаем для конечных значений n

$$k\sqrt{n}(\vartheta^* - \vartheta_0) = \frac{\frac{1}{k\sqrt{n}} \sum_1^n \left(\frac{\partial \ln f_i}{\partial \vartheta}\right)_0}{-\frac{B_1}{k^2} - \frac{\alpha}{2} B_2 \frac{\vartheta^* - \vartheta_0}{k^2}} \tag{4.5}$$

Из вышесказанного следует, что знаменатель дроби в правой части этого равенства сходится по вероятности к 1.

Далее, согласно формуле (4.4), $\left(\frac{\partial \ln f}{\partial \vartheta}\right)_0$ есть случайная величина со средним значением 0 и стандартным отклонением k. В силу теоремы Линдеберга-Леви, сумма $\sum_i^n \left(\frac{\partial \ln f_i}{\partial \vartheta}\right)_0$ асимптотически нормальна $\mathcal{N}(0, nk^2)$, и, следовательно, числитель дроби в правой части равенства (4.5) асимптотически нормален $\mathcal{N}(0, 1)$.

Наконец, из теоремы сходимости следует, что $k\sqrt{n}(\vartheta^* - \vartheta_0)$ — величина асимптотически нормальная $\mathcal{N}(0, 1)$, так что ϑ^* асимптотически нормальна $\mathcal{N}(\vartheta_0, \frac{1}{nk^2})$, где $k^2 = \mathsf{E}\left(\frac{\partial \ln f}{\partial \vartheta}\right)_0^2$.

---

[1]Пусть независимые случайные величины $\xi_1, \ldots, \xi_n$ одинаково распределены и имеют конечное среднее значение m. Тогда величина $\overline{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$ сходится по вероятности к m. Для доказательства используются свойства характеристических функций (сходимость к константе).

Можно ввести коэффициент эффективности (уровень эффективности), равный

$$e(\vartheta^*) = \frac{\min \mathsf{D}\,\vartheta^*}{\mathsf{D}\,\vartheta^*}.$$

Здесь числитель взят из неравенства Рао-Крамера.

Асимптотическая эффективность e_∞ оценки ϑ^* находится по формулам

$$e_n = \frac{1}{\mathsf{D}\vartheta_n^* \mathsf{E}\left(\frac{\partial \ln L_n}{\partial \vartheta}\right)^2}, \quad \mathsf{D}\vartheta_n^* \sim \frac{1}{nk^2}, \quad e_\infty = \frac{k^2}{\mathsf{E}\left(\frac{\partial \ln f}{\partial \vartheta}\right)_0^2} = 1,$$

так что наша теорема доказана. Соответствующая теорема дли дискретного распределения доказывается аналогичным образом. □

В случае нескольких неизвестных параметров следует ввести условия, служащие непосредственным обобщением условий 1) - 3). Тогда можно доказать, таким же способом, как и выше, используя многомерную форму теоремы Линдеберга-Леви, что уравнения правдоподобия имеют систему решений, являющихся асимптотически нормальными и совместно асимптотически эффективными оценками для параметров.

5 Достаточные статистики

Статистическая структура

Пусть $\mathcal{P} = \{P\}$ — некоторое семейство вероятностных мер (распределений) на измеримом пространстве $(X, \mathcal{A})$. Тройка $(X, \mathcal{A}, \mathcal{P})$ называется **статистической структурой**.

Замечания. 1. Если $\mathcal{P}$ состоит из одного элемента, то это вероятностное пространство.

2. Часто семейство $\mathcal{P}$ параметризуется $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$. Множество Θ — параметрическое пространство (пространство параметров). Естественно считать, что $\vartheta_1 \neq \vartheta_2 \Rightarrow P_{\vartheta_1} \neq P_{\vartheta_2}$.

3. В математической статистике пространство $(X, \mathcal{A})$ интерпретируется как пространство возможных наблюдений, или как выборочное пространство.

Говорят, что статистическая структура $(X, \mathcal{A}, \mathcal{P})$ доминируется σ-конечной мерой μ на $(X, \mathcal{A})$, если все меры $P_\vartheta \in \mathcal{P}$ абсолютно непрерывны относительно меры μ, т.е. для всякой меры $P_\vartheta \in \mathcal{P}$ существует функция (плотность P_ϑ по μ) $L(x|\vartheta)$ от $x \in X$, такая, что $P_\vartheta(A) = \int\limits_A L(x|\vartheta)\mu(dx)$ для всех $A \in \mathcal{A}$. В этом случае функция $X * \Theta \ni (x, \vartheta) \mapsto L(x|\vartheta) \in [0, \infty)$ называется функцией правдоподобия (likehood).

5.1 Методы нахождения эффективных оценок.

Статистика

Всякая измеримая функция от наблюдения называется статистикой. Более точно, если $(X, \mathcal{A}, \mathcal{P})$ — статистическая структура, то измеримое отображение T измеримого пространства $(X, \mathcal{A})$ в некоторое измеримое пространство $(Y, \mathcal{B})$ называется статистикой. (Она же случайная величина в вероятностном пространстве $(X, \mathcal{A}, P)$, $P \in \mathcal{P}$).

Две статистики $T_1 : X \to Y$ и $T_2 : X \to Y$ называются эквивалентными, если событие $A = \{T_1 \neq T_2\} \in \mathcal{A}$ является $\mathcal{P}$-пренебрежимым, т.е. $P(A) = 0$ для всех $P \in \mathcal{P}$.

Две статистики T_1 и T_2 на $(X, \mathcal{A}, \mathcal{P})$ называются независимыми, если для всякого $P \in \mathcal{P}$ независимы сл. величины T_1 и T_2, рассматриваемые в вероятностном пространстве $(X, \mathcal{A}, P)$.

Статистика $T(\xi)$ на $(X, \mathcal{A}, \mathcal{P})$ называется интегрируемой, если для всякого распределения P_ϑ, $\vartheta \in \Theta$ сл. в. $T(\xi)$, рассматриваемая в вероятностном пространстве $(X, \mathcal{A}, P_\vartheta)$, интегрируема. Математическое ожидание сл. в. T, соответствующее распределению P_ϑ, обозначается как $\mathsf{E}_\vartheta T$ или $\mathsf{E}_\vartheta T(\xi)$.

Достаточная статистика

Статистика $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ называется достаточной, если при заданном значении статистики T распределение наблюдения x не зависит от ϑ, т.е. более точно, для всякого $A \in \mathcal{A}$ величина $P_\vartheta\{x \in A | T(x)\}$ не зависит от ϑ (и значит, не несет никакой информации относительно ϑ)[1]. Следующая теорема дает удобный способ отыскания достаточной статистики.

[1]Или точнее: если существует вариант условных функций распределения при заданных T, который не зависит от ϑ.

Теорема факторизации (Неймана-Фишера).

Пусть L — функция правдоподобия. Статистика $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ является достаточной, если и только если существует $\mathcal{A}$-измеримая неотрицательная функция h на X и $\mathcal{B}$-измеримая неотрицательная функция g_ϑ на Y такие, что

$$L(x|\vartheta) = g_\vartheta(T(x))h(x) \tag{5.1}$$

для всех $\vartheta \in \Theta$, $x \in X$.

Доказательство. (Ограничимся дискретным случаем). $L(x|\vartheta) = P_\vartheta(\xi = x)$. Если T — достаточная статистика, $x \in X$ и $T(x) = t \in Y$, то

$$L(x|\vartheta) = P_\vartheta(\xi = x) = P_\vartheta(\xi = x, T(\xi) = t) =$$

$$= P_\vartheta(T(\xi) = t)P_\vartheta(\xi = x|T(\xi) = t) = g_\vartheta(t)h(x) = g_\vartheta(T(x))h(x),$$

где $g_\vartheta(t) = P_\vartheta(T(\xi) = t)$ а $h(x) = P_\vartheta(\xi = x|T(\xi) = t)$ не зависит от ϑ.

Если же функция правдоподобия L имеет вид $L(x|\vartheta) = g_\vartheta(T(x))h(x)$, то при $T(x) = t$ и $P_\vartheta(T(\xi) = t) > 0$ получаем

$$P_\vartheta(\xi = x|T(\xi) = t) = \frac{P_\vartheta(\xi = x, T(\xi) = t)}{P_\vartheta(T(\xi) = t)} = \frac{P_\vartheta(\xi = x)}{P_\vartheta(T(\xi) = t)} =$$

$$= \frac{P_\vartheta(\xi = x)}{\sum\limits_{y:T(y)=t} P_\vartheta(\xi = y)} = \frac{g_\vartheta(t)\ h(x)}{\sum\limits_{y:T(y)=t} g_\vartheta(t)h(y)} = \frac{h(x)}{\sum\limits_{y\in T^{-1}(t)} h(y)},$$

а последнее выражение, очевидно, не зависит от ϑ.

<u>Полная статистика</u>

Статистика T называется (ограниченно) **полной**, если для всякой (ограниченной) числовой статистики $f(T)$ из $\mathsf{E}_\vartheta f(T(\xi)) = 0$, $\forall\vartheta$ следует $f(T(\xi)) = 0$ $\mathcal{P}$-п.в. т.е.

$$P_\vartheta\{f(T(\xi)) = 0\} = 1,\ \forall\vartheta. \tag{5.2}$$

Для некоторых распределений полноту статистики можно доказать непосредственно (например, для равномерного, $U[0, \vartheta]$), для большинства распределений из так называемого экспоненциального семейства этот факт доказывается с помощью теоремы:

Теорема (*). Пусть статистическая структура $(X, \mathcal{A}, \mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\})$ допускает функцию правдоподобия вида

$$\frac{dP_\vartheta}{d\mu}(x) = L(x|\vartheta) = c(\vartheta)\exp\{\sum_{k=1}^{m}\gamma_k(\vartheta)T_k(x)\}$$

(экспоненциальное семейство распределений) и существует подмножество $\Theta_0 \subset \Theta$ такое, что образ отображения

$$\Theta_0 \ni \vartheta \mapsto \gamma(\vartheta) = \{\gamma_1(\vartheta), ..., \gamma_m(\vartheta)\} \in R^m$$

содержит хотя бы одну точку вместе с ее окрестностью и $c(\vartheta) \neq 0$, если ϑ принадлежит этой окрестности. *Тогда статистика* $T(x) = \{T_1(x), ..., T_m(x)\}$ *является полной (и достаточной).*

Доказательство. Можно считать, что $\gamma(\Theta_0)$ содержит "параллелепипед"

$$R = \{(\gamma_1, ..., \gamma_m) : -C < \gamma_k < C;\ k = 1, ..., m\},\ C > 0,$$

Пусть

$$\mathsf{E}_\vartheta f(T) = 0, \forall\vartheta \in \Theta_0 \tag{5.3}$$

для некоторой статистики $f(T)$. Положим $f(t) = f^+(t) - f^-(t)$, $f^\pm(t) \geqslant 0$, и обозначим через λ образ меры μ при отображении $T : (X, \mathcal{A}) \to (R^m, \mathbf{B}^m)$, т.е. $\lambda(B) = \mu(T^{-1}(B))$ для всех $B \in \mathbf{B}^m$. Тогда для всех $\gamma \in R$

$$\int e^{(\gamma,t)} f^+(t)\lambda(dt) = \int e^{(\gamma,t)} f^-(t)\lambda(dt),$$

где $(\gamma, t) = \sum\limits_{1}^{m}\gamma_k t_k$ и, в частности,

$$\int f^+(t)\lambda(dt) = \int f^-(t)\lambda(dt);$$

при этом последние интегралы без ограничения общности можно считать равными единице (умножив их, например, на $Const \neq 0$). Полагая

$$P^\pm(B) = \int\limits_B f^\pm(t)\lambda(dt),\ B \in \mathbf{B}^m,$$

имеем, что P^+ и P^- суть вероятностные меры на $(R^m, \mathbf{B}^m)$ и при этом

$$\int e^{(\gamma,t)} P^+(dt) = \int e^{(\gamma,t)} P^-(dt), \ \forall \gamma \in R.$$

Отсюда следует, что эти интегралы определены в полосе

$$\{\gamma = u + iv : u \in R^m, \ v \in R^m\}$$

m-мерной комплексной плоскости и, более того, в этой полосе являются аналитическими функциями от γ. По теореме об аналитическом продолжении эти интегралы совпадают в указанной полосе и, в частности,

$$\int\limits_{R^m} e^{i(v,t)} P^+(dt) = \int\limits_{R^m} e^{i(v,t)} P^-(dt), \ v \in R^m.$$

Последние интегралы представляют собой характеристические функции распределения P^+ и P^- соответственно, а из их совпадения следует $P^+ = P^-$ и, значит, $f^+(t) = f^-(t)$ п.в. по мере λ , т.е. $f(t) = 0$ λ-п.в. Таким образом, равенство (5.3) влечет (5.2). $\square$

Свободная статистика

Множество $A \in \mathcal{A}$ называется свободным (относительно семейства $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ вероятностных мер на $(X, \mathcal{A})$), если $P_\vartheta(A)$ не зависит от $\vartheta \in \Theta$. Статистика $U : (X, \mathcal{A}) \to (Z, \mathcal{C})$ называется свободной, если распределение этой статистики не зависит от $\vartheta \in \Theta$, т.е. $\{x : U(x) \in C\}$ есть свободное множество для всех $C \in \mathcal{C}$.

Замечание. Интуитивно ясно, что свободная статистика не несет никакой информации относительно истинного значения параметра ϑ. Напротив, достаточная статистика содержит в себе всю информацию (столько информации, сколько в самом наблюдении) относительно параметра ϑ. Часто достаточная (и ограниченно полная) статистика T и свободная статистика U дополняют друг друга в том смысле, что отображение $x \mapsto (T(x), U(x))$ биективно (и статистики T и U независимы как случайные величины).

Теорема Басу (D.Basu). Пусть $T(\xi)$ — достаточная ограниченно полная статистика и $U(\xi)$ — свободная статистика. Тогда статистики $T(\xi)$ и $U(\xi)$ независимы.

Доказательство. Так как статистика $T : (X, \mathcal{A}) \to (Y, \mathcal{B})$ достаточная, а статистика $U : (X, \mathcal{A}) \to (Z, \mathcal{C})$ свободная, то для всякого $C \in \mathcal{C}$)

$$P_\vartheta\{U(\xi) \in C | T(\xi)\} - P_\vartheta\{U(\xi) \in C\} = g(T(\xi))$$

есть свободная ограниченная статистика и

$$\mathsf{E}_\vartheta g(T(\xi)) = 0, \ \forall \vartheta,$$

откуда в силу ограниченной полноты статистики $T(\xi)$ имеем

$$P_\vartheta\{U(\xi) \in C | T(\xi)\} = P_\vartheta\{U(\xi) \in C\}, \ P_\vartheta - \text{п.в.}, \ \forall \vartheta.$$

Это влечет независимость сл.в. $T(\xi)$ и $U(\xi)$ в вероятностном пространстве $(X, \mathcal{A}, P_\vartheta)$ для всех ϑ, так как при $B \in \mathcal{B}$ имеем

$$\begin{aligned} P_\vartheta\{T(\xi) \in B, U(\xi) \in C\} &= \mathsf{E}_\vartheta \mathbf{I}_{\{T(\xi) \in B\}} P_\vartheta\{U(\xi) \in C | T(\xi)\} = \\ &= \mathsf{E}_\vartheta \mathbf{I}_{\{T(\xi) \in B\}} P_\vartheta\{U(\xi) \in C\} = P_\vartheta\{U(\xi) \in C\} \mathsf{E}_\vartheta \mathbf{I}_{\{T(\xi) \in B\}} = \\ &= P_\vartheta\{U(\xi) \in C\} P_\vartheta\{T(\xi) \in B\}. \quad \square \end{aligned}$$

Пример. Для $\mathcal{N}(\mu, \sigma^2)$ рассмотрим статистики

$T = (\overline{x}, s^2)$ и $U = (\frac{x_1 - \overline{x}}{s}, \dots, \frac{x_n - \overline{x}}{s})$, $\overline{x} = \frac{1}{n}\sum\limits_1^n x_i$, $s^2 = \sum\limits_1^n (x_i - \overline{x})^2$. U — свободная статистика и T и U независимы.

Статистика T_1 **подчинена**[1] T_2, если T_1 есть измеримая функция от T_2, $T_1 = \varphi(T_2)$.

Статистики $T = T(x)$ и $T' = T'(x)$ **эквивалентны**, если существуют измеримые функции $f(\cdot)$ и $g(\cdot)$, такие, что $T = f(T')$ и $T' = g(T)$ $\mathcal{P}$-п.в.

Пример. Для $\mathcal{N}(0, \sigma^2)$ рассмотрим статистики

$$\begin{array}{ll} T_1 = (x_1, x_2, \dots, x_n), & T_2 = (x_1^2, x_2^2, \dots, x_n^2), \\ T_3 = (x_1^2 + x_2^2 + \dots + x_m^2, x_{m+1}^2 + \dots + x_n^2), & T_4 = x_1^2 + x_2^2 + \dots + x_n^2. \end{array}$$

Эти статистики по-разному редуцируют данные. Больше всего — последняя (минимальная).

Достаточная статистика T называется минимальной, если она дает наибольшую редукцию данных среди всех достаточных статистик, т.е. если для любой достаточной статистики U существует функция h такая, что $T = h(U)$ ($\mathcal{P}$-п.в.).

[1] англ. ancillary — подчиненная, вспомогательная (А.А.Боровков, Математическая статистика).

Минимальные достаточные статистики существуют, если X — евклидово, а семейство $\mathcal{P}$ — доминируемо.

Всякая полная достаточная статистика является минимальной достаточной статистикой (но не наоборот!).

Утверждение. Если минимальная достаточная статистика существует, то для того, чтобы достаточная статистика была полной, необходимо, чтобы она была минимальной.

Доказательство. Предположим, что $T = h(S)$ есть минимальная достаточная статистика, а S — полная. Рассмотрим функцию $\psi(S) = S - \mathsf{E}_\vartheta(S|T)$ и пусть λ — распределение (мера) S. Тогда

$$\mathsf{E}_\vartheta \psi(S) = \int \psi(s)\lambda(ds) = 0,$$

откуда в силу полноты S следует $S = g(T) = \mathsf{E}_\vartheta(S|T) \pmod{\lambda}$. □

Пример построения минимальной достаточной статистики.

Утверждение. Пусть $\mathcal{P}$ — конечное семейство с плотностями $p_0(x), p_1(x), ..., p_k(x)$ с одним и тем же носителем. Тогда

$$T_1(x) = \Big(\frac{p_1(x)}{p_0(x)}, ..., \frac{p_k(x)}{p_0(x)}\Big)$$

— минимальная достаточная статистика.

Доказательство. Из теоремы о факторизации следует с очевидностью, что T — достаточная статистика, если $\forall\vartheta$, ϑ_0 отношение $\frac{p_\vartheta(x)}{p_{\vartheta_0}(x)}$ есть функция от T.

Но тогда $T_1(x)$ есть функция от T. □

Следствие. Утверждение легко обобщается на счетный случай.

Лемма. Если $\mathcal{P}$ — семейство с общим носителем и $\mathcal{P}_0 \subset \mathcal{P}$, T — минимальная достаточная статистика для $\mathcal{P}_0$ и достаточная для $\mathcal{P}$, то она же минимальная достаточная статистика для $\mathcal{P}$.

Доказательство. Если T_1 — достаточная статистика для $\mathcal{P}$, то она достаточна и для $\mathcal{P}_0$, откуда следует $T = f(T_1)$ □.

Пример. Рассмотрим $\mathcal{P} = \{\mathcal{N}(\vartheta, 1)\}$. Пусть $\mathcal{P}_0 = \{\mathcal{N}(\vartheta_0, 1), \mathcal{N}(\vartheta_1, 1)\}$ Тогда

$$\frac{p_{\vartheta_1}(x)}{p_{\vartheta_0}(x)} = e^{\frac{1}{2}\left(\sum(x_i-\vartheta_0)^2 - \sum(x_i-\vartheta_1)^2\right)},$$

так что $T = \sum x_i$ — есть достаточная минимальная статистика для всего класса $\mathcal{P}$.

Замечание. Минимальная достаточная статистика не обязана быть полной. Соответствующие примеры легко получаются, когда размерность статистики больше, чем размерность параметра. Например, совместная плотность минимальной достаточной статистики $S = (x_{(1)}, x_{(n)})$ для семейства $U(\vartheta, 1+\vartheta)$ равна

$$g_\vartheta(u, v) = \begin{cases} n(n-1)(v-u)^{n-2}, & \vartheta \leqslant u < v \leqslant \vartheta + 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Если взять функцию $\varphi(v-u)$ и сделать ортогональное преобразование $(v-u)/\sqrt{2} = t$, $(v+u)/\sqrt{2} = z$, то интеграл по треугольнику $\vartheta \leqslant u < v \leqslant \vartheta + 1$ будет равен

$$\int \varphi(v-u) g_\vartheta(u,v)dudv = n(n-1)\int_0^1 \varphi(x)x^{n-2}(1-x)dx, \quad (x = \sqrt{2}t).$$

Интеграл в правой части от ϑ не зависит и легко подобрать функцию $\varphi(x) \not\equiv 0$, обращающую его в нуль. Например, для $\varphi(x) = ax + b$ интеграл равен $a\frac{n-1}{n+1} + b$.

Выпуклые функции потерь

Пусть функция потерь[1] $\mathcal{L}(\vartheta, d)$ выпукла по d. Напомним, что функция $\varphi(x)$ на (a, b) называется выпуклой, если для любых $a < x < y < b$ и $0 < \gamma < 1$

$$\varphi(\gamma x + (1-\gamma)y) \leqslant \gamma\varphi(x) + (1-\gamma)\varphi(y).$$

Если неравенство строгое, то функция строго выпукла. Для дифференцируемых функций для выпуклости необходимо и достаточно

$$\varphi'(x) \leqslant \varphi'(y),\ a < x < y < b$$

или $\varphi''(x) \geqslant 0,\ a < x < b$.

Для выпуклых функций справедливо неравенство Иенсена:

$$\varphi(\mathsf{E}\xi) \leqslant \mathsf{E}\varphi(\xi). \tag{5.4}$$

[1] Последствия оценивания $\tau(\vartheta)$ величиной d.

Примеры: $1/\mathsf{E}\xi < \mathsf{E}(1/\xi)$, $\mathsf{E}(\log\xi) < \log(\mathsf{E}\xi)$.

Неравенство Иенсена. Пусть функция $\varphi(x)$, $-\infty < x < \infty$ выпукла, т. е. $\varphi(\alpha x + (1-\alpha)z) \leqslant \alpha\varphi(x) + (1-\alpha)\varphi(z)$, $0 \leqslant \alpha \leqslant 1$. Для $x < y < z$ можно записать очевидное неравенство (при $y = \alpha x + (1-\alpha)z$):

$$\frac{\varphi(y)-\varphi(x)}{y-x} \leqslant \frac{\varphi(z)-\varphi(y)}{z-y}, \tag{5.5}$$

откуда следует

$$\overline{\lim_{x:x<y}} \frac{\varphi(y)-\varphi(x)}{y-x} \leqslant c_1(y) \leqslant c_2(y) \leqslant \underline{\lim_{z:y<z}} \frac{\varphi(z)-\varphi(y)}{z-y}. \tag{5.6}$$

Выбирая $c_1(y) \leqslant c(y) \leqslant c_2(y)$[1], имеем $(x-y)c(y) + \varphi(y) \leqslant \varphi(x)$, $-\infty < x < \infty$. Другими словами: для всякой точки x_0 существует число $c = c(x_0)$, такое, что $\varphi(x) \geqslant \varphi(x_0) + c(x - x_0)$ для всех x. Полагая теперь $x = \xi$, $x_0 = \mathsf{E}\xi$ и применяя E к получившемуся неравенству, получаем неравенство (5.4).

Теорема (Рао-Блекуэлла). Пусть ξ — вектор наблюдений из распределения $P_\vartheta \in \mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ и пусть статистика T достаточна для $\mathcal{P}$. Пусть δ есть некоторая оценка для $\tau(\vartheta)$ и пусть функция потерь $\mathcal{L}(\vartheta, d)$ строго выпукла по d. Тогда если δ имеет конечные математическое ожидание и риск $R(\vartheta,\delta) = \mathsf{E}L(\vartheta,\delta(\xi)) < \infty$ и если $\eta(T) = \mathsf{E}(\delta(\xi)|T)$, то для риска оценки $\eta(T)$ справедливо неравенство

$$R(\vartheta,\eta) < R(\vartheta,\delta), \tag{5.7}$$

если только с вероятностью 1 не выполняется равенство $\delta(\xi) = \eta(T)$.

Доказательство. Положим в неравенстве Иенсена $\varphi(d) = \mathcal{L}(\vartheta, d)$, $\delta = \delta(\xi)$, причем в нем используется $P_\vartheta^{\xi|t}(x|t)$ — условное распределение ξ при $T = t$. Тогда

$$\mathcal{L}(\vartheta,\eta(t)) < \mathsf{E}(\mathcal{L}(\vartheta,\delta(\xi))|t),$$

если только с вероятностью 1 не выполняется $\delta(\xi) = \eta(T)$.

Взяв математические ожидания от обеих сторон этого неравенства, получаем (5.7) □.

Замечание. Если $\mathcal{L}(\vartheta,\delta(\xi)) = (\tau(\vartheta) - \delta(\xi))^2$, то в этом случае получаем, что дисперсия оценки η не превосходит дисперсии оценки δ, а повторное усреднение уже не улучшает оценку.

Особенно сильные результаты дает этот метод в случае, когда $T(\xi)$ — полная достаточная статистика.

Теорема (А.Н.Колмогоров). Пусть $T(\xi)$ — полная достаточная статистика. Тогда $\eta(T)$ оптимально оценивает $\tau(\vartheta)$ тогда и только тогда, когда

$$\mathsf{E}_\vartheta \eta(T(\xi)) = \tau(\vartheta), \tag{5.8}$$

т. е. $\eta(T)$ — несмещенная оценка.

Доказательство. Пусть $T_1(\xi) = \varphi_1(T(\xi))$ — несмещенная оценка $\tau(\vartheta)$ и $T(\xi)$ — полная достаточная статистика. Предположим, что существует другая несмещенная оценка $\varphi_2(T(\xi))$. В этом случае для любого ϑ

$$E_\vartheta(\varphi_1(T(\xi)) - \varphi_2(T(\xi))) = 0 \quad \Longrightarrow \quad \varphi_1(\xi) - \varphi_2(\xi) = 0$$

с вероятностью 1, откуда следует единственность оценки. Равномерная минимальность риска (дисперсии) следует из теоремы Рао-Блекуэлла. □

Вывод: указанная процедура вычисления условного математического ожидания любую несмещенную оценку превращает в н.о.м.д.

Пример.

Дана независимая выборка объема n из экспоненциального распределения ($\lambda e^{-\lambda x}$, $x > 0$). Найти н.о.м.д. для значения функции распределения $F(a)$. Поскольку $L(x,\lambda) = \lambda^n e^{-\lambda \sum x_i}$, очевидно, что достаточная статистика $T(x) = \sum x_i$. Распределение величины $T(\xi)$ определяется[2] плотностью вероятности $p_T(s) = \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!}$.

[1] одно или оба неравенства могут быть строгими.

[2] Эту формулу можно получить с помощью характеристической функции суммы независимых сл. величин

Выберем оценку величины $F(a)$ в виде

$$t(\xi) = t(\xi_1) = \begin{cases} 1, & \xi_1 < a \\ 0, & \xi_1 > a \end{cases} \qquad \mathsf{E}_\vartheta t(\xi_1) = P\{\xi_1 < a\} = F(a). \tag{5.9}$$

Теперь найдем условную плотность вероятности

$$p_{\xi_1|T-\xi_1}(y|s-y) = \frac{p_{\xi_1}(y)p_{T-\xi_1}(s-y)}{p_T(s)} = \frac{\lambda e^{-\lambda y}\lambda^n(s-y)^{n-2}e^{-\lambda(s-y)}(n-1)!}{\lambda^n s^{n-1}e^{-\lambda s}(n-2)!} =$$
$$= (n-1)\frac{(1-y/s)^{n-2}}{s}, \quad 0 \leqslant y \leqslant s.$$

Наконец, искомая оценка есть условное математическое ожидание $t(\xi_1)$:

$$t_1(T) = \int\limits_0^{\min(a,T)} (n-1)(1-y/s)^{(n-2)}\frac{dy}{s}\Big|_{s=T} = 1 - \left(1 - \frac{\min(a,T)}{T}\right)^{(n-1)}.$$

Несмещенность.

Можно попытаться облегчить нахождение оптимальных оценок путем сужения их класса. Одним из условий беспристрастности, применимым к оценкам, является условие несмещенности оценок[1]:

$$\mathsf{E}_\vartheta\delta(\xi) = \tau(\vartheta), \ \forall\vartheta \in \Theta.$$

Не для всех функций $\tau(\cdot)$ такие оценки существуют. Например, для биномиального распределения несмещенной оценки для $g(p) = \frac{1}{p}$ не существует, так как должно было бы быть

$$\sum_{k=0}^{n}\delta(k)C_n^k p^k q^{n-k} = \frac{1}{p},$$

что не выполняется при $p \to 0$.

Если существует несмещенная оценка для $\tau(\cdot)$, такая функция называется *допускающей несмещенную оценку* (ДНО).

Лемма. Пусть ξ имеет распределение из семейства $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ и пусть T — полная достаточная статистика для $\mathcal{P}$.

Тогда каждая ДНО-функция имеет, и при том только одну ($\mathcal{P}$ -п.н.), несмещенную оценку, которая является функцией от T.

Доказательство. Существование следует из теоремы Рао-Блэкуэлла, так как любая несмещенная оценка, не являющаяся функцией от T, улучшается ее условным математическим ожиданием $\eta(T) = \mathsf{E}\{\delta(\xi)|T\}$ при фиксированном T и она остается несмещенной. Единственность следует из того, что в противном случае разность $f(T) = \eta_1(T) - \eta_2(T)$ равна нулю $\mathcal{P}$-п.н. в силу полноты T.

Теорема. Пусть ξ имеет распределение из семейства $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ и пусть T — полная достаточная статистика для $\mathcal{P}$.

Тогда

1. Для каждой ДНО-функции существует несмещенная оценка, которая равномерно минимизирует риск для любой функции потерь $\mathcal{L}(\vartheta, d)$, выпуклой относительно d, в частности, эта оценка является НРМД (или НОМД);

2. вышеуказанная оценка — единственная несмещенная оценка, которая является функцией от T; это единственная несмещенная оценка с минимальным риском при условии, что риск конечен и $\mathcal{L}(\vartheta, d)$ строго выпукла по d.

3. Вычисление условного математического ожидания.

[1]Средние **погрешности** недо- и переоценки равны. Бывает еще так называемая медианная несмещенность — **частоты** недо- и переоценки равны: $P_\vartheta(\delta(X) < \tau(\vartheta)) = P_\vartheta(\delta(X) > \tau(\vartheta)), \forall\vartheta$.

НРМД оценка может быть получена как условное математическое ожидание $\mathsf{E}[\delta(\xi)|T]$, где $\delta(\xi)$ — любая несмещенная оценка для $\tau(\vartheta)$ (ее можно выбрать так, чтобы сделать вычисление $\mathsf{E}[\delta(\xi)|T]$ как можно проще).

Пример. Распределение $U(0,\vartheta)$. $\xi_{(n)} = \max\{\xi_1,...,\xi_n\}$ — полная достаточная статистика. Так как $\mathsf{E}\xi_1 = \vartheta/2$, то $\mathsf{E}[\xi_1|\xi_{(n)} = t]$ будет НРМД оценкой для $\vartheta/2$. Если $\xi_{(n)} = t$, то $P\{\xi_1 = t\} = 1/n$ и с вероятностью $(n-1)/n$ величина ξ_1 равномерно распределена на $(0,t)$[1].

Следовательно,

$$\mathsf{E}[\xi_1|t] = \frac{1}{n}t + \frac{n-1}{n}\frac{t}{2} = \frac{n+1}{n}\frac{t}{2}.$$

Таким образом, $[(n+1)/n]T/2$ и $[(n+1)/n]T$ суть НРМД оценки для $\vartheta/2$ и ϑ соответственно.

Если снять условие выпуклости функции потерь, например, $\mathcal{L}(\vartheta,d) \leqslant M$, то можно придумать такую несмещенную оценку $\delta(\xi)$, что $R(\vartheta_0,\delta_n) \to 0$ для произвольного значения $\vartheta = \vartheta_0$.

Теорема. Пусть функция потерь $\mathcal{L}(\vartheta,d)$ для оценивания $\tau(\vartheta)$ ограничена, $\mathcal{L}(\vartheta,d) \leqslant M$ и пусть $\mathcal{L}(\vartheta,\tau(\vartheta)) = 0$ для всех ϑ. Допустим также, что $\tau(\vartheta)$ — ДНО-функция и пусть ϑ_0 — произвольное значение ϑ. Тогда существует последовательность несмещенных оценок δ_n, для которых $R(\vartheta_0,\delta_n) \to 0$.

Доказательство. Поскольку $\tau(\vartheta)$ — ДНО-функция, существует некоторая несмещенная оценка $\delta(\xi)$. Для любого $0 < \pi < 1$ пусть

$$\delta'_\pi(\xi) = \begin{cases} \tau(\vartheta_0) & \text{с вероятностью } 1-\pi, \\ \frac{1}{\pi}[\delta(\xi) - \tau(\vartheta_0)] + \tau(\vartheta_0) & \text{с вероятностью } \pi. \end{cases}$$

Тогда $\delta'_\pi(\xi)$ — несмещенная оценка для всех π и всех ϑ, так как

$$\mathsf{E}_\vartheta \delta'_\pi(\xi) = (1-\pi)\tau(\vartheta_0) + \frac{\pi}{\pi}[\tau(\vartheta) - \tau(\vartheta_0)] + \pi\tau(\vartheta_0) = \tau(\vartheta).$$

При $\vartheta = \vartheta_0$ риск $R(\vartheta_0,\delta'_\pi)$ равен сумме $(1-\pi)\times 0$ и умноженной на π ожидаемой потери от оценки $\frac{1}{\pi}[\delta(\xi) - \tau(\vartheta_0)] + \tau(\vartheta_0)$, так что

$$R(\vartheta_0,\delta'_\pi) \leqslant \pi M.$$

Если $\pi \to 0$, то $R(\vartheta_0,\delta'_\pi) \to 0$. □.

(Басу доказал этот факт для более общего случая невыпуклых функций потерь.)

Этот результат показывает, что за исключением тривиальных случаев для ограниченных функций потерь не существует несмещенных оценок не только с равномерно минимальным риском, но и с локально минимальным риском, поскольку при каждом ϑ_0 риск может быть сделан произвольно малым даже для несмещенных оценок. Затруднение, связанное с невыпуклыми функциями потерь, возникает из-за возможности сколь угодно больших ошибок, так как при $\pi \to 0$ ошибка $|\delta'_\pi - \tau(\vartheta_0)| \to 0$.

Для больших объемов выборки начинает влиять локальное поведение функции потерь вблизи истинного значения $\tau(\vartheta)$, поэтому процедура минимизации функции риска равносильна минимизации $\mathsf{E}[\delta(\xi) - \tau(\vartheta)]^2$.

Замечание. Такого рода трудностей не возникает при использовании медианной несмещенности.

5.2 Методы нахождения НРМД оценок

Можно пользоваться рассмотренным ранее теоремами в несколько иных формулировках. Теорема Рао-Блэкуэлла: "Оптимальная оценка, если она существует, является функцией от достаточной статистики" и Колмогорова: "Если существует полная достаточная статистика, то всякая функция от нее является оптимальной оценкой своего математического ожидания".

1. Решение уравнений для оценки δ. Если T — полная достаточная статистика, то НРМД оценка любой ДНО-функции $\tau(\vartheta)$ однозначно определяется совокупностью уравнений для всех $\vartheta \in \Theta$.

Примеры.

а). Для распределения Бернулли ($P\{\xi = 1\} = p$, $P\{\xi = 0\} = 1-p$) легко показать, что $T(\xi) = \sum\limits_{i=1}^{n} \xi_i$ является полной достаточной статистикой. Поэтому любую оценку функции $\tau(p)$ следует искать в виде $\delta(T)$. Пусть $\tau(p) = pq = p(1-p)$. Поскольку T имеет биномиальное распределение, уравнение для нахождения $\delta(T)$ запишем в виде (условие несмещенности)

$$\mathsf{E}\delta(T) = \sum_{t=0}^{n} \delta(t) C_n^t p^t q^{n-t} = pq$$

[1]Это следует из того, что по условию, n-мерная случайная величина ξ имеет равномерное распределение на n-мерном кубе $Q(\vartheta) = \{x = (x_1,...,x_n) : 0 < x_1 < \vartheta, ..., 0 < x_n < \vartheta\}$. Для $0 < t < \vartheta$ множество уровня $T(\xi) = t$ — это та часть поверхности куба $Q(t)$, что лежит в положительном октанте. Мера $s(\cdot)$ на этой поверхности — это обычная $(n-1)$-мерная мера Лебега. Отсюда условная плотность ξ при данном $= t$ постоянна на поверхности уровня $T(\xi_1,...,\xi_n) = t$. Поэтому условное распределение ξ при данном T — равномерное (на указанной поверхности).

для всех $0 < p < 1$. Обозначим $\rho = p/q$, тогда $q = (1+\rho)^{-1}$ и

$$\sum_{t=0}^{n} \delta(t) C_n^t \, \rho^t = \rho\,(1+\rho)^{n-2} = \sum_{t=1}^{n-1} C_{n-2}^{t-1} \, \rho^t, \ 0 < \rho < \infty.$$

Сравнивая коэффициенты при степенях ρ, получаем $\delta(t) = \frac{t(n-t)}{n(n-1)}$.

б). Рассмотрим распределение Пуассона с параметром ϑ и оценим функцию $\tau(\vartheta) = \vartheta^2$. Функция $T(\xi) = \sum\limits_{i=1}^{n} \xi_i$ — полная достаточная статистика. Рассмотрим статистику $T_1 = C(n)T(T-1)$ и найдем неслучайную функцию $C(n)$. Из условия несмещенности получаем

$$\begin{gathered}\mathsf{E}_\vartheta C(n)T(T-1) = C(n)(\mathsf{E}_\vartheta T^2 - \mathsf{E}_\vartheta T) = \\ = C(n)(n^2\vartheta^2 + n\vartheta - n\vartheta) = n^2 C(n)\vartheta^2 = \vartheta^2,\end{gathered}$$

Таким образом, при $C(n) = n^{-2}$ наша оценка оптимальная.

2. Метод моментов (более старый и простой метод).

Вычисляются моменты как функции параметров, точные моменты заменяются выборочными, затем решается система уравнений относительно параметров.

Пример. Рассмотрим распределение $U(\vartheta_1, \vartheta_2)$.

$$\begin{gathered}\mathsf{E}_\vartheta \xi_1 = \frac{\vartheta_1 + \vartheta_2}{2} = A_1 = \frac{1}{n}\sum \xi_i = \overline{\xi}, \\ \mathsf{E}_\vartheta \xi_1^2 = \frac{\vartheta_1^2 + \vartheta_1\vartheta_2 + \vartheta_2^2}{3} = A_2 = \frac{1}{n}\sum \xi_i^2 = \overline{\xi^2}.\end{gathered}$$

Эта система эквивалентна следующей

$$\begin{gathered}\vartheta_1 + \vartheta_2 = 2\overline{\xi}, \\ \vartheta_1\vartheta_2 = 4\,\overline{\xi}^{\,2} - 3\overline{\xi^2}.\end{gathered}$$

ϑ_1 и ϑ_2 являются корнями уравнения $t^2 - 2\overline{\xi}t + 4\,\overline{\xi}^{\,2} - 3\overline{\xi^2} = 0$ и

$$\begin{gathered}\widehat{\vartheta}_1 = \overline{\xi} - \sqrt{3(\overline{\xi^2} - \overline{\xi}^{\,2})} = \overline{\xi} - \sqrt{3S^2}, \\ \widehat{\vartheta}_2 = \overline{\xi} + \sqrt{3(\overline{\xi^2} - \overline{\xi}^{\,2})} = \overline{\xi} + \sqrt{3S^2},\end{gathered}$$

где $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi})^2$.

6 Линейное оценивание.

6.1 Теорема Гаусса-Маркова.

Предположим, что наблюдению доступны лишь линейные комбинации неизвестных величин (наблюдения *косвенные*)

$$\xi_i = \sum_{j=1}^{k} a_{ij}\alpha_j + \nu_i, \ i = 1, 2, \ldots, n. \tag{6.1}$$

Пусть ν_i — независимые сл. величины, $\mathsf{E}\mu = 0$, $\mathsf{D}\nu = \sigma^2$, $i = 1, 2, \ldots, n$.

Требуется оценить α_j, $j = 1, 2, \ldots, k$, точнее, найти линейные (**1**) несмещенные (**2**) оценки $\widehat{\alpha}_j$ с минимальной дисперсией (**3**).

Запишем равенство (6.1) в виде

$$\xi = \sum_{j=1}^{k} a_j \alpha_j + \nu, \tag{6.2}$$

где $a_j = (a_{1j}, a_{2j}, \ldots, a_{nj})^{\mathsf{T}}$, $\xi, \nu, a_j \in \mathcal{R}_n$, $n \geqslant k$, векторы-столбцы a_j линейно независимы, или в виде

$$\xi = A\alpha + \nu, \quad \alpha \in \mathcal{R}_k, \tag{6.3}$$

причем $\mathsf{E}\nu = 0$, $\mathsf{E}\nu\nu^{\mathsf{T}} = \sigma^2 I$.

1. (Линейность) Будем искать оценку $\widehat{\alpha}_j$ в виде $\widehat{\alpha}_j = \sum\limits_{i=1}^{n} b_{ji}\xi_i = (b_j, \xi)$, $b_j = (b_{j1}, b_{j1}, \ldots, b_{jn})^{\mathsf{T}}$.

2. Требование несмещенности дает:

$$\mathsf{E}\widehat{\alpha}_j = \sum_{i=1}^{n} b_{ji} \sum_{s=1}^{k} a_{is}\alpha_s = \sum_{s=1}^{k} \left(\sum_{i=1}^{n} b_{ji} a_{is} \right) \alpha_s = \alpha_j. \tag{6.4}$$

Отсюда $(b_j, a_s) = \delta_{js}$, $j, s = 1, 2, \ldots, k$.

3. Вычислим дисперсию $\mathsf{D}\widehat{\alpha}_j = \mathsf{D} \sum\limits_{i=1}^{n} b_{ji}\xi_i = \sigma^2 \sum\limits_{i=1}^{n} b_{ji}^2 = \sigma^2 ||b_j||^2$.

Требование минимальности дисперсии приводит к следующей задаче на условный экстремум:

Для каждого j найти $\min ||b_j||^2$ при условии $(b_j, a_s) = \delta_{js}$, $j, s = 1, 2, \ldots, k$. Воспользуемся методом множителей Лагранжа. Напомним, что для нахождения минимума $\varphi(x)$ при условиях $g_i(x) = 0$, $i = 1, 2, \ldots, m$, нужно, чтобы градиент $\mathrm{grad}\varphi(x)$ был ортогонален всем поверхностям $g_i(x) = 0$, $i = 1, 2, \ldots, m$, т.е. градиент $\mathrm{grad}\varphi(x)$ может быть разложен по векторам $\mathrm{grad} g_i(x))$, $i = 1, 2, \ldots, m$: $\mathrm{grad} \left(\varphi(x) - \sum\limits_{i=1}^{m} \lambda_i g_i(x) \right) = 0$ при некоторых λ_i. Выражение в скобках — так называемая функция Лагранжа. Введем функцию Лагранжа

$$\mathcal{L} = ||b_j||^2 - 2 \sum_{s=1}^{k} \lambda_{js} (b_j, a_s) \tag{6.5}$$

и, дифференцируя по b_{ji}, получаем $2b_{ji} - 2 \sum\limits_{s=1}^{k} \lambda_{js} a_{si}$ или $b_j = \sum\limits_{s=1}^{k} \lambda_{js} a_s$. Используем условие несмещенности: $(b_j, a_p) = \sum\limits_{s=1}^{k} \lambda_{js} (a_s, a_p) = \delta_{jp}$.

Отсюда $\lambda_{js} = (a_j, a_s)^-$ и окончательно[1]

$$\widehat{\alpha}_j = \sum_{s=1}^{k} (a_j, a_s)^- (a_s, \xi).$$

Поскольку в векторно-матричной форме $(a_j, a_s) = (A^{\mathsf{T}} A)_{js}$, то

$$\widehat{\alpha} = (A^* A)^{-1} A^{\mathsf{T}} \xi. \tag{6.6}$$

Найдем матрицу ковариаций $\widehat{\alpha}$. Поскольку

$$\widehat{\alpha} - \alpha = (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} \xi - \alpha = (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} (A\alpha + \nu) - \alpha =$$
$$= (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} A\alpha + (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} \nu - \alpha = (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} \nu,$$

то

$$\mathsf{E}(\widehat{\alpha} - \alpha)(\widehat{\alpha} - \alpha)^{\mathsf{T}} = \mathsf{E}(A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}} \nu\nu^{\mathsf{T}} A (A^{\mathsf{T}} A)^{-1} = \sigma^2 (A^{\mathsf{T}} A)^{-1}. \tag{6.7}$$

[1]Знак $^-$ говорит о том, что берется соответствующий элемент матрицы, обратной к матрице $||(a_j, a_s)||$.

Рассмотрим *метод наименьших квадратов*. Пусть $\widetilde{\alpha}$ выбираются из условия[1]

$$\sum_{i=1}^{n}(\xi_i - \sum_{j=1}^{k} a_{ij}\alpha_j)^2 \sim \min_{\alpha_j}.$$

Дифференцируя по α_s, получим

$$2\sum_{i=1}^{n}(\xi_i - \sum_{j=1}^{k} a_{ij}\widetilde{\alpha}_j)a_{is} = 0 \Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{k} a_{ij}a_{is}\widetilde{\alpha}_j = \sum_{i=1}^{n} a_{is}\xi_i. \tag{6.8}$$

Отсюда получаем

$$\widetilde{\alpha} = (A^*A)^{-1}A^*\xi,$$

т.е. ту же оценку,что и $\widehat{\alpha}$ ($\widetilde{\alpha} = \widehat{\alpha}$).

Таким образом, справедлива **Теорема Гаусса-Маркова**:

Пусть ξ измеряется по схеме (6.1). Тогда ЛНОМД дается формулой

Как оценить σ^2?

Заметим, что из (6.8) следует $(\xi - A\widehat{\alpha})a_s = 0$, $s = 1, 2, \ldots, k$, т.е.

$$(I - A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}})\xi \perp L(a_1, \ldots, a_k) = (I - \Pi_a)\xi.$$

Таким образом, $\Pi_a = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ — отогональный проектор на $L(a_1, \ldots, a_k)$ — линейную оболочку векторов a_i (это можно проверить непосредственно).

Пусть $k < n$. Обозначим

$$\begin{aligned}
&s^2 = ||\xi - A\alpha||^2 = ||\nu||^2,\\
&s_1^2 = ||\xi - \Pi_a\xi||^2 = ||\xi - \Pi_a(A\alpha + \nu)||^2 = ||(I - \Pi_a)\nu||^2,\\
&s_2^2 = ||\Pi_a\xi - A\alpha||^2 = ||A(\widehat{\alpha} - \alpha)||^2 = ||\Pi_a(\xi - A\alpha)||^2 = ||\Pi_a\nu||^2.
\end{aligned}$$

Далее, $\mathsf{E}s^2 = \mathrm{tr}\sigma^2 I = n\sigma^2$, $\mathsf{E}s_1^2 = \sigma^2\mathrm{tr}(I - \Pi_a) = \sigma^2(n-k)$.

Отсюда $\widehat{\sigma^2} = \frac{1}{n-k}s_1^2 = \frac{1}{n-k}||\xi - \Pi_a\xi||^2$ — несмещеная оценка σ^2.

Доверительные множества в нормальной регрессии.

Доверительные множества — аналог интервалов в интервальных оценках.

Пусть $\nu \sim \mathcal{N}(0, \sigma^2 I)$. Тогда $s_2^2 = ||\Pi_a\xi - A\alpha||^2 = ||\Pi_a\nu||^2 = \sigma^2\chi_k^2$, $s_1^2 = ||\xi - \Pi_a\xi||^2 = ||(I - \Pi_a)\nu||^2 = \sigma^2\chi_{n-k}^2$ и независимы, поэтому

$$\frac{\frac{1}{k}\,s_2^2}{\frac{1}{n-k}s_1^2} = \frac{\frac{1}{k}\chi_k^2}{\frac{1}{n-k}\chi_{n-k}^2} = F_{k,n-k}.$$

($F_{k,n-k}$ —распределение Снедекора-Фишера).

Пусть $P\{F_{k,n-k} \leqslant \varepsilon\} = \gamma_F(\varepsilon)$, тогда с вероятностью $\gamma_F(\varepsilon)$

$$||A(\alpha - \widehat{\alpha})||^2 = (A^{\mathsf{T}}A(\alpha - \widehat{\alpha}), (\alpha - \widehat{\alpha})) \leqslant \varepsilon\frac{k}{n-k}||(I - \Pi_a)\xi||^2. \tag{6.9}$$

Левая часть неравенства (6.9) представляет собой квадратичную форму относительно координат α с матрицей $A^{\mathsf{T}}A > 0$, поэтому (6.9) определяет в координатах α_j эллипсоид с центром $\widehat{\alpha}$ (доверительный эллипсоид Хотеллинга).

Если нам нужно оценить одну координату α_j, то вспомним, что ее дисперсия равна $\sigma^2(a_j, a_j)^-$, поэтому $\frac{\alpha_j - \widehat{\alpha}_j}{\sqrt{\sigma^2(a_j,a_j)^-}} \sim \mathcal{N}(0,1)$, а

$$\frac{\alpha_j - \widehat{\alpha}_j}{\sqrt{(a_j, a_j)^- \frac{1}{n-k}||(I - \Pi_a)\xi||^2}} = t_{n-k},$$

[1] Здесь не делается никаких предположений о ξ_i, $i = 1, 2, \ldots, n$.

и если $P\{|t_{n-k}|| < \varepsilon\} = \gamma_t(\varepsilon)$, то с вероятностью $\gamma_t(\varepsilon)$ неравенство

$$|\alpha_j - \widehat{\alpha}_j| \leqslant \varepsilon \sqrt{\frac{(a_j, a_j)^{-}||(I - \Pi_a)\xi||^2}{n-k}}$$

дает интервальную оценку α_j.

6.2 Задачи редукции измерений.

1° Постановка задачи несмещенной редукции измерений.

Для схемы измерений $\xi = Af + \nu$, $\mathcal{M}\nu = 0$, $\mathcal{M}\nu\nu^{\mathsf{T}} = \sigma^2 I$ ставится **задача несмещенной редукции**:

$$\inf\{\mathcal{M}||R\xi - f||^2 \mid R,\ RA = I\} = \inf\{\sigma^2 \mathrm{tr} RR^{\mathsf{T}} \mid R,\ RA = I\}\} = h_0.$$

Решаем уравнение $RA = I$: $R = R_0 + Y$, где $R_0 = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$, а Y — решение уравнения $YA = 0 \Leftrightarrow Y\Pi_a = 0 \Leftrightarrow Y = Z(I - \Pi_a)$, $\forall Z$.

Т.о., общее решение $R = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} + Z(I - \Pi_a)$.

В этом случае $\mathrm{tr} RR^{\mathsf{T}} = \mathrm{tr}(A^{\mathsf{T}}A)^{-1} + \mathrm{tr} Z(I - \Pi_a)Z^{\mathsf{T}}$ и inf достигается на $R = R_0 = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ и равен $h_0 = \sigma^2 \mathrm{tr}(A^{\mathsf{T}}A)^{-1}$. Очевидно, этот результат совпадает с результатом, полученным в теореме Гаусса-Маркова.

В этом случае $R\xi = f + R\nu$, где $R\nu$ — шум, суммарная энергия которого равна h_0.

2° Задача редукции с ограничением на уровень шума. Часто шум, полученный при решении задачи несмещенной редукции, неприемлемо велик. Вспомним, что ошибка складывается из двух:

$$R\xi = f + (RA - I)f + R\nu.$$

Введем расстояние между матрицами[1] (операторами) A и B: $\rho^2(A - B) = \mathrm{tr}(A - B)(A - B)^{\mathsf{T}}$.

6.3 Синтез прибора с ограничением на уровень шума

Рассмотрим задачу

$$\inf\{\mathrm{tr}(RA - I)(RA - I)^{\mathsf{T}} \mid R,\ \mathcal{M}||R\nu||^2 \leqslant \varepsilon\}. \tag{6.10}$$

Заметим, что $\mathcal{M}||R\nu||^2 = \sigma^2 \mathrm{tr} RR^{\mathsf{T}} = \sigma^2 \sum\limits_{i,j} r_{ij}^2$.

Далее представим два случая:

1. $h_0 \leqslant \varepsilon$. В этом случае условие в задаче (6.10) выполняется и $R_0 = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ — есть решение, так как любое $R = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} + Z(I - \Pi_a)$ минимизирует $||RA - I||_2^2$.

2. Введем систему координат в пространстве матричных элементов (изобразим лишь два!):

Очевидно, решение есть точка касания, где $\sigma^2 \mathrm{tr} RR^{\mathsf{T}} = \varepsilon$ (равно!). Тогда решаем задачу методом множителей (одного!) Лагранжа. Функция Лагранжа

$$L(R) = \mathrm{tr}(RA - I)(RA - I)^{\mathsf{T}} + \omega\sigma^2 \mathrm{tr} RR^{\mathsf{T}}.$$

$$\nabla_R L = 2(RA - I)A^{\mathsf{T}} + 2\omega\sigma^2 R = 0.$$

[1] Можно также ввести скалярное произведение $(AB)_2 = \mathrm{tr} AB^{\mathsf{T}}$ и норму $||A||_2 = \{\mathrm{tr} AA^{\mathsf{T}}\}^{1/2}$, которая называется нормой Гильберта-Шмидта.

$R(AA^\top + \omega\sigma^2 I) = A^\top$, $R = R(\omega) = A^\top(AA^\top + \omega\sigma^2 I)^{-1} = (A^\top A + \omega\sigma^2 I)^{-1}A^\top$. Обозначим $h = \sigma^2 \mathrm{tr} R(\omega) R^\top(\omega) = \sigma^2 \mathrm{tr}(A^\top A + \omega\sigma^2 I)^{-1} A^\top A (A^\top A + \omega\sigma^2 I)^{-1}$ и[1] $g(\omega) = \mathrm{tr}(RA - I)(RA - I)^\top = \omega^2\sigma^4 \mathrm{tr}(A^\top A + \omega\sigma^2 I)^{-2}$.

Пусть далее $\{e_i\}$ — ортонормированный базис из собственных векторов оператора $A^\top A$: $A^\top A e_i = \lambda_i e_i$, $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_k > 0$[2]. Тогда

$$h = \sigma^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + \omega\sigma^2)^2}, \quad g = \sigma^4 \sum_{i=1}^{k} \frac{\omega^2}{(\lambda_i + \omega\sigma^2)^2}.$$

Вычислим производные: $\frac{dh}{d\omega} = -2\sigma^4 \sum\limits_{i=1}^{k} \frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3} < 0$, $0 < \omega < \infty$, $h \underset{\omega\to 0}{\longrightarrow} h_0 = \sigma^2 \sum\limits_{i=1}^{k} \frac{1}{\lambda_i} = \sigma^2 \mathrm{tr}(A^\top A)^{-1}$, $h \underset{\omega\to 0}{\longrightarrow} 0$. Поэтому уравнение $h(\omega) = \varepsilon$ при $\varepsilon < \varepsilon_0 = h_0$ имеет единственное решение.

Кроме того, $\frac{dg}{d\omega} = \sigma^4 \sum\limits_{i=1}^{k} \left[\frac{2\omega}{(\lambda_i+\omega\sigma^2)^2} - \frac{2\omega^2\sigma^2}{(\lambda_i+\omega\sigma^2)^3}\right] = 2\omega\sigma^4 \sum\limits_{i=1}^{k} \frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3}$ и поэтому имеет место дифференциальный закон сохранения:

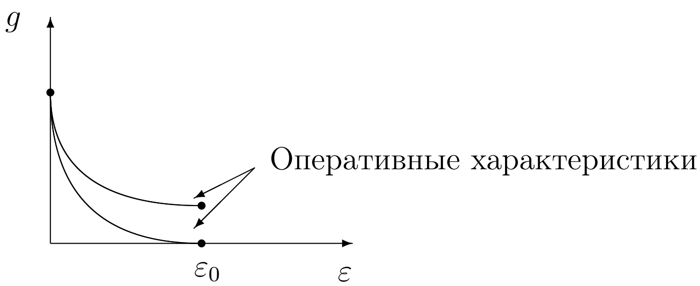
$$\omega \frac{dh}{d\omega} + \frac{dg}{d\omega} = 0. \tag{6.11}$$

Итак, общее решение задачи (6.10) имеет вид:

$$R = \begin{cases} R(\omega) = (A^\top A + \omega\sigma^2 I)^{-1}, & 0 < \varepsilon < \varepsilon_0 = h_0 = \sigma^2 \mathrm{tr}(A^\top A)^{-1}, \\ 0 & \varepsilon - 0, \\ R_0 = (A^\top A)^{-1} A^\top, & \varepsilon \geqslant \varepsilon_0 = h_0, \end{cases} \tag{6.12}$$

при этом выполняется (6.11).

Зависимость g от ε носит название оперативной характеристики. При этом характеристика, график которой лежит ниже, соответствует равномерно лучшему прибору.



7 Проверка статистических гипотез

7.1 Постановка задачи

Пусть $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ — некоторое семейство вероятностных мер (распределений) на измеримом пространстве $(X, \mathcal{A})$ и пусть с ним связана некоторая гипотеза H (непротиворечивое утверждение относительно параметра ϑ). Альтернативу обозначим K. Будем предполагать, что если параметр ϑ известен, то можно сказать, верна гипотеза или нет[3]. Это означает, что распределения класса $\mathcal{P}$ разбиваются на два множества, которые мы будем обозначать теми же буквами: $\mathcal{P} = H \cup K$, причем, если $P_\vartheta \in H$, то гипотеза верна, если $P_\vartheta \in K$, то гипотеза неверна (верна альтернатива).

[1]Здесь используется равенство $I - RA = I - (A^\top A + \omega\sigma^2 I)^{-1} A^\top A = \omega\sigma^2 (A^\top A + \omega\sigma^2 I)^{-1}$.

[2]$\mathrm{rank}\, A = k \leqslant n$.

[3]В данном случае речь идет о так называемых параметрических гипотезах. Существуют также и непараметрические гипотезы — относительно распределения в целом (например, относительно функции распределения).

Если Θ обозначает множество значений параметра ϑ, то предыдущее разбиение индуцирует разбиение ϑ: $\Theta = \Theta_H \cup \Theta_K$, причем $H = \{P_\vartheta, \vartheta \in \Theta_H\}$ и $K = \{P_\vartheta, \vartheta \in \Theta_K\}$. Далее гипотеза как утверждение и множество H отождествляются.

Поскольку вывод о справедливости гипотезы предполагается делать в терминах наблюдений случайной величины ξ, которая контролируется распределением $P_\vartheta, \vartheta \in \Theta$, естественно на множестве значений ξ определить решающую функцию $\varphi(\cdot)$, принимающую два значения: d_H, если гипотеза принимается и d_K, если гипотеза не принимается. Без ограничения общности можно считать, что $d_H = 0$, а $d_K = 1$.

Тем самым выборочное пространство разбивается на два непересекающихся множества:

$$S_H = \{x : \varphi(x) = 0\},\ S_K = \{x : \varphi(x) = 1\}.$$

Множество S_K называется критическим. Если наблюдаемое значение ξ попадает в S_K, то гипотеза отвергается. S_H — множество принятия гипотезы H.

Для каждого значения ξ представляется четыре возможности:

1. Гипотеза принята ($\varphi(x) = 0$), параметр $\vartheta \in \Theta_H$ — ошибки нет.
2. Гипотеза отвергнута ($\varphi(x) = 1$), параметр $\vartheta \in \Theta_K$ — ошибки также нет.
3. Гипотеза отвергнута ($\varphi(x) = 1$), параметр $\vartheta \in \Theta_H$ — ошибка 1-го рода, ее вероятность равна

$$P_\vartheta\{\varphi(\xi) = 1\} = P_\vartheta\{\xi \in S_K\},\ \vartheta \in \Theta_H.$$

4. Гипотеза принята ($\varphi(x) = 0$), параметр $\vartheta \in \Theta_K$ — ошибка 2-го рода, ее вероятность равна

$$P_\vartheta\{\varphi(\xi) = 0\} = P_\vartheta\{\xi \in S_H\},\ \vartheta \in \Theta_K.$$

Одновременно уменьшить обе ошибки, как правило, трудно, поэтому обычно задают *границу* для вероятности отклонения H, когда гипотеза на самом деле верна (т.е. ошибку первого рода):

$$P_\vartheta\{\varphi(\xi) = 1\} = P_\vartheta\{\xi \in S_K\} \leqslant \alpha,\ \forall\vartheta \in \Theta_H.$$

Число α называют уровнем значимости, а число

$$\sup_{\Theta_H} P_\vartheta\{\xi \in S_K\}$$

(для удобства) размером критерия или критической области[1]. При этом желательно сделать минимальной вероятность $P_\vartheta\{\xi \in S_H\}$, $\vartheta \in \Theta_K$ (ошибку 2-го рода), или, что то же самое, сделать максимальной вероятность

$$\beta = P_\vartheta\{\xi \in S_K\} = 1 - P_\vartheta\{\xi \in S_H\},\ \vartheta \in \Theta_K$$

ее отвергнуть, когда она на самом деле неверна. Рассматриваемая как функция $\vartheta \in \Theta_K$ при фиксированном значении α, она называется мощностью критерия для H при альтернативе K. В общем случае $\beta(\vartheta)$, $\vartheta \in \Theta$ называется функцией мощности критерия H.

Если на практике мощность слишком мала, то следует увеличить уровень значимости α, сбалансировав вероятность отвергнуть H, если гипотеза верна, и отвергнуть K, если гипотеза неверна. Если мы априори уверены в гипотезе H, то для ее отклонения нужны веские доводы. В этом случае следует выбирать низкий уровень значимости. Вероятность ошибиться, отвергнув H, при этом мала.

Пусть выбран уровень значимости α, тогда задача состоит в выборе критической функции φ, такой, что мощность $\beta(\vartheta)$ для всех $\vartheta \in \Theta_K$ максимальна при условии, что $\mathsf{E}_\vartheta\varphi(\xi) \leqslant \alpha$, $\vartheta \in \Theta_H$.

При этом мы сталкиваемся с характерной трудностью, состоящей в том, что как правило, критическая функция (критерий), максимизирующая мощность при некоторой альтернативе

[1] Сходным понятием является так называемая *надежность критерия*, равная случайной величине $\alpha(\xi) = \min\{\alpha | \xi \in S_\alpha,\ P_H(S_\alpha) = \alpha\}$. Если же критерий равномерно наиболее мощный, то $\alpha(\xi)$ — *надежность гипотезы*.

$\vartheta \in \Theta_K$, зависит от этой альтернативы. Поэтому необходимы дополнительные соображения о том, что следует понимать под оптимальной решающей процедурой.

Может оказаться, что один и тот же критерий максимизирует мощность для всех альтернатив из K. Такие критерии называются *равномерно наиболее мощными.*

Предварительный пример. Пусть гипотеза и альтернатива каждая содержат по одному распределению («простая» гипотеза и «простая» альтернатива). То и другое предполагаются дискретным. Задача построения критического множества эквивалентна вариационной задаче:

$$\sum_{x \in S_K} P_K(x) \sim \max, \quad \sum_{x \in S_K} P_H(x) \leqslant \alpha.$$

Нетрудно видеть, что в S_K должны быть включены точки $x_1, x_2, ...$, упорядоченные по величине отношения $t(x) = P_K(x)/P_H(x)$:

$$t(x_1) \geqslant t(x_2) \geqslant ....$$

В S_K включается максимальное число таких точек, ограниченное условием

$$P_H\{\xi \in S_K\} = \sum_{t(x)>c} P_H(x) \leqslant \alpha.$$

Однако может оказаться, что включив очередную точку в S_K, мы не достигаем α , а включив следующую, превосходим α. Эта трудность преодолевается переходом к *рандомизированным* критериям. С помощью рандомизации можно «расщепить» очередную точку, взяв в S_K такую ее часть, чтобы получить суммарную вероятность в точности равную α, не нарушая при этом порядка точек.

Рандомизированный критерий строится следующим образом. Пусть в точке x вероятность отклонения гипотезы равна $\varphi(x)$, а вероятность принятия равна $1 - \varphi(x)$. Если наблюдение $\xi = x$, то производится случайный эксперимент с двумя исходами r и $\overline{r}$, имеющими вероятности $\varphi(x)$ и $1 - \varphi(x)$. Если выпадает r, то гипотеза отвергается, если же выпадает $\overline{r}$ — принимается.

Для *простой* гипотезы и *простой* альтернативы всегда существует наиболее мощный критерий.

Теорема (фундаментальная лемма Неймана-Пирсона).

Пусть P_H и P_K — распределения вероятностей, обладающие плотностями p_H и p_K соответственно по отношению к некоторой мере μ (например, $\mu = p_H + p_K$). Тогда

1°. (Существование) Для проверки $H : P_H$ при конкурирующей гипотезе $K : P_K$ найдется критерий φ и константа λ такие, что

$$\mathsf{E}_H \varphi(\xi) = \alpha \tag{7.1}$$

и

$$\varphi(x) = \begin{cases} 1, & \text{когда } p_K > \lambda p_H, \\ 0, & \text{когда } p_K < \lambda p_H. \end{cases} \tag{7.2}$$

2°. (Достаточное условие для критерия наибольшей мощности) Если критерий удовлетворяет требованиям (7.1) и (7.2) при некотором λ, то он является наиболее мощным критерием уровня α для проверки распределения P_H при конкурирующем P_K.

3°. (Необходимое условие для критерия наибольшей мощности) Если φ наиболее мощный критерий уровня α для проверки распределения P_H при конкурирующем P_K, то при некотором λ он удовлетворяет (7.2) почти всюду по мере μ. Он также удовлетворяет (7.1), кроме случая, когда существует критерий размера $< \alpha$ и мощности 1.

Пусть $0 < \alpha < 1$. Обозначим

$$\alpha(z) = P_H\{p_K(\xi) \geqslant z p_H(\xi)\}. \tag{7.3}$$

Так как при вычислении вероятности P_H достаточно рассматривать лишь точки, в которых $p_H > 0$, то

$$\alpha(z) = P_H\{p_K(\xi) \geqslant zp_H(\xi)\} = P_H\{\frac{p_K(\xi)}{p_H(\xi)} \geqslant z\}. \tag{7.4}$$

Следовательно, $1-\alpha(z) = P_H\{\frac{p_K(\xi)}{p_H(\xi)} < z\}$ и, таким образом, $1-\alpha(z)$ — функция распределения случайной величины $\frac{p_K(\xi)}{p_H(\xi)}$. Поэтому функция $\alpha(z)$ непрерывна слева:

$$\alpha(z) = P_H \bigcap_{n=1}^{\infty}\{\frac{p_K(\xi)}{p_H(\xi)} < z - \frac{1}{n}\} = \lim_{n\to\infty} \alpha(z - \frac{1}{n}), \tag{7.5}$$

не возрастает, причем $\alpha(-\infty) = 1$, $\alpha(\infty) = 0$, и

$$\alpha(z) - \alpha(z+0) = P_H\{\frac{p_K(\xi)}{p_H(\xi)} = z\}, \quad P_H\{\frac{p_K(\xi)}{p_H(\xi)} > z\} = \alpha(z+0). \tag{7.6}$$

1°. Существование. Пусть задано α, $0 < \alpha < 1$. Определим z_0 из условия: $\alpha(z_0) \geqslant \alpha \geqslant \alpha(z_0+0)$ и рассмотрим критерий:

$$\varphi(x) = \begin{cases} 1, & p_K(x) > z_0 p_H(x) \\ \frac{\alpha - \alpha(z_0+0)}{\alpha(z_0) - \alpha(z_0+0)}, & p_K(x) = z_0 p_H(x) \\ 0. & p_K(x) < z_0 p_H(x). \end{cases} \tag{7.7}$$

Заметим прежде всего, что функция $\varphi(x)$ определена $P_H + P_K$-почти всюду. Действительно, $\varphi(x)$ не определена лишь в случае $\alpha(z_0) - \alpha(z_0+0) = 0$ в тех точках x, в которых $p_K(x) = z_0 p_H(x)$. Но при этом

$$P_H\{p_K(\xi) = z_0 p_H(\xi)\} = \alpha(z_0) - \alpha(z_0+0) = 0. \tag{7.8}$$

Из сказанного следует, что существует $\mathsf{E}_H\varphi(\xi)$, причем («размер» критерия)

$$\mathsf{E}_H\varphi(\xi) = P_H\{\frac{p_K(\xi)}{p_H(\xi)} > z_0\} + \frac{\alpha - \alpha(z_0+0)}{\alpha(z_0) - \alpha(z_0+0)} P_H\{\frac{p_K(\xi)}{p_H(\xi)} = z_0\} = \alpha.$$

Пункт 1° доказан, так как можно положить $\lambda = z_0$.

2°. (Достаточность.) Пусть критерий φ удовлетворяет условиям (7.1) и (7.2) и φ^* — любой другой критерий с уровнем $\alpha_1 \leqslant \alpha$. Обозначим

$$S^+ = \{x : \varphi - \varphi^* > 0\}, \quad S^- = \{x : \varphi - \varphi^* < 0\}. \tag{7.9}$$

В точках $S^+ : \varphi(x) - \varphi^*(x) > 0$, следовательно, $p_K(x) > \lambda p_H(x)$, т.к. $\varphi(x) > 0$, т.е. $\varphi(x) \neq 0$. В точках $S^- : \varphi(x) < \varphi^*(x) \leqslant 1$, следовательно, $p_K(x) \leqslant \lambda p_H(x)$, т.к. $\varphi(x) < 1$, т.е. $\varphi(x) \neq 1$. Поэтому

$$\int\limits_X (\varphi - \varphi^*)(p_K - \lambda p_H) d\mu = \int\limits_{S^+ \cup S^-} (\varphi - \varphi^*)(p_K - \lambda p_H) d\mu \geqslant 0,$$

откуда следует, что

$$\mathsf{E}_K(\varphi - \varphi^*) = \int (\varphi - \varphi^*) p_K d\mu \geqslant \lambda \int (\varphi - \varphi^*) p_H d\mu =$$

$$= \lambda \mathsf{E}_H(\varphi - \varphi^*) \geqslant \lambda(\alpha - \alpha_1) \geqslant 0,$$

т.е., что φ — наиболее мощный критерий уровня α для проверки P_H против P_K.

3°. (Необходимость.) Пусть φ^* — наиболее мощный критерий уровня $\alpha_1 \leqslant \alpha$ и φ удовлетворяет (7.1) и (7.2). Обозначим S пересечение множества $S^+ \cup S^-$ (см. (7.9)), на котором $\varphi \neq \varphi^*$, и множества $\{x : p_K(x) \neq \lambda p_H(x)\}$. Последнее взято в силу того, что в (8.7) нет условия с равенством $p_K(x) = \lambda p_H(x)$. Допустим $\mu(S) > 0$. Из положительности на S произведения $(\varphi - \varphi^*)(p_K - \lambda p_H)$ следует

$$\int\limits_{S^+ \cup S^-} (\varphi - \varphi^*)(p_K - \lambda p_H) d\mu = \int\limits_{S} (\varphi - \varphi^*)(p_K - \lambda p_H) d\mu > 0$$

т.е., что φ - более мощный критерий, чем φ^*. Из этого противоречия заключаем, что $\mu(S) = 0$.

Если бы φ^* имел размер, меньший α, и мощность, меньшую единицы, то в критическую область можно было бы добавить точки (или части точек), так, чтобы либо мощность стала равной единице, либо размер стал равным α. Таким образом, или $\mathsf{E}_H \varphi^* = \alpha$ или $\mathsf{E}_K \varphi^* = 1$. $\square$

В качестве следствия может быть получен следующий результат.

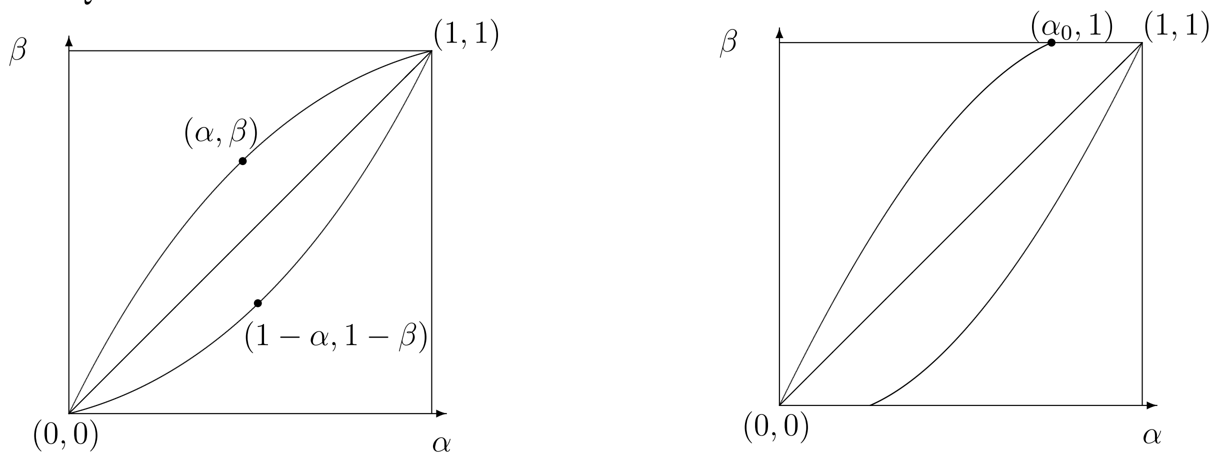
Теорема. Пусть β — мощность наиболее мощного критерия уровня α для проверки P_H против P_K. Тогда[1] $\beta > \alpha$ за исключением случая $P_H = P_K$.

Доказательство. Если положить $\varphi^*(x) = \alpha$, то $\mathsf{E}_H \varphi^* = \alpha$ и $\mathsf{E}_K \varphi^* = \alpha = \beta$. Тогда для наиболее мощного критерия φ: $\beta = \mathsf{E}_K \varphi \geqslant \mathsf{E}_K \varphi^* = \alpha$.

Если $\alpha = \beta$, то φ^* — наиболее мощный критерий. Но тогда по предыдущей теореме он удовлетворяет (8.7). Поэтому $p_K(x) = \lambda p_H(x) \mod \mu$ и, следовательно, $P_H = P_K$ $\square$

Замечание. Пусть Q — множество точек α, β, таких, что существует критерий φ, необязательно наиболее мощный, для которого $\mathsf{E}_H \varphi = \alpha$, $\mathsf{E}_K \varphi = \beta$. Множество Q, как легко проверить, выпукло, содержит точки (0,0) и (1,1) и вместе с точкой α, β содержит также $1-\alpha, 1-\beta$. Можно показать, что Q замкнуто.

Существование наиболее мощного критерия для каждого α является следствием замкнутости Q.



Пример. Найти наиболее мощный критерий уровня $\alpha = 0.05$ для гипотезы $\xi \sim U[-1,1]$ против альтернативы $\xi \sim \mathcal{N}(0,1)$ по одному измерению ξ. Определить мощность этого критерия.

Поскольку вне $[-1,1]$ $p_K(x) > \lambda p_H(x) = 0$ при любом λ, в критическое множество войдет внешняя часть промежутка $[-1,1]$. Кроме того, в критическое множество может войти (при $p_K(0) = \frac{1}{\sqrt{2\pi}} > \frac{\lambda}{2}$) симметричная окрестность нуля, т.е. $S_K = (-\infty, -1) \cup (-\varepsilon, \varepsilon) \cup (1, +\infty)$, а из условия $\int\limits_{S_K} p_H(x) dx = \alpha$ получаем, что $\varepsilon = \alpha$. После этого нетрудно посчитать мощность $\beta = \int\limits_{S_K} p_K(x) dx = 1 - 2\Phi(-\alpha) + 2\Phi(-1) = 0.357188$.

[1] Это свойство критерия называется *несмещенностью*.

7.2 Продолжение темы проверка статистических гипотез.

Ранее были рассмотрены некоторые задачи интервального оценивания. В более общем случае речь идет о построении доверительных подмножеств пространства параметров.

Мы хотим на одном примере установить связь между задачей построения доверительных множеств и задачей проверки параметрических гипотез.

Пусть $\xi \sim \mathcal{N}(\vartheta_0, \sigma^2)$ и пусть $\sigma > 0$ известна. Тогда $\frac{\sum(\xi_j - \vartheta_0)}{\sqrt{n}\sigma} \sim \mathcal{N}(0, 1)$ и

$$P_\vartheta \left\{ \left| \frac{\sum(\xi_j - \vartheta_0)}{\sqrt{n}\sigma} \right| < z_{1-\alpha/2} \right\} = 1 - \alpha,$$

т. е.

$$P_{\vartheta_0} \{\vartheta \in S(\xi)\} = 1 - \alpha,$$

где

$$S(\xi) = \left\{ \xi = (\xi_1, ..., \xi_n) : \left(\frac{1}{n} \sum \xi_j - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}, \frac{1}{n} \sum \xi_j + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right) \right\}.$$

Если параметр ϑ не принадлежит множеству S, то либо гипотеза H о значении параметра $\vartheta = \vartheta_0$ неверна, либо мы должны допустить ошибку, вероятность которой не превышает α. Таким образом, критическим является дополнительное к S множество в $\mathcal{R}^n$: $S_K = \overline{S}$, то есть,

$$S_K(\xi) = \left\{ \xi = (\xi_1, ..., \xi_n) : \left(\left| \frac{1}{n} \sum \xi_j - \vartheta \right| > \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right) \right\}. \tag{7.10}$$

Другой путь к нахождению критерия (при $\vartheta \in \mathcal{R}^1$) — это попытаться найти критическое множество наилучшее для альтернатив в известном смысле близких к (нулевой) гипотезе.

Рассмотрим задачу проверки гипотез в случае, когда плотность имеет вид $f(x, \vartheta)$, $\vartheta \in \mathcal{R}^1$ и речь идет о гипотезе $H_0 : \vartheta = \vartheta_0$. Рассмотрим нерандомизированный критерий и соответствующее критическое множество S_K уровня α.

$$\int\limits_{S_K} f(x, \vartheta_0) dx = \alpha. \tag{7.11}$$

Функция мощности критерия определяется равенством

$$\beta(\vartheta) = \int\limits_{S_K} f(x, \vartheta) dx, \ \vartheta \in \mathcal{R}^1. \tag{7.12}$$

Предположим, что $\beta(\vartheta)$ может быть разложена в ряд

$$\beta(\vartheta) = \alpha + (\vartheta - \vartheta_0)\beta'(\vartheta_0) + \frac{(\vartheta - \vartheta_0)^2}{2}\beta''(\vartheta_0) + ... \tag{7.13}$$

Если K — класс односторонних альтернатив $\vartheta > \vartheta_0$, $K = \{P_\vartheta : \vartheta > \vartheta_0\}$, то для получения *локально наиболее мощного* одностороннего критерия следует максимизировать $\beta'(\vartheta_0)$, или

$$\beta'(\vartheta_0) = \int\limits_{S_K} \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} dx \sim \max, \tag{7.14}$$

если предположить возможность дифференцирования под знаком интеграла.

Если $K = \{P_\vartheta : \vartheta < \vartheta_0\}$, то $\beta'(\vartheta_0)$ следует минимизировать, а в случае двустороннего класса альтернатив $K = \{P_\vartheta : \vartheta \neq \vartheta_0\}$ следует наложить условие *локальной несмещенности* $\beta'(\vartheta_0) = 0$ и максимизировать $\beta''(\vartheta_0)$. Такой критерий называется *локально наиболее мощным несмещенным*.

Во всех рассмотренных случаях критическое множество может быть найдено на основе следующей леммы.

Лемма (Нейман-Пирсон).

Пусть $f_0, f_1, ..., f_m$ — μ-интегрируемые на $\mathcal{R}^n$ функции и S — измеримое подмножество $\mathcal{R}^n$, для которого

$$\int\limits_S f_j(x)d\mu(x) = C_j, \quad j = 1, 2, .., m, \tag{7.15}$$

где C_j, $j = 1, 2, ..., m$ — заданные числа. Пусть далее существуют постоянные $k_1, ..., k_m$, такие, что для измеримого множества $S_0 \subset \mathcal{R}^n$, в точках которого $f_0(x) \geqslant k_1 f_1(x) + ... + k_m f_m(x)$ и вне которого $f_0(x) \leqslant k_1 f_1(x) + ... + k_m f_m(x)$. Кроме того, для $S_0 \subset \mathcal{R}^n$ также выполнены условия (8.5).

Тогда

$$\int\limits_{S_0} f_0(x)d\mu(x) \geqslant \int\limits_S f_0(x)d\mu(x). \tag{7.16}$$

Доказательство. Согласно условиям леммы

$$\int\limits_{S_0} f_0(x)d\mu(x) - \int\limits_S f_0(x)d\mu(x) = \int\limits_{S_0 \setminus S \cap S_0} f_0(x)d\mu(x) - \int\limits_{S \setminus S \cap S_0} f_0(x)d\mu(x) \geqslant$$

$$\geqslant \int\limits_{S_0 \setminus S \cap S_0} \sum_1^m k_j f_j(x)d\mu(x) - \int\limits_{S \setminus S \cap S_0} \sum_1^m k_j f_j(x)d\mu(x) = 0.$$

При этом последнее равенство нулю следует из равенств

$$\int\limits_{S_0 \setminus S \cap S_0} f_j(x)d\mu(x) = \int\limits_{S \setminus S \cap S_0} f_j(x)d\mu(x), \quad j = 1, \ldots, m,$$

которые, в свою очередь, следуют из

$$\int\limits_{S_0} f_j(x)d\mu(x) = \int\limits_S f_j(x)d\mu(x) = C_j, \quad j = 1, \ldots, m. \quad \square$$

Вернемся к задаче построения критического множества для локальных критериев.

Теорема. Для случая односторонних альтернатив получаем следующий результат. Пусть S_K $K = \{P_\vartheta : \vartheta > \vartheta_0\}$. Определим критическое множество S_K равенством

$$S_K = \Big\{x : \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} \geqslant k f(x, \vartheta_0)\Big\}, \tag{7.17}$$

где постоянная k определяется условием $\int\limits_{S_K} f(x, \vartheta_0)dx = \alpha$. Тогда

$$\int\limits_{S_K} \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} dx \geqslant \int\limits_S \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} dx$$

для любого другого критического множества S размера (уровня) α.

Доказательство. Достаточно сослаться на лемму, в которой

$$f_0(x) = \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0}; \quad f_1(x) = f(x, \vartheta_0).$$

Аналогично, если $K = \{P_\vartheta : \vartheta < \vartheta_0\}$, то

$$S_K = \Big\{x : \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} \leqslant k f(x, \vartheta_0)\Big\}.$$

Наконец, в случае двусторонних альтернатив $K = \{P_\vartheta : \vartheta \neq \vartheta_0\}$

$$\beta(\vartheta_0) = \int\limits_{S_K} f(x, \vartheta_0)dx = \alpha; \ \beta'(\vartheta_0) = \int\limits_{S_K} \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0}dx = 0 \tag{7.18}$$

и условие

$$\beta''(\vartheta_0) = \int\limits_{S_K} \frac{\partial^2 f(x, \vartheta_0)}{\partial^2 \vartheta_0}dx \sim \max$$

максимизирует критическое множество S_K, определяемое равенством

$$S_K = \{x : \frac{\partial^2 f(x, \vartheta_0)}{\partial^2 \vartheta_0} \geqslant k_1 \frac{\partial f(x, \vartheta_0)}{\partial \vartheta_0} + k_2 f(x, \vartheta_0)\},$$

где k_1 и k_2 определяется условием (7.18). □

Заметим, что мы получили достаточность решений при условии существования надлежащих k_j, $j = 1, 2$. При некоторых предположениях относительно плотностей Данциг и Вальд доказали как необходимость, так и существование соответствующих k_j, $j = 1, 2$.

Рассмотрим теперь задачу проверки гипотезы $\mu = \mu_0$ при двусторонней альтернативе $K = \{\mathcal{N}(\mu, \sigma^2) : \mu \neq \mu_0\}$ и фиксированным $\sigma > 0$.

Так как

$$f(x, \mu_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2}\sum(x_j - \mu_0)^2)\};$$

$$\frac{1}{f(x, \mu_0)}\frac{\partial f(x, \mu_0)}{\partial \mu_0} = \frac{1}{\sigma^2}\sum(x_j - \mu_0)$$

$$\frac{1}{f(x, \mu_0)}\frac{\partial^2 f(x, \mu_0)}{\partial^2 \mu_0} = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\left[\sum(x_j - \mu_0)\right]^2,$$

то критическое множество имеет вид

$$-\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\left[\sum(x_j - \mu_0)\right]^2 \geqslant \frac{\overline{k}_1}{\sigma^2}\sum(x_j - \mu_0) + \overline{k}_2,$$

или что то же самое,

$$S_K = \left\{x : \frac{1}{\sigma^4}\left[\sum(x_j - \mu_0)\right]^2 \geqslant k_1 \sum(x_j - \mu_0) + k_2\right\}. \tag{7.19}$$

Решение неравенства (7.19) имеет вид $\sum(x_j - \mu_0) < C_1$, $\sum(x_j - \mu_0) > C_2$, где $C_1 < C_2$, но в силу условия

$$\int\limits_{S_K} f(x, \mu_0)\sum(x_j - \mu_0)dx = 0,$$

$|C_1| = |C_2| = C$ так что константу k_1 в (7.19) можно положить равной нулю.

Окончательно,

$$S_K = \left\{x : \left|\frac{\sum(x_j - \mu_0)}{\sqrt{n\sigma^2}}\right| \geqslant C\right\}. \tag{7.20}$$

При этом постоянную $C = \sqrt{\frac{\sigma^2 k_2}{n}}$ можно определить, учитывая, что $\left[\frac{\sum(x_j - \mu_0)}{\sqrt{n\sigma^2}}\right] \sim \mathcal{N}(0, 1)$ и

$$\int\limits_{S_K} f(x, \mu_0)dx = 2(1 - \Phi(C) = \alpha,$$

таким образом, $C = z_{1-\alpha/2}$— квантиль уровня $1 - \alpha/2$.

Критические множества, определенные равенствами (7.10) и (7.20) совпадают. Поскольку последнее не зависит от альтернативы $\mu = \mu_1$, задаваемый (7.20) критерий является не только локально, но и РНМ критерием.

Принцип отношения правдоподобия

Пусть $L(x, \vartheta)$ — функция правдоподобия в задаче проверки гипотезы $H : \{\vartheta \in \Theta_H\}$ против альтернативы $K : \{\vartheta \in \Theta_K\}$. *Критерий отношения правдоподобия* определяется статистикой

$$\lambda = \lambda(x) = \frac{\sup\limits_{\vartheta \in \Theta_H} L(x, \vartheta)}{\sup\limits_{\vartheta \in \Theta_H \cup \Theta_K} L(x, \vartheta)}. \tag{7.21}$$

Очевидно, $0 \leqslant \lambda \leqslant 1$. Критическое множество S_K объема α для проверки гипотезы $\vartheta \in \Theta_H$ имеет вид $\lambda(x) < C_\alpha$, где постоянная C_α определяется условием

$$\int\limits_{S_K} L(x, \vartheta) dx \leqslant \alpha, \ \vartheta \in \Theta_H. \tag{7.22}$$

Здесь всюду x и ϑ многомерные.

Тест $\lambda < C$ интуитивно означает, что если правдоподобие $L(x, \vartheta)$ при расширении области Θ_H до $\Theta_H \cup \Theta_K$ значительно возрастает, то "наиболее вероятное"ϑ не принадлежит Θ_H, т.е. гипотезу следует отвергнуть.

Наоборот, если $L(x, \vartheta)$ не возрастает или возрастает незначительно, то "наиболее вероятное" ϑ принадлежит Θ_H, т.е. гипотезу не следует отвергать.

Пример. Рассмотрим пример проверки гипотезы $\vartheta_H = \{\mu = \mu_0, \sigma^2 > 0\}$ против альтернативы $\vartheta_K = \{\mu \neq \mu_0, \sigma^2 > 0\}$ на основании выборки $x_1, ..., x_n$ из $\mathcal{N}(\mu, \sigma^2)$. В этом случае

$$\max_{\vartheta \in \Theta_H} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2} \sum (x_j - \mu_0)^2)\} = \left\{\frac{1}{2\pi \frac{1}{n} \sum (x_j - \mu_0)^2}\right\}^{n/2} \exp\left(-\frac{n}{2}\right),$$

$$\max_{\vartheta \in \Theta_H \cup \vartheta \in \Theta_K} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2} \sum (x_j - \mu_0)^2)\} = \\ = \left\{\frac{1}{2\pi \frac{1}{n} \sum (x_j - \overline{x})^2}\right\}^{n/2} \exp\left(-\frac{n}{2}\right),$$

так что $\lambda = \{\sum (x_j - \overline{x})^2 / \sum (x_j - \mu_0)^2\}^{n/2}$.

Поскольку

$$\|x - \overline{\mu}_0\|^2 = \|(I - \Pi)(x - \overline{\mu}_0)\|^2 + \|\Pi(x - \overline{\mu}_0)\|^2,$$

где $\overline{\mu}_0 = (\mu_0, \ldots, \mu_0)$, а Π — ортогональный проектор на $\frac{\overline{\mu}_0}{||\overline{\mu}_0||}$, или, что то же самое,

$$\sum (x_j - \mu_0)^2 = \sum (x_j - \overline{x})^2 + n(\overline{x} - \mu_0)^2,$$

то

$$\lambda = \left\{1 + \frac{n(\overline{x} - \mu_0)^2}{\sum (x_j - \overline{x})^2)}\right\}^{-n/2} = \left\{1 + \left[\frac{\sqrt{n}(\overline{x} - \mu_0)}{\|(I - \Pi)x\| / \sqrt{n-1}}\right]^2 \frac{1}{n-1}\right\}^{-n/2},$$

где статистика $\frac{\sqrt{n}(\overline{x} - \mu_0)}{\|(I - \Pi)x\| / \sqrt{n-1}}$ при гипотезе $\mu = \mu_0$ контролируется распределением Стьюдента с $n-1$ степенью свободы. Тем самым критерий отношения правдоподобия $\lambda < C$ эквивалентен критерию $|t_{n-1}| = \frac{\sqrt{n}|\overline{x} - \mu_0|}{\|(I - \Pi)x\| / \sqrt{n-1}} > \widetilde{C}$ и является подобным[1] и несмещенным ($\beta > \alpha$).

[1]Тесты, в которых критическое множество не зависит от свободных параметров, называются *подобными выборочному пространству*, или просто *подобными*.

Критерий отношения правдоподобия в теории нормальной (линейной) регрессии

Наиболее важное приложение принцип отношения правдоподобия при конечном объеме выборки находит в теории нормальной регрессии. В качестве иллюстрации рассмотрим задачу проверки гипотезы о равенстве нулю некоторых коэффициентов в схеме нормальной регрессии. Именно, пусть требуется выбрать между двумя возможностями:

$$H : \xi \in \mathcal{N}(\alpha_1 e_1 + ... + \alpha_k e_k, \|\sigma^2 \delta_{ij}\|);\ e_q = (e_{q1}, ..., e_{qn}),$$

$$K : \xi \in \mathcal{N}(\alpha_1 e_1 + ... + \alpha_k e_k + ... + \alpha_s e_s, \|\sigma^2 \delta_{ij}\|);\ i, j = 1, 2, ..., n.$$

Так как

$$\max_{\Theta} L(x, \vartheta) = \left\{ \frac{1}{2\pi \frac{1}{n} \min\limits_{\alpha} \sum\limits_j (x_j - \sum\limits_q \alpha_q e_{qj})^2} \right\}^{n/2} \exp\left(-\frac{n}{2}\right),$$

то получаем следующее выражение для отношения правдоподобия

$$\lambda = \left\{ \frac{\|x - \Pi_k x\|^2}{\|x - \Pi_s x\|^2} \right\}^{-n/2} = \left\{ 1 + \frac{\|(\Pi_s - \Pi_k)x\|^2}{\|x - \Pi_s x\|^2} \right\}^{-n/2}.$$

Если верна гипотеза H, то $\|x - \Pi_s x\|^2 \sim \sigma^2 \chi_{n-s}$, $\|(\Pi_s - \Pi_k x)\|^2 \sim \sigma^2 \chi_{s-k}$ и, следовательно, в этом случае

$$\frac{\frac{\|(\Pi_s - \Pi_k x)\|^2}{s-k}}{\frac{\|x - \Pi_s x\|^2}{n-s}} = F_{s-k,\ n-s}.$$

Поэтому критерий $\lambda < C$ эквивалентен F-критерию: $F_{s-k,\ n-s} \geqslant C_\alpha$, где C_α определяется из условия $P\{F_{s-k,\ n-s} \geqslant C_\alpha\} = \alpha$. Последняя вероятность может быть получена из таблиц распределения Снедекора-Фишера.

Также, как принцип нахождения оценок максимального правдоподобия, критерий, основанный на отношении правдоподобия имеет "приблизительный" характер, тем не менее, он обладает хорошими асимптотическим свойствами, в частности, состоятельностью.

Пусть S — критическое множество гипотезы $\vartheta \in \Theta_H$ при альтернативе $\vartheta \in \Theta_K$. Соответствующий критерий называется состоятельным критерием объема α, если

$$\lim_{n \to \infty} P_\vartheta(S) \leqslant \alpha,\ \vartheta \in \Theta_H\ \lim_{n \to \infty} P_\vartheta(S) = 1,\ \vartheta \in \Theta_K.$$

Критерий χ^2.

1. Рассмотрим полиномиальное распределение с параметрами p_k и найдем матрицу ковариаций.

$$\mathsf{E}\xi_i = np_i, \qquad i = 1, \ldots, r,$$

$$\mathsf{E}\xi_i\xi_j = p_i p_j \frac{\partial^2}{\partial p_i \partial p_j} \left(\sum_{k=1}^{r} p_k \right)^n = n(n-1) p_i p_j, \qquad i, j = 1, \ldots, r,$$

$$\mathsf{E}\xi_i^2 = p_i^2 \frac{\partial^2}{\partial p_i^2} \left(\sum_{k=1}^{r} p_k \right)^n + \mathsf{E}\xi_i = n(n-1) p_i^2 + np_i, \qquad i = 1, \ldots, r.$$

Отсюда

$$\mathrm{cov}\xi_i\xi_j = n(\delta_{ij} p_i - p_i p_j), \qquad i, j = 1, \ldots, r.$$

В дальнейшем будем рассматривать вектор $\xi = (\xi_1, \ldots, \xi_{r-1})$ с $r - 1$ независимыми координатами и его положительно определенную ковариационную матрицу $\mathrm{cov}\xi_i\xi_j$ размера $r - 1$.

2. Легко проверить умножением, что справедлива формула

$$(A - xx^*)^{-1} = (A)^{-1} + \frac{(A)^{-1}x((A)^{-1}x)^*}{1 - x^*(A)^{-1}x},$$

где $A = A^*$ — обратимая матрица, x — вектор-столбец, $*$ — знак сопряжения (транспонирования).

3. Полагая $A = ||\delta_{ij}p_i||$, а $x = (p_1, p_2, \ldots, p_{r-1})^*$ получим

$$||\mathrm{cov}\xi_i\xi_j||^{-1} = \frac{1}{n}\left(A^{-1} + \frac{ee^*}{p_r}\right) = \frac{1}{n}\left\|\frac{\delta_{ij}}{p_i} + \frac{1}{p_r}\right\|,$$

где $e = A^{-1}x = (1, 1, \ldots, 1)^*$ — вектор размерности $(r-1)$.

4. Сформируем квадратичную форму $(\xi_i - np_i)^*||\mathrm{cov}\xi_i\xi_j||^{-1}(\xi_i - np_i)$, которая сходится к χ^2_{r-1} распределению:

$$\sum_{i,j=1}^{r-1}(\xi_i - np_i)^*||\mathrm{cov}\xi_i\xi_j||^{-1}(\xi_j - np_j) =$$

$$= \sum_{i=1}^{r-1}\frac{(\xi_i - np_i)^2}{np_i} + \frac{(\xi_r - np_r)^2}{np_r} = \sum_{i=1}^{r}\frac{(\xi_i - np_i)^2}{np_i}.$$

Окончательно имеем (обычно в книгах пишут n_i вместо ξ_i)

$$\sum_{i=1}^{r}\frac{(n_i - np_i)^2}{np_i} \sim \chi^2_{r-1}.$$

Таким образом, получаем асимптотический (при $n \to \infty$) критерий для гипотезы $\vartheta = p$ против альтернативы $\vartheta \neq p$:

$$\sum_{j=1}^{r}\frac{(n_j - np_j)^2}{np_j} > C_\alpha, \tag{7.23}$$

где

$$P(\chi^2_{r-1} > C_\alpha) = \alpha.$$

Этот критерий является примером асимптотически непараметрического критерия, т.к. предельное распределение используемых в нем статистик является «абсолютным», т.е. никак не связано с природой исходного распределения. Отсюда виден подход к решению задачи проверки непараметрических гипотез: используется специальный прием параметризации — группировка данных.

Область возможных значений наблюдаемых величин разбивается на r непересекающихся областей и вместо наблюдения указывается лишь тот интервал, в который это наблюдение попало. Проведенная редукция выборки $x \in \mathcal{R}^n$ к вектору ξ называется группировкой данных. Ясно, что при этом происходит частичная потеря информации, которая, впрочем, уменьшается при дроблении областей. К другим недостаткам этого метода относится необъективный характер выбора областей, зависящий от выборки и/или от исследователя.

8 Теория статистических решений

Рассмотрим типичную ситуацию, в которой возникает задача принятия решения. Предположим, что нам известны возможные «состояния природы» $\vartheta \in \Theta$ например, $\vartheta_1, \ldots, \vartheta_k$ и определены возможные «действия» $d \in D$, например, $d_1, \ldots, d_N$, которые связаны с состояниями природы таким образом, что действие d_i, выполненное при состоянии природы ϑ_j влечет потери $l(\vartheta_i, d_j)$ (или другие «неприятности», оцениваемые числом $l(\vartheta_i, d_j)$, причем значения риска, сопутствующие каждой комбинации ϑ_i, d_j 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 котором риск $l(\vartheta, d)$ минимален. В данном случае *правило действия* состоит в наблюдении за состоянием природы и принятии определенного *решения о действии*, если минимум $l(\vartheta, d)$ как функции $d \in D$ достигается на одном действии d_i. Если минимум $l(\vartheta, d)$, соответствующий состоянию природы ϑ достигается на нескольких $d \in D$, скажем на $d_{i_1}, \dots, d_{i_m}$, то можно выполнить любое из них. Но можно также воспользоваться экспериментом с m случайными исходами $\alpha_1, \alpha_2, \dots, \alpha_m$, $p(\alpha_1)+p(\alpha_2)+\dots+p(\alpha_m) = 1$. 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 рандомизированного правила риск $l(\vartheta, d)$ при фиксированном ϑ является случайной величиной, но

$$\mathsf{E}l(\vartheta, d) = \sum_{t=1}^{m} l(\vartheta, d_i)p_t = l(\vartheta, d_{i_t}) = l(\vartheta), \quad i = i_1, \dots, i_m.$$

На самом деле, конечно, состояние природы в момент принятия решения обычно неизвестно. Если, однако, о состоянии природы неизвестно ничего (в том числе, неизвестно множество Θ возможных состояний природы), то нет и задачи принятия решения: можно принять любое решение, так как в терминах риска невозможно привести аргументы в пользу какого-нибудь одного из них. Если же известно множество Θ всевозможных состояний природы, то оптимальное правило можно определить, например, как решение $d^* \in D$ задачи

$$c^* = \max_{\vartheta\in\Theta} l(\vartheta, d^*) = \min_{d\in D} \max_{\vartheta\in\Theta} l(\vartheta, d), \tag{8.1}$$

минимизирующее в (8.1) максимальный риск $\max\limits_{\vartheta\in\Theta} l(\vartheta, d) = l(\vartheta(d), d)$, $d \in D$, 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 решения $d_1, \dots, d_N$ каждый раз принимаются случайно с некоторыми 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

$$\max_{\vartheta\in\Theta} \mathsf{E}l(\vartheta, \delta^*) = \min_{\delta} \max_{\vartheta\in\Theta} \mathsf{E}l(\vartheta, \delta), \tag{8.2}$$

в которой $\mathsf{E}l(\vartheta, \delta) = \sum\limits_{i=1}^{N} p_i l(\vartheta, d_i)$ и $\min\limits_{\delta}$ вычисляется на множестве $\mathcal{P} = \{(p_1, \dots, p_N),\ p_i \geqslant 0,\ i = 1, \dots, N,\ p_1 + \dots, +p_N = 1\}$ всех распределений δ.

Для $\Theta = \{\vartheta_1, \dots, \vartheta_k\}$ определим *ожидаемый (маргинальный) риск*, отвечающий состоянию природы ϑ_t,

$$l_t(p) = \mathsf{E}l(\vartheta_t, \delta) = \sum_{i=1}^{N} p_i l(\vartheta_t, d_i), \quad t = 1, \dots, k, \tag{8.3}$$

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 $p = p^{(i)} = (0, \dots, 0, p_i = 1, 0, \dots, 0)$, $i = 1, \dots, N$.

В задаче (8.1) требуется найти точку $l(p^*) = (l_1(p^*), \dots, l_k(p^*)) = (\mathsf{E}l(\vartheta_1, \delta^*), \dots, \mathsf{E}l(\vartheta_k, \delta^*)) \equiv (\sum_{i=1}^N p_i^* l(\vartheta_1, d_i), \dots, \sum_{i=1}^N p_i^* l(\vartheta_k, d_i)) \in \mathcal{L} = \{l(p), p \in \mathcal{P}\} = \mathrm{co}\{l^{(1)}, \dots, l^{(N)}\}$, максимальная координата которой минимальна[1], $\max\limits_{1\leqslant t\leqslant k} l_t(p^*) = \min\limits_{p\in\mathcal{P}} \max\limits_{1\leqslant t\leqslant k} l_t(p) = \min\limits_{p\in\mathcal{P}} \max\limits_{1\leqslant t\leqslant k} l_t$, $l = (l_1, \dots, l_k)$. 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 (6.10), записанная в виде

$$\begin{aligned} c_r^* = \max(l_1^*, \dots, l_k^*) = \min\{\max\{l_1, \dots, l_k\} | l \in \mathcal{L}\} = \\ = \min\{\max\{l_1(p), \dots, l_k(p)\} | p \in \mathcal{P}\}, \end{aligned} \tag{8.4}$$

всегда имеет решение.

На рис. 1 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

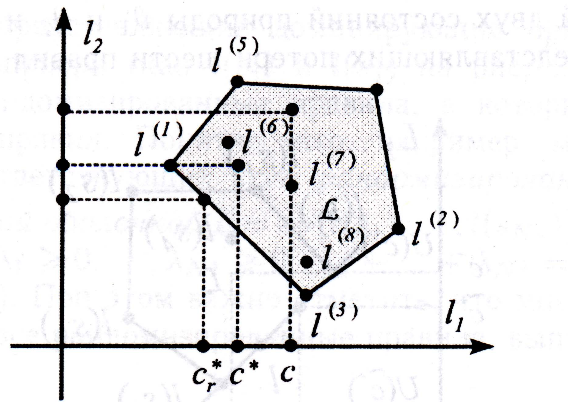


Рис. 1. Множество $\mathcal{L} = \mathrm{co}(l^{(1)}, \dots, l^{(8)})$; множество $\{(l_1, l_2) \in \mathcal{R}^2, \quad \max(l_1, l_2) = c\}$, множество $\{(l_1, l_2) = \max(l_1^*, l_2^*) = c^*\}$, где $l_1^* = l(\vartheta_1, d^*)$, $l_2^* = l(\vartheta_2, d^*)$, d^* — решение задачи (8.1); c_r^* — значение минимума в задаче (8.4), определяющее распределение оптимального рандомизированного действия δ^*, $p_2^* = p_4^* = p_5^* = p_6^* = p_7^* = p_8^* = 0$, $p_1^* l_1^{(1)} + p_3^* l_1^{(3)} = p_1^* l_2^{(1)} + p_3^* l_2^{(3)} = c_r^*$, $p_1^* + p_3^* = 1$; $\max\limits_{s=1,2} \mathsf{E}l(\vartheta_s, \delta^*) = c_r^* \leqslant \max\limits_{s=1,2} l(\vartheta_s, d^*) = c^*$, т.е. ожидаемый риск c_r^*, сопутствующий рандомизированному решению δ^* в (8.4), меньше риска c^*, сопутствующего решению d^* в (8.1)

.

которой возможны наблюдения над 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, $s(\cdot): \ X \to D$. Если

$$s(x) = d, \tag{8.5}$$

то правило s при наблюдении x предписывает действие d. В данном случае всего N^q отображений $s_i(\cdot): \ \{x_1, \dots, x_q\} \ \to \{d_1, \dots, d_N\}$, множество всех таких отображений (чистых правил) обозначим S.

При этом с каждым правилом s связано разбиение (которое мы также обозначим s) множества наблюдений X на подмножества $D_1, \dots, D_N$: $X = D_1 + \dots + D_N$, где

$$D_j = \{x \in X, \ s(x) = d_j\}, \quad j = 1, \dots, N.$$

[1] $\mathrm{co}(l^{(1)}, \dots, l^{(N)}) = \{l^{(1)} p_1, \dots, l^{(N)} p_N, \quad p = (p_1, \dots, p_N) \in \mathcal{P}\}$ — 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. 1.

Каждое *правило действия* сопряжено с риском и, естественно, лучшим является то, которому сопутствует меньший риск. Задача сводится к выбору лучшего правила.

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

$$p_s(d|\vartheta) = \sum_{x:s(x)=d} p(x|\vartheta), \quad d \in D, \quad \vartheta \in \Theta, \quad s(\cdot) \in S, \tag{8.6}$$

и вычислить ожидаемый риск потерь, сопутствующий применению правила s в состоянии природы $\vartheta_i \in \Theta$,

$$L_i(s) = \sum_{t=1}^{N} l(\vartheta_i, d_t) p_s(d_t|\vartheta_i), \quad i = 1, \ldots, k. \tag{8.7}$$

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 $(L_1(s), \ldots, L_k(s))$, $l(s) = (L_1(s), \ldots, L_k(s))$, $s \in S$. На рис. 2, как и на рис. 1, представлен случай двух состояний природы ϑ_1 и ϑ_2 и изображены шесть точек в $\mathcal{R}^2$, представляющих потери шести правил решения[1]. (Координаты каждой точки являются соответствующими

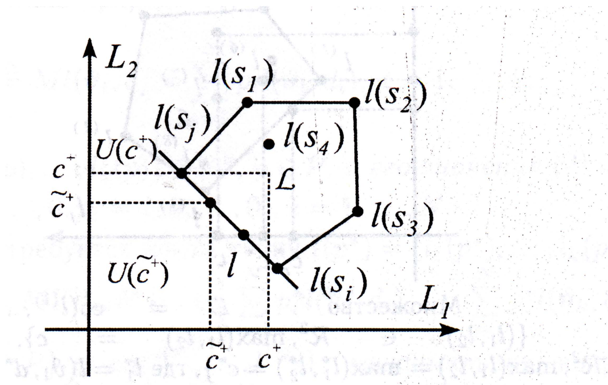


Рис. 2. Шесть точек представляют потери, отвечающие шести отмеченным чистым правилам $s_1, \ldots, s_4, \ldots, s_i, \ldots, s_j$, множество $U\widetilde{c}^+$ определяет минимаксный риск, равный $\widetilde{c}^+$, c^+ — минимальный риск, полученный как решение задачи (8.9).

потерями.)

Если для правил s_1 и s_2 $l(s_1) = (L_1(s_1), L_2(s_1)) \leqslant (L_1(s_2), L_2(s_2)) = l(s_2)$, что означает: $L_1(s_1) \leqslant L_1(s_2)$, $L_2(s_1) \leqslant L_2(s_2)$, то говорят, что правило s_1 *доминирует* над s_2 (см. рис. 2). 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 $s_1, s_2, \ldots, s_i, \ldots, s_j$, и их выпуклая оболочка $\mathcal{L} = \text{co}, \{l(s_1), l(s_2), l(s_3), \ldots, l(s_i), \ldots, l(s_j)\}$, представлены на рис. 2. Из сказанного следует, что нас могут интересовать лишь правила s_i и s_j которым на рис. 2 соответствуют точки $l(s_i)$ и $l(s_j)$, для которых нет других точек, 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 $\widetilde{s}$, 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 $\widetilde{s}$, даются равенствами

$$\begin{aligned} \mathsf{E}L_1(\widetilde{s}) &= p_i L_1(s_i) + p_j L_1(s_j) \quad \text{при} \quad \vartheta = \vartheta_1, \\ \mathsf{E}L_2(\widetilde{s}) &= p_i L_2(s_i) + p_j L_2(s_j) \quad \text{при} \quad \vartheta = \vartheta_2, \end{aligned} \tag{8.8}$$

[1]Остальные из $N^q - 6$, как и $l(s_4)$, лежат в пятиугольнике $\mathcal{L} = \text{co}\{l(s_1) \ldots, l(s_{N^q})\}$ и не показаны, поскольку не влияют на выбор оптимального правила.

а вероятности p_i, p_j определяются как решение линейных уравнений $\mathsf{E}L_1(\widetilde{s}) = \mathsf{E}L_2(\widetilde{s})$, $p_1 + p_2 = 1$ при условии $p_1 \geqslant 0$, $p_2 \geqslant 0$.

Если точки, представляющие доминирующие правила, не могут быть выделены априори (как $l(s_i)$ и $l(s_j)$ на рис. 2), то следует рассматривать рандомизированные правила, в которых учитываются все N^q чистых правил. Понятно, что, например, множество точек на рис. 2, соответствующих всем рандомизированным правилам, является выпуклой оболочкой $\mathcal{L} = \mathrm{co}\{l(s_1), \ldots, l(s_{N^q})\} = \{\lambda_1 l(s_1), \ldots, \lambda_{N^q} l(s_{N^q})\}$, $l(s_1) \geqslant 0, \ldots, l(s_{N^q}) \geqslant 0$, $l(s_1) + \cdots + l(s_{N^q}) = 1$, натянутой на $l(s_1), \ldots, l(s_{N^q})$. При этом важно отметить что множество $\mathcal{L}$ точек, представляющих все рандомизированные правила, выпукло и замкнуто в $\mathcal{R}^2$.

2. Минимаксное правило решения. Рассмотрим *минимаксное правило действия* s^+, минимизирующее максимальный риск среди $L_i(s)$, $i = 1, \ldots, k$ в (8.7) на множестве S всех известных правил

$$\max_{1\leqslant i\leqslant k} L_i(s^+) = \min_{s\in S} \max_{1\leqslant i\leqslant k} L_i(s). \tag{8.9}$$

Если семейство множеств $U(C) = \{l = (L_1, \ldots, L_k) \in \mathcal{R}^k, \ \max\limits_{1\leqslant i\leqslant k}(L_i) \leqslant c\}$, $c \in \mathcal{R}^1$, см. рис. 2, то правило $s^+ \in S$ определяется как соответствующее точке $l(s^+) \in U(c^+)$, где c^+ — минимальное значение $c \in \mathcal{R}^1$, при котором $l(s^+) \in U(c)$, $c^+ = \min\{c \in \mathcal{R}^1, \{l(s), s \in S\} \cap U(c) \neq \varnothing\}$. Минимальный риск $\max\limits_{1\leqslant i\leqslant k} L_i(s^+) = c^+$.

Пусть рандомизированное правило $\widetilde{s}$ состоит из чистых правил $s_1, s_2, \ldots, s_t$, применяемых с вероятностями $p_1, p_2, \ldots, p_t$. Тогда ожидаемый маргинальный риск, связанный с применением $\widetilde{s}$ в состоянии ϑ_i, равен

$$\mathsf{E}L_i(\widetilde{s}) = \sum_{j=1}^{t} L_i(s_j)p_j = \sum_{j=1}^{t}\sum_{m=1}^{N} l(\vartheta_i, d_m)p_{s_j}(d_m|\vartheta_i)p_j, \quad i = 1, \ldots, k. \tag{8.10}$$

Минимаксное рандомизированное правило $\widetilde{s}^*$, минимизирующее максимальный ожидаемый маргинальный риск среды $\mathsf{E}L_i(\widetilde{s})$, $i = 1, \ldots, k$, в (8.10) определим условием

$$\max_{1\leqslant i\leqslant k} \mathsf{E}L_i(\widetilde{s}^+) = \min_{\widetilde{s}\in\widetilde{S}} \max_{1\leqslant i\leqslant k} \mathsf{E}L_i(\widetilde{s}), \tag{8.11}$$

в котором $\widetilde{S}$ — класс всех рандомизированных правил. Минимаксное рандомизированное правило $\widetilde{s}^+$ определится как соответствующее точке $l(\widetilde{s}^+) \in U(\widetilde{c}^+)$, где $\widetilde{c}^+$ — минимальное значение $c \in \mathcal{R}^1$, при котором $\mathcal{L} \cap U(c) \neq \varnothing$, $c^+ = \min\{c \in \mathcal{R}^1, \mathcal{L} \cap U(c) \neq \varnothing\}$; минимальный ожидаемый риск $\max\limits_{1\leqslant i\leqslant k} \mathsf{E}L_i(\widetilde{s}^+) = \widetilde{c}^+$. Так как $\{l(s), s \in S\} = \{l(s_1), \ldots, l(s_{N^q})\} \subset \mathcal{L} = \mathrm{co}\{l(s_1), \ldots, l(s_{N^q})\}$, то в любом случае $\widetilde{c}^+ \leqslant c^+$ см. рис. 3, поэтому *минимаксным называется рандомизированное правило* $\widetilde{s}^+$*, а задача* (8.9) *и правило* s^+*, доминируемое* $\widetilde{s}^+$*, обычно не рассматриваются.*

Согласно равенствам (8.8) вероятности p_i^*, p_j^*, определяющие искомое рандомизированное правило $\widetilde{s}^+$ на рис. 2 удовлетворяют системе линейных уравнений $\widetilde{c}^+ = p_i^* L_1(s_i) + p_j^* L_1(s_j) = p_i^* L_2(s_i) + p_j^* L_2(s_j)$, $p_1^* + P_2^* = 1$ (при условии $p_i^* \geqslant 0$, $p_j^* \geqslant 0$).

На рис. 3 приведены примеры ситуаций, в которых $c_1^+ = \widetilde{c}_1{}^+$, $c_2^+ < \widetilde{c}_2{}^+$ и $c_3^+ = \widetilde{c}_3{}^+$.

3. Байесовское правило решения. Байесовское правило применяется в случае, когда состояние природы ϑ случайно и известны априорные вероятности состояния природы $p(\vartheta_1), \ldots, p(\vartheta_k)$. *Байесовским называется правило решения* s^* (*называемое также байесовской стратегией*), *минимизирующее ожидаемый* **байесовский риск**

$$L(s^*) = \sum_{i=1}^{k} L_i(s^*)p(\vartheta_i) = \sum_{t=1}^{N}\sum_{i=1}^{k} l(\vartheta_i, d_t)p_s(d_t|\vartheta_i)p(\vartheta_i). \tag{8.12}$$

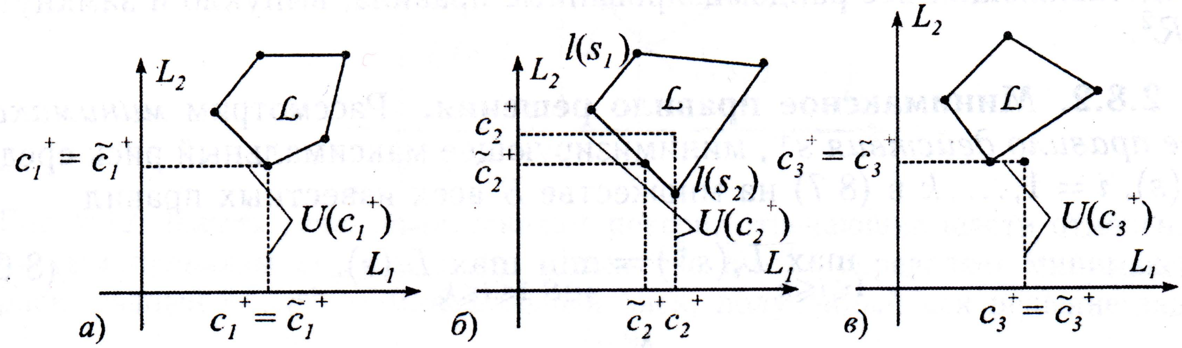


Рис. 3. Минимальные риски c_1^+, c_2^+, c_3^+. Риск $\widetilde{c}_2^+$, соответствующий рандомизированному правилу, основанному на чистых правилах s_1, s_2, $\widetilde{c}_2^+ < c_2^+$.

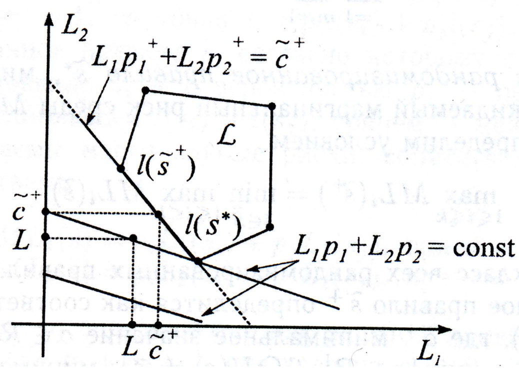


Рис. 4. Байесовское правило s^* отвечает точке $l(s^*)$, через которую проходит прямая $L_1p_1 + L_2p_2 = const$ при значении $const$, равном ожидаемому байесовскому риску L. Прямая $L_1p_1 + L_2p_2 = \widetilde{c}^+$ отвечает априорным вероятностям состояний природы p_1^+, p_2^+, при которых байесовский риск максимален и совпадает с минимаксным $\widetilde{c}^+$.

Рассмотрим графическое решение задачи (8.12) в случае $\Theta = \{\vartheta_1, \vartheta_2\}$. Определим на плоскости $\{(L_1, L_2)\}$ семейство прямых $p(\vartheta_1)L_1 + p(\vartheta_2)L_2 = const$, $p(\vartheta_1) + p(\vartheta_2) = 1$. 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. 4, где $p_1 = p(\vartheta_1)$ и $p_2 = p(\vartheta_2)$, 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 $const = L = p(\vartheta_1)L_1 + p(\vartheta_2)L_2$ 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. 4, задачу (8.11), рис. 2.

4. Байесовская стратегия в случае невозможности наблюдений над природой. Если задано априорное распределение вероятностей состояний природы $p(\vartheta_1), \ldots, p(\vartheta_k)$, но наблюдения над природой невозможны, *то ожидаемый риск, связанный с действием d_j ра-*

[1]Так подобранное распределение вероятностей состояний природы называется наименее благоприятным.

вен

$$L(d_j) = \sum_{i=1}^{k} l(\vartheta_i, d_j) p(\vartheta_i) \tag{8.13}$$

и байесово действие определится из условия

$$L(d_j) \sim \min_j . \tag{8.14}$$

Рассмотрим плоскость $\{(p, L)\}$ и зададим распределение состояний в виде $p(\vartheta_1) = p$, $p(\vartheta_2) = 1 - p$, где p — параметр, $0 \leqslant p \leqslant 1$. Тогда в зависимости от значения p байесовское действие будет d_1, d_2 или d_3, как то показано на рис. 5, где представлены три прямые $(p, L), l(\vartheta_1, d_j)p + l(\vartheta_2, d_j)(1-p) = L$, $j = 1, 2, 3$, определяющие зависимости ожидаемых рисков $L(d_1, p)$, $L(d_2, p)$ и $L(d_3, p)$ обусловленные действиями d_1, d_2 и d_3 от параметра p. Для каждого значения p байесовским будет действие $d_{i(p)}$, для которого $L(d_{i(p)}, p) = \min_j L(d_j)$, см. условие (8.14) и рис. 5. В точках пересечения прямых возможна рандомизация.

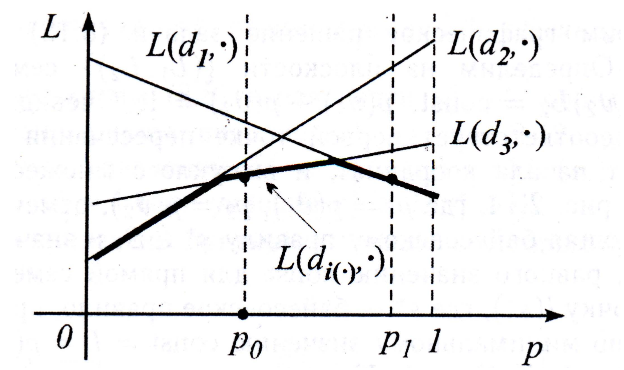


Рис. 5. При $p = p_0$ байесово действие есть d_3, при p_1 есть d_1, $p \in [0, 1]$ — выделенная кусочно-линейная кривая.

Заметим, что в рассмотренной ситуации определяется не правило решения, а непосредственно байесовское 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 наблюдения над природой.*

5. Байесово действие. Рассмотрим вместе с правилом s соответствующее разбиение s множества наблюдений

$$X = D_1 +, \ldots, D_N, \quad D_j = \{x \in X,\ s(x) = d_j\} \quad j = 1, \ldots, N. \tag{8.15}$$

Согласно (8.15) вероятность действия d_m в состоянии природы ϑ_j равна

$$p_s(d_m|\vartheta_j) = P(\{x \in D_m|\vartheta_j\}) = \sum_{x \in D_m} p(x|\vartheta_j), ; \ m = 1, \ldots, N, \quad j = 1, \ldots, k, \tag{8.16}$$

и выражение (8.12) для ожидаемого риска, свойственного правилу $s \in S$ может быть переписано в виде

$$L(s) = \sum_{i=1}^{k} \sum_{t=1}^{N} l(\vartheta_i, d_t) p_s(d_t|\vartheta_i) p(\vartheta_i) = \sum_{t=1}^{N} \sum_{i=1}^{k} l(\vartheta_i, d_t) p(\vartheta_i) \sum_{x \in D_t} p(x|\vartheta_i). \tag{8.17}$$

Как следует из (8.16) и (8.17), для того, чтобы *риск* (8.17) *был минимальным, необходимо и достаточно, чтобы в разбиении* (8.15)

$$D_t \subset \Big\{ x \in X,\ \ \sum_{i=1}^{k} l(\vartheta_i, d_t) p(x|\vartheta_i) p(\vartheta_i) \leqslant \leqslant \sum_{i=1}^{k} l(\vartheta_i, d_j) p(x|\vartheta_i) p(\vartheta_i), \quad j = 1, \ldots, N \Big\},\ \ t = 1, \ldots, N. \tag{8.18}$$

Соответственно, по наблюдению x следует принять решение d_t, если $x \in D_t$, $t = 1, \ldots, N$, (8.18). Если с каждым наблюдением будет связано такое байесовское действие, то ожидаемый риск (8.17) и правило s, соответствующее так определенному разбиению $X = \sum\limits_{j=1}^{N} D_j$, также будут байесовскими.

Покажем, что решение этой задачи сводится к решению предыдущей с помощью байесовского пересчета априорных вероятностей в апостериорные, при условии, что при наблюдении над природой получено значение $x \in X$. Решение d_t при условии, что наблюдено x, приводит к условному ожидаемому риску

$$L_(d_t|x) = \sum_{i=1}^{k} l(\vartheta_i, d_t) p(\vartheta_i|x) = \sum_{i=1}^{k} l(\vartheta_i, d_t) \frac{p(\vartheta_i) p(x|\vartheta_i)}{\sum\limits_{j=1}^{k} p(\vartheta_j) p(x|\vartheta_j)}. \tag{8.19}$$

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) совпадает с (8.13), если в последнем априорную вероятность $p(\vartheta_j)$ заменить на апостериорную $p(\vartheta_j|x)$, $j = 1, \ldots, k$.

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 $s : X \to D$.

Рассмотрим произвольное правило s. Согласно (8.19), правилу $s(\cdot)$ при условии, что наблюдено x, сопутствует условный ожидаемый риск

$$L(s(x)|x) = \sum_{i=1}^{k} l(\vartheta_i, s(x)) p(\vartheta_i|x) = \sum_{i=1}^{k} l(\vartheta_i, s(x)) \frac{p(\vartheta_i) p(x|\vartheta_i)}{\sum\limits_{j=1}^{k} p(\vartheta_j) p(x|\vartheta_j)}. \tag{8.20}$$

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

$$L(s) = \mathsf{E} L(s(\xi)|\xi). \tag{8.21}$$

Действительно, математическое ожидание E можно представить в виде $\mathsf{E} = \sum\limits_{i=1}^{k} p(\vartheta_i) \mathsf{E}_i$, где E_i — оператор условного математического ожидания при условии, что наблюдения x распределены согласно $p(x|\vartheta_i)$, $x \in X$, или, иначе говоря, E_i — оператор условного математического ожидания при условии, что природа находится в состоянии ϑ_i. Пусть $D_j = \{x : s(x) = d_j\}$ и $\chi_{D_j}(x)$ — индикаторная функция D_j, так что

$$P\{s(x) = d_j|\vartheta_i)\} = \sum_{x \in D_j} p(x|\vartheta_i) = \sum_{x \in X} \chi_{D_j}(x) p(x|\vartheta_i), \quad i = 1, \ldots, k, \quad j = 1, \ldots, N. \tag{8.22}$$

Теперь доказательство следует из цепочки равенств:

$$\mathsf{E}L(s(\xi)|\xi) = \sum_{i=1}^{k} p(\vartheta_i)\ \mathsf{E}_i \left(\sum_{j=1}^{N} \chi_{D_j} L(s(x)|x) \right) =$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N} p(\vartheta_i) \sum_{x} p(x|\vartheta_i) \chi_{D_j}(x) L_s(x) =$$

$$= \sum_{j=1}^{N} \sum_{x \in X} \chi_{D_j}(x) \sum_{i=1}^{k} l(\vartheta_i, s(x)) p(\vartheta_i) p(x|\vartheta_i) =$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{k} l(\vartheta_i, d_j) p(d_j|\vartheta_i) p(\vartheta_i) = L(s).$$

где использованы равенства (8.22) и $l(\vartheta_i, s(x)) = l(\vartheta_i, d_j)$,$x \in D_j$, $j = 1, \ldots, N$.

Приведем теперь формальное доказательство того, что *последовательность байесовских действий действительно определяет байесовское правило решения (байесовскую стратегию).*

Теорема.

Пусть $\mathcal{K}$ — класс правил s, таких, что решение $s(x) = d_t$ принимается для $x \in X$, удовлетворяющих неравенствам

$$L(d_t|x) \leqslant L(d_j|x), \quad j = 1, \ldots, N,$$

где $L(d_t|x)$ — условная потеря, определенная в (8.19). Тогда для всякого правила $s \in \mathcal{K}$ выполняется неравенство $L(s) \leqslant L(s')$, где s' — произвольное правило. Иными словами всякое правило из $\mathcal{K}$ является байесовским.

Доказательство. Обозначим $l(x) = \min\limits_t L(d_t|x)$. Тогда $\mathsf{E}l(\xi) = \mathsf{E}L(s(\xi)|\xi)$ при $s \in \mathcal{K}$. Действительно,

$$\mathsf{E}l(\xi) = \mathsf{E} \sum_{j=1}^{N} \chi_{D_j}(x) \min_t L(d_t|\xi) = \mathsf{E} \sum_{j} \chi_{D_j} L(s(\xi)|\xi) =$$

$$= \mathsf{E} \sum_{j} \chi_{D_j} L(s(\xi)|\xi) = \mathsf{E} L_{s(x)}(x) = L(s),$$

где использовано, что при $x \in D_t$

$$L(d_t|x) \leqslant L(d_i|x) \ \Rightarrow \ l(x) = L(d_t|x), \quad s(x) = d_t.$$

Отсюда следует

$$L(s) = \mathsf{E} \min_t L(d_t|\xi) \leqslant \mathsf{E} \min_t L(s^*(\xi)|\xi) = L(s^*).$$

□

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. Байесовское действие теперь является байесовским решением.

7. Правило решения, минимизирующая ожидаемое число ошибок. В задаче решения риск чаще всего оценивается посредством количества ошибок. Рассмотрим в этой связи некоторые характерные решающие правила.

Зададим функцию, определяющую риск потерь, условием

$$l(\vartheta_i, d_j) = 1 - \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Тогда в (8.19)

$$L_{d_t}(x) = \sum_i l(\vartheta_i, d_t) p(\vartheta_i|x) = 1 - p(\vartheta_t|x) \tag{8.23}$$

и соответственно ожидаемый риск дается равенством (8.12)

$$L(s) = \sum_{i,j=1}^{k} (1 - \delta_{ij}) p(d_j|\vartheta_i) p(\vartheta_i) = \sum_{i=1}^{k} p(\vartheta_i)(1 - p(d_i|\vartheta_i)). \tag{8.24}$$

Но последнее выражение совпадает *с математическим ожиданием доли ошибочных решений*. Следовательно, *байесовское правило решения в этом случае совпадает с правилом, минимизирующей ожидаемое число ошибок решения*.

По данному наблюдению $x \in X$ байесовская стратегия решения 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 наблюдено

$$\begin{gathered} \xi = x \in D_t \subset \{x \in X : s(x) = d_t\} = \\ = \{x \in X : p(\vartheta_t|x) \geqslant p(\vartheta_j|x),\ j = 1, \dots, k\}, \ \ t = 1, \dots, k. \end{gathered} \tag{8.25}$$

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 25 можно также записать

$$D_t = \{x :\ s(x) = d_t\} = \{x :\ p(\vartheta_t|x) \geqslant p(\vartheta_j|x),\ j = 1, \dots, N\}. \tag{8.26}$$

Заметим, что вероятность верно опознать состояние природы ϑ_t равна

$$P\{s(\xi) = d_t|\vartheta_t\} = \sum_{x \in D_t} p(x|\vartheta_t) = p(d_t|\vartheta_t), \ \ t = 1, \dots, k,$$

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 состояний природы одинаковы, то в (8.26)

$$D_t = \{x \in X :\ p(x|\vartheta_t) \geqslant p(x|\vartheta_j)\}, \ \ t = 1, \dots, k,$$

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.

Литература

1. Ю.П.Пытьев, И.А.Шишмарев. Теория вероятностей, математическая статистика и элементы теории возможностей для физиков. Физический факультет МГУ им. М.В.Ломоносова, 2010.

2. А.И.Кибзун, Е.Р.Горяинов, А.В.Наумов, А.Н.Сиротин. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. Физматлит, 2002.

3. Г.П.Климов. Теория вероятностей и математическая статистика. Изд-во МГУ, 1983.

4. А.А.Боровков. Математическая статистика. — М.: Наука, 1984.

5. Леман Э. Проверка статистических гипотез. — М.: Наука, 1979.

6. Д.А.Коршунов, Н.И.Чернова. Сборник задач и упражнений по математической статистике. Новосибирск. Изд-во Института математики, 2004.

7. Д.М.Чибисов, В.И.Пагурова. Задачи по математической статистике. — М.: Изд-во Моск. ун-та, 1990.

Содержание