НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МЕХАНИКО–МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

кафедра теории вероятностей и математической статистики

И. С. БОРИСОВ

# ЛЕКЦИИ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Новосибирск – 2010

# Введение

***Математическая статистика*** — раздел математики, который посвящен методам обработки результатов реальных стохастических экспериментов. Математическая статистика тесно связана с теорией вероятностей, поскольку также имеет дело со стохастическим экспериментом. Однако математическая статистика занимается в известном смысле обратными задачами теории вероятностей. Как нам хорошо известно, в теории вероятностей исходным объектом было вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, и мы могли говорить о вероятностях появления тех или иных событий, связанных, скажем, с последовательностью независимых испытаний. В то же время, исходными данными в математической статистике являются конечные наборы результатов уже проведенных в одних и тех же условиях стохастических экспериментов (стало быть, независимых и одинаково распределенных) $x_1 = \xi_1(\omega)$,..., $x_n = \xi_n(\omega)$, где $\omega$ — некоторый элементарный исход вероятностного пространства $\Omega$, описывающего всю серию проводимых испытаний, а величины $x_i$ могут принимать значения в любом измеримом пространстве $\mathfrak{X}$. При этом *распределение отдельного эксперимента* $P(A) = \mathbf{P}(\xi_1 \in A)$ *предполагается неизвестным частично или полностью*. Задача математической статистики как раз и состоит в восстановлении (или оценке) распределения $P(\cdot)$ как можно более точно.

Курс состоит из двух разделов:

1. Теория построения оценок неизвестных параметров наблюдаемых распределений.

2. Проверка статистических гипотез.

**Определение.** Вектор $X = (x_1, \ldots, x_n)$ называется *выборкой объема $n$ из неизвестного распределения*. Декартова степень $\mathfrak{X}^n$ с соответствующей $\sigma$-алгеброй подмножеств называется *выборочным пространством*.

**Замечание.** Как мы видим, с одной стороны, выборка — *неслучайный* вектор

$$X = (x_1, \ldots, x_n) = (\xi_1(\omega), \ldots, \xi_n(\omega)),$$

представляющий собой обычную $n$-ку чисел или элементов более сложной природы. Повторно проведя $n$ раз эксперимент в тех же условиях, получим новые значения $(\xi_1(\omega'), \ldots, \xi_n(\omega'))$, для которых, вообще говоря, $\xi_i(\omega') \neq \xi_i(\omega)$ хотя бы при одном $i$. Так что если нас будут интересовать всевозможные значения выборочного вектора и их распределение (если иметь в виду многократные серии одних и тех же испытаний), то выборку $X$ будем интерпретировать как *$n$-мерный случайный вектор с независимыми одинаково распределенными координатами*. При этом, следуя традиции математической статистики, в этом случае мы оставим прежнее обозначение и для случайного выборочного вектора. Так что запись типа $\mathbf{P}(x_1 \in A_1, \ldots, x_n \in A_n)$ не должна приводить к недоразумению.

Стоит также отметить, что по большому счету суть всех статистических процедур сводится к поиску "почти достоверных" измеримых подмножеств в выборочном пространстве, т. е. подмножеств, в которые выборочный вектор $X$ попадает с вероятностью, близкой к 1. При этом применяется *основной принцип статистики*: имеющаяся в вашем распоряжении "неслучайная" выборка $(x_1, \ldots, x_n)$ *наверняка* находится именно в этом подмножестве.

**Определение.** *Эмпирическое распределение, построенное по выборке объема $n$*:

$$\mathbf{P}_n^*(A) = \frac{\#\{\, x_i \mid x_i \in A \,\}}{n}.$$

Легко видеть, что $\mathbf{P}_n^*$ — дискретная вероятностная мера с атомами в точках $\{x_i;\ i \leq n\}$ и равными массами $1/n$.

**Определение.** *Выборочной характеристикой* называется измеримый $k$-мерный (принимающий значения в пространстве $\mathbb{R}^k$) функционал $G(\mathbf{P}_n^*)$.

**Пример.** $G(\mathbf{P}_n^*) = \int\limits_{\mathfrak{X}} f(x)\,\mathbf{P}_n^*(dx)$ — выборочный момент "порядка $f$", где $f(x)$ — скалярная измеримая функция. Очевидно, $G(\mathbf{P}_n^*) = \frac{1}{n}\sum\limits_{i=1}^{n} f(x_i)$. В математической статистике для этой величины используется обозначение $\overline{f(x)}$. Пусть $\mathfrak{X} = \mathbb{R}$. Если $f(x) = x^k$, то говорят о *$k$-ом выборочном моменте*. При этом $\overline{x}$ традиционно называют *выборочным средним*.

Приведем еще пару примеров:

$$G_1(\mathbf{P}_n^*) = g_1(\overline{f_1(x)}, \dots \overline{f_k(x)}, \overline{x}), \quad G_2(\mathbf{P}_n^*) = g_2(\overline{f_1(x) - f_2(\overline{x})}),$$

где $g_j(\cdot)$ и $f_j(t)$ — произвольные измеримые функции, заданные на соответствующих пространствах. Скажем, *выборочные центральные моменты* $\overline{(x-\overline{x})^k}$, в частности, *выборочная дисперсия* $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^2} - (\overline{x})^2$ имеют как раз указанную структуру.

## Основная теорема статистики

Далее предполагается, что наблюдения $x_i$ принимают значения в произвольном измеримом пространстве $(\mathfrak{X}, \mathcal{B})$, где $\mathfrak{X}$ — множество произвольной природы, а $\mathcal{B}$ — некоторая $\sigma$-алгебра его подмножеств.

**Теорема** (ЗБЧ для эмпирического распределения). *Для каждого множества $A \in \mathcal{B}$ при $n \to \infty$ имеет место предельное соотношение*

$$\mathbf{P}_n^*(A) \xrightarrow[\text{п.н.}]{} \mathcal{P}_{x_1}(A)$$

.

Доказательство. Заметим, что $\#\{\, x_i \mid x_i \in A \,\} = \sum\limits_{i=1}^{n} I(x_i \in A)$, где $I(x_i \in A)$ — бернуллиевские случайные величины. Тогда в силу УЗБЧ

$$\frac{\#\{\, x_i \mid x_i \in A \,\}}{n} = \sum_{i=1}^{n} \frac{I(x_i \in A)}{n} \xrightarrow[\text{п.н.}]{} \mathbf{P}(x_i \in A) \equiv \mathcal{P}_{x_1}(A).$$

Другими словами, для любого фиксированного множества $A$ эмпирическое распределение является сильно состоятельной оценкой для неизвестного истинного распределения. □

**Замечание.** Приведенная теорема имеет дело с поточечной сходимостью рассматриваемых функций множества. Покажем, что равномерная аппроксимация в рамках этой теоремы, вообще говоря, не имеет места. Равномерная аппроксимация означает, что

$$\sup_{A \in \mathcal{B}} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \xrightarrow[\text{п.н.}]{} 0.$$

Предположим, что $\mathcal{B}$ содержит все конечные подмножества $\mathfrak{X}$, и распределение $x_1$ не имеет ни одного атома. Пусть $A^* = \{\, x_1, x_2, \dots, x_n \,\}$, тогда $\mathbf{P}_n^*(A^*) = 1$ и $\mathcal{P}_{x_1}(A^*) = 0$. Следовательно,

$$\sup_{A \in \mathcal{B}} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \geq |\mathbf{P}_n^*(A^*) - \mathcal{P}_{x_1}(A^*)| \equiv 1.$$

# Теоремы Гливенко — Кантелли.

Пусть $x_1 \in (\mathfrak{X}, \mathcal{B})$, и $\mathcal{A} \subseteq \mathcal{B}$ - класс измеримых подмножеств. Будем говорить, что для класса $\mathcal{A}$ выполняется *теорема Гливенко — Кантелли*, если

$$\sup_{A\in\mathcal{A}} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \underset{\text{п.н.}}{\longrightarrow} 0 \text{ при } n \to \infty.$$

**Определение.** Класс $\mathcal{A}$ имеет *конечную двустороннюю $\varepsilon$-энтропию относительно распределения $\mathcal{P}_{x_1}$*, если $\forall \varepsilon > 0$ существует конечный набор $\{A_i^{\pm}\}_{i=1}^{N(\varepsilon)} \in \mathcal{B}$ со следующим свойством: $\forall A \in \mathcal{A} \quad \exists A_{i_0}^{\pm}$ такие, что $A_{i_0}^- \subseteq A \subseteq A_{i_0}^+$ и $\mathcal{P}_{x_1}(A_{i_0}^+ \setminus A_{i_0}^-) \leqslant \varepsilon$.

**Замечание.** Вспомним некоторые определения анализа. Пусть $(S, \varrho)$ — метрическое пространство. Множество $M \subseteq S$ называется $\varepsilon$-сетью для $S$, если для любой точки $x \in S$ существует хотя бы одна точка $y \in M$ такая, что $\varrho(x, y) \leqslant \varepsilon$.

Множество $X \subseteq S$ называется вполне ограниченным, если при любом $\varepsilon > 0$ для него существует *конечная* $\varepsilon$-сеть. При этом определяется число $H(\varepsilon) = \log\min N(\varepsilon)$, называемое $\varepsilon$-энтропией множества $S$.

**Пример.** Пусть $S = \{A\} \equiv \mathcal{A}$ — класс множеств. Это пространство можно сделать метрическим, введя бинарный функционал $\varrho(A, B) = \mathcal{P}_{x_1}(A \triangle B)$, где $A \triangle B \equiv (A \setminus B) \cup (B \setminus A)$.

**Упражнение.** *Проверить, что введенный функционал $\varrho(A, B)$ удовлетворяет всем аксиомам метрики, кроме одной: $\varrho(A, B) = 0 \Leftrightarrow A = B$. Чтобы получить метрику, надо провести факторизацию пространства, отождествив все множества $A$ и $B$, для которых $\mathcal{P}_{x_1}(A \triangle B) = 0$.*

Заметим, что определение класса $\mathcal{A}$ с двусторонней $\varepsilon$-энтропией похоже на определение вполне ограниченного класса множеств. Отличие только в слове "двусторонняя". Скажем, для общих метрических решеток это означает, что по точкам $\varepsilon$-сети мы можем "подобраться" с двух сторон — "справа" и "слева" — к любой точке рассматриваемого множества на расстояние не более $\varepsilon$. Поэтому для краткости всюду в дальнейшем будем называть класс $\mathcal{A}$ с указанным свойством *вполне ограниченным относительно $\mathcal{P}$*.

**Теорема** (*Обобщенная теорема Гливенко — Кантелли*). *Если класс $\mathcal{A}$ вполне ограничен относительно $\mathcal{P}_{x_1}$, то верна теорема Гливенко — Кантелли.*

Доказательство. Пусть $\varepsilon > 0$. Обозначим $\triangle_n = \sup\limits_{A\in\mathcal{A}} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)|$. Подкласс множеств $B_i = \{\, A \in \mathcal{A} \mid A_i^- \subseteq A \subseteq A_i^+ \,\}$ будем называть *брикетом* (иногда в литературе для него используют обозначение $[A_i^-, A_i^+]$). Заметим, что $\mathcal{A} = \bigcup\limits_{i\leqslant N(\varepsilon)} B_i$ — покрытие (но не разбиение) класса $\mathcal{A}$. Тогда $\triangle_n = \max\limits_{i\leqslant N(\varepsilon)} \sup\limits_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)|$. Утверждение теоремы будет доказано, если $\limsup\limits_{n\to\infty} \sup\limits_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \leqslant \delta(\varepsilon)$, и $\delta(\varepsilon) \to 0$ при $\varepsilon \to 0$. Имеем

$$\begin{aligned}
\sup_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| &= \sup_{A\in B_i} |\mathbf{P}_n^*(A) \pm \mathbf{P}_n^*(A_i^-) - \mathcal{P}_{x_1}(A) \pm \mathcal{P}_{x_1}(A_i^-)| \leqslant \\
&\leqslant \sup_{A\in B_i} \left(|\mathbf{P}_n^*(A) - \mathbf{P}_n^*(A_i^-)| + |\mathbf{P}_n^*(A_i^-) - \mathcal{P}_{x_1}(A_i^-)| + |\mathcal{P}_{x_1}(A) - \mathcal{P}_{x_1}(A_i^-)|\right) \leqslant \\
&\leqslant \sup_{A\in B_i} |\mathbf{P}_n^*(A) - \mathbf{P}_n^*(A_i^-)| + |\mathbf{P}_n^*(A_i^-) - \mathcal{P}_{x_1}(A_i^-)| + \sup_{A\in B_i} |\mathcal{P}_{x_1}(A) - \mathcal{P}_{x_1}(A_i^-)|.
\end{aligned}$$

Первое неравенство в этой цепочке следствие применения неравенства треугольника для модуля, второе — для sup. Так как $A_i^-$ минимальная точка брикета $B_i$ в смысле частичного порядка, а $A_i^+$ — максимальная, то $0 \leqslant \mathbf{P}_n^*(A) - \mathbf{P}_n^*(A_i^-) \leqslant \mathbf{P}_n^*(A_i^+ \setminus A_i^-)$. Аналогично, $0 \leqslant \leqslant \mathcal{P}_{x_1}(A) - \mathcal{P}_{x_1}(A_i^-) \leqslant \mathcal{P}_{x_1}(A_i^+ \setminus A_i^-) \leqslant \varepsilon$. Тогда

$$\sup_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \leqslant \mathbf{P}_n^*(A_i^+ \setminus A_i^-) + |\mathbf{P}_n^*(A_i^-) - \mathcal{P}_{x_1}(A_i^-)| + \mathcal{P}_{x_1}(A_i^+ \setminus A_i^-).$$

Согласно УЗБЧ для эмпирического распределения $\limsup\limits_{n\to\infty} |\mathbf{P}_n^*(A^-) - \mathcal{P}_{x_1}(A^-)| = 0$ и

$$\limsup_{n\to\infty} \mathbf{P}_n^*(A_i^+ \setminus A_i^-) = \lim_{n\to\infty} \mathbf{P}_n^*(A_i^+ \setminus A_i^-) = \mathcal{P}_{x_1}(A_i^+ \setminus A_i^-) \leqslant \varepsilon.$$

В итоге получаем

$$\limsup_{n\to\infty} \sup_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| \leqslant 2\varepsilon.$$

Заметим, что правая часть неравенства не зависит от номера $i$, так что

$$\limsup_{n\to\infty} \max_{i\leqslant N(\varepsilon)} \sup_{A\in B_i} |\mathbf{P}_n^*(A) - \mathcal{P}_{x_1}(A)| = \limsup_{n\to\infty} \triangle_n \leqslant 2\varepsilon.$$

В силу произвольности $\varepsilon$ выполнено $\triangle_n \underset{\text{п.н.}}{\longrightarrow} 0$ при $n \to \infty$. □

**Определение.** Пусть $(x_1, \ldots, x_n) \in \mathbb{R}^n$. *Эмпирической функцией распределения $F_n^*(t)$, построенной по выборке объема $n$*, называется функция $F_n^*(t) = \mathbf{P}_n^*(-\infty, t)$. Таким образом, $F_n^*(t)$ — это доля выборочных наблюдений, попавших строго левее точки $t$.

Если упорядочить все элементы выборки $(x_1, \ldots, x_n)$ в порядке возрастания, то получится новая последовательность — *вариационный ряд*: $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$. Случайная величина $x_{(k)}$ называется *$k$-й порядковой статистикой* .

**Теорема** (Классическая теорема Гливенко — Кантелли). *Для любой функции распределения $F_{x_1}$ наблюдаемой случайной величины*

$$\sup_{t\in\mathbb{R}} |F_n^*(t) - F_{x_1}(t)| \underset{\text{п.н.}}{\longrightarrow} 0$$

*при $n \to \infty$.*

Доказательство. Заметим, что в нашем случае $(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B})$. Введем в рассмотрение класс $\mathcal{A} = \{(-\infty, t);\ t \in \mathbb{R}\}$. Тогда теорема будет доказана, если класс $\mathcal{A}$ вполне ограничен относительно любой функции распределения $F_{x_1}$. Мы ограничимся рассмотрением непрерывной функции распределения $F_{x_1}$. Укажем алгоритм построения конечной двусторонней $\varepsilon$-сети. Разобьем отрезок $[0,1]$ на оси ординат на $m$ равных частей и полагаем $\varepsilon = \frac{1}{m}$. Вспомним определение квантильного преобразования: $F_{x_1}^{-1}(t) = \inf\{y \mid F_{x_1}(y) \geqslant t\}$ при $t \in (0,1)$. Заметим, что $\{F_{x_1}^{-1}(\frac{k}{m})\}_{k=1}^m$ — конечный набор точек. Кроме того, положим $F_{x_1}^{-1}(1) = \infty$ и $F_{x_1}^{-1}(0) = -\infty$. С помощью набора $\{F_{x_1}^{-1}(\frac{k}{m})\}_{k=0}^m$ можно построить искомую двустороннюю $\varepsilon$-сеть. Для каждого $t \in \mathbb{R}$ существует $k_0$ такое, что $F_{x_1}^{-1}(\frac{k_0}{m}) \leqslant t < F_{x_1}^{-1}(\frac{k_0+1}{m})$. Рассмотрим два луча:

$$A_{k_0}^- = \Big(-\infty, F_{x_1}^{-1}\Big(\frac{k_0}{m}\Big)\Big);$$

$$A_{k_0}^+ = \Big(-\infty, F_{x_1}^{-1}\Big(\frac{k_0+1}{m}\Big)\Big).$$

Очевидно, что $\mathcal{P}_{x_1}(A_{k_0}^+ \setminus A_{k_0}^-) \leqslant \varepsilon$. Таким образом классическая теорема Гливенко — Кантелли для непрерывной функции распределения доказана. □

**Замечание.** В случае разрывной функции распределения построенный класс $\{A_k^\pm\}_{k=0}^m$ не будет удовлетворять определению вполне ограниченного класса множеств. Пусть $t$ — точка разрыва и $\varepsilon \ll h$, где $h$ — величина разрыва. Тогда $\mathcal{P}_{x_1}(A_{k_0}^+ \setminus A_{k_0}^-) \geqslant h$, где $A_{k_0}^-$ и $A_{k_0}^+$ — *любая* пара лучей вида $(-\infty, t_k)$, «окоймляющих» с двух сторон луч $A = (-\infty, t)$.

Чтобы установить справедливость теоремы предлагается класс $\{A_k^\pm\}_{k=0}^m$ брать немного другими, а именно добавить к только что введенным открытым интервалом еще и замкнутые справа:

$$A_{k_0}^- = \Big(-\infty, F_{x_1}^{-1}\Big(\frac{k_0}{m}\Big)\Big];$$

$$A_{k_0}^{+} = \Big(-\infty, F_{x_1}^{-1}\Big(\frac{k_0+1}{m}\Big)\Big].$$

**Упражнение.** *Доказать, что в результате объединения этих двух конечных классов получится двусторонняя конечная $\varepsilon$-сеть для любой функции распределения.*

**Теорема**. *Пусть класс $\mathcal{A}$ вполне ограничен относительно распределений $\mathcal{P}_1$ и $\mathcal{P}_2$. Тогда для каждого $\alpha \in (0,1)$ класс $\mathcal{A}$ вполне ограниченно относительно смеси $\alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2$.*

Доказательство. Пусть $\varepsilon > 0$. Построим $\{A_i^{\pm}\}_{i\leqslant N_1(\varepsilon)}$ и $\{B_j^{\pm}\}_{j\leqslant N_2(\varepsilon)}$ — двусторонние конечные $\varepsilon$-сети относительно распределений $\mathcal{P}_1$ и $\mathcal{P}_2$ соответственно. Определим новые элементы $C_l^{\pm}$ по формулам:

$$C_l^{+} \in \{A_i^{+} \cap B_j^{+}, i \leqslant N_1(\varepsilon), j \leqslant N_2(\varepsilon)\};$$

$$C_l^{-} \in \{A_i^{-} \cup B_j^{-}, i \leqslant N_1(\varepsilon), j \leqslant N_2(\varepsilon)\}.$$

Из множества $\{C_l^{\pm}\}$, содержащего не более $2N_1N_2$ элементов, можно получить необходимую двустороннюю $\varepsilon$-сеть для любого $\alpha$.

Пусть $A \in \mathcal{A}$. Выберем номера $i_0$ и $j_0$ так, чтобы $A_{i_0}^{-} \subseteq A \subseteq A_{i_0}^{+}$ и $B_{j_0}^{-} \subseteq A \subseteq B_{j_0}^{+}$. Заметим, что $A_{i_0}^{-} \cup B_{j_0}^{-} \subseteq A \subseteq A_{i_0}^{+} \cap B_{j_0}^{+}$. Положим $C_{l_0}^{+} = A_{i_0}^{+} \cap B_{j_0}^{+}$, $C_{l_0}^{-} = A_{i_0}^{-} \cup B_{j_0}^{-}$. Необходимо проверить, что мера симметрической разности не превосходит $\varepsilon$.

Имеем

$$\mathcal{P}(C_{l_0}^{+} \setminus C_{l_0}^{-}) = \alpha\mathcal{P}_1(C_{l_0}^{+} \setminus C_{i_0}^{-}) + (1-\alpha)\mathcal{P}_2(C_{l_0}^{+} \setminus C_{l_0}^{-})$$
$$\leqslant \alpha\mathcal{P}_1(A_{i_0}^{+} \setminus A_{i_0}^{-}) + (1-\alpha)\mathcal{P}_2(B_{j_0}^{+} \setminus B_{j_0}^{-}) \leqslant \alpha\varepsilon + (1-\alpha)\varepsilon \leqslant \varepsilon.$$

Первое неравенство в этой цепочке следует из того, что $C_{l_0}^{+} \subseteq A_{i_0}^{+}$ и $B_{j_0}^{-} \subseteq C_{l_0}^{-}$; $C_{l_0}^{+} \subseteq B_{j_0}^{+}$ и $B_{j_0}^{-} \subseteq C_{l_0}^{-}$, а также из монотонности вероятностной меры $\mathcal{P}$. □

**Теорема**. *Класс $\mathcal{B}$ вполне ограничен относительно любого дискретного распределения $\mathcal{P}_{\partial}$.*

**Замечание.** Согласно теореме Лебега любое распределение в измеримом пространстве может быть представлено в виде смеси: $\mathcal{P} = \alpha\mathcal{P}_{\text{д}} + (1-\alpha)\mathcal{P}_{\text{н}}$, где $\mathcal{P}_{\text{д}}$ — дискретное распределение; $\mathcal{P}_{\text{н}}$ — распределение, не имеющее атомов (в одномерном случае $\mathcal{P}_{\text{н}}$ — классическое непрерывное распределение). Тогда на основании двух последних теорем все дальнейшее исследование можно сводить к неатомарным распределениям. В частности, можно не доказывать отдельно классическую теорему Гливенко — Кантелли для случая разрывной функции распределения.

Доказательство. В случае дискретного распределения имеем не более чем счетный набор атомов $\{a_i\}$ таких, что $\mathcal{P}_{\text{д}}(a_i) = p_i > 0$. Для любого множества $A \subseteq \mathcal{B}$ имеем $\mathcal{P}_{\text{д}}(A) = \sum\limits_{i:a_i\in A} p_i$. Вместо $\mathcal{B}$ рассмотрим класс $\mathcal{B}_{\text{д}} = \{\, A \cap \{a_i\} \mid A \in \mathcal{B}\}$, состоящий из всевозможных подмножеств совокупности $\{a_i\}$. Иногда для этого класса используют стандартное обозначение теории множеств $\mathcal{B}_{\text{д}} = 2^{\{a_i\}}$. Для конечных множеств это обозначение оправданно, так как известно, что $\#(2^{\{a_i\}_{i\leqslant N}}) = 2^N$. Для любого $\varepsilon > 0$ найдется $N(\varepsilon)$ такое, что $\sum\limits_{i>N(\varepsilon)} p_i \leqslant \varepsilon$. Необходимо построить конечный набор пар $\{A_i^{\pm}\}_{i\leqslant N(\varepsilon)} \in \mathcal{B}_{\text{д}}$ со следующим свойством: для любого $A \in \mathcal{B}_{\text{д}}$ найдется номер $i$ такой, что $A_i^{-} \subseteq A \subseteq A_i^{+}$ и $\mathcal{P}_{\text{д}}(A_i^{+} \setminus A_i^{-}) \leqslant \varepsilon$. Положим $\{A_i^{-}\}_{i\leqslant 2^{N(\varepsilon)}} = 2^{\{a_1,\ldots,a_{N(\varepsilon)}\}}$. Определим правые концы брикетов по формуле $A_i^{+} = A_i^{-} \cup \{a_{N(\varepsilon)+1}, a_{N(\varepsilon)+2}, \ldots\}$. Для каждого $A \in \mathcal{B}_{\text{д}}$ выберем левый конец брикета: $A_i^{-} = A \cap \{a_1, \ldots, a_{N(\varepsilon)}\}$. Таким образом, искомая конечная двусторонняя $\varepsilon$- сеть построена, так как $\mathcal{P}_{\text{д}}(A_i^{+} \setminus A_i^{-}) = \sum\limits_{i>N(\varepsilon)} p_i \leqslant \varepsilon$. □

Рассмотрим многомерный вариант классической теоремы Гливенко — Кантелли для векторнозначных наблюдений $x_i \in \mathbb{R}^k$ с произвольной совместной функцией распределения

$$F(t_1, \ldots, t_k) = P(x_1^{(1)} < t_1, \ldots, x_1^{(k)} < t_k).$$

Определим эмпирическую функцию распределения по формуле

$$F_n^*(t_1, \ldots, t_k) = \frac{1}{n} \sum_{i=1}^{n} I(x_i^{(1)} < t_1, \ldots, x_i^{(k)} < t_k).$$

**Теорема.** *Имеет место предельное соотношение*

$$\sup_{\bar{t} \equiv (t_1, \ldots, t_k) \in \mathbb{R}^k} |F_n^*(\bar{t}) - F(\bar{t})| \underset{\text{п.н.}}{\longrightarrow} 0 \text{ при } n \to \infty.$$

**Упражнение.** *Доказать теорему Гливенко — Кантелли в* $\mathbb{R}^k$.

## Понятие оценки неизвестного параметра.

**Определение.** *Параметром* наблюдаемого распределения называется значение того или иного функционала от этого распределения: $\theta = G(\mathcal{P}_{x_1}) \in \mathbb{R}^k$, где $\mathcal{P}_{x_1}(A) = \mathbf{P}(x_1 \in A)$.

**Определение.** *Параметрическое семейство распределений* $\{\mathcal{P}_\theta\}_{\theta \in \Theta}$ — это класс распределений известной функциональной формы, содержащей один или несколько скалярных параметров. При этом множество $\Theta$ называется *параметрическим*.

Ясно, что эти два определения тесно связаны между собой. Обычно в статистических задачах подразумевается, что параметрическое множество находится во взаимно-однозначном соответствии с параметрическим семейством распределений — иначе оценить параметр по выборке, вообще говоря, будет невозможно. Поэтому параметры таких параметрических семейств удовлетворяют первому из вышеприведенных определений.

**Примеры.** $\{\pi_\lambda\}_{\lambda > 0}$ — семейство пуассоновских распределений с параметром $\theta = \lambda$. Для этого семейства $\lambda = \mathbf{E}x_1$ — интегральный функционал от наблюдаемого распределения. Аналогичные представления имеют место для параметров нормального распределения. Квантиль порядка $t$ неизвестного распределения $F$ при любом фиксированном $t \in [0, 1]$ также является примером такого вида функционалов.

**Определение.** *Оценка* $\theta_n^* = \theta_n^*(x_1, \ldots, x_n)$ *неизвестного параметра* $\theta$ — это выборочная характеристика $\theta_n^* = G(\mathbf{P}_n^*)$, которая в том или ином смысле приближает неизвестный параметр $\theta$.

**Определение.** Оценка $\theta_n^*$ называется *состоятельной,* если $\theta_n^* \underset{p}{\to} \theta$ при $n \to \infty$.

**Определение.** Оценка $\theta_n^*$ называется *сильно состоятельной,* если $\theta_n^* \underset{\text{п.н.}}{\longrightarrow} \theta^*$ при $n \to \infty$.

**Замечание.** Пусть $\mathfrak{X}^\infty$ — пространство последовательностей $(x_1, x_2, \ldots)$, а $X_\infty$ — выборка бесконечного объема. Будем считать, что вектор $X_n$ получен в результате проецирования $X_\infty$ на первые $n$ координат. Сходимость почти наверное понимается относительно распределения $\mathbf{P}$ в $(\mathfrak{X}^\infty, \mathcal{B}^\infty, \mathbf{P})$. В силу теоремы Колмогорова о согласованных распределениях, такое распределение всегда существует.

**Пример.** Пусть $\mathbf{E}|f(x_1)| < \infty$. Используя ЗБЧ в форме Хинчина, получаем:

$$\frac{1}{n} \sum_{i=1}^{n} f(x_1) = \overline{f(x)} = \theta_n^* \underset{p}{\to} \mathbf{E}f(x_1) = \theta.$$

**Замечание.** В курсе теории вероятностей УЗБЧ был доказан в предположении существования $\mathbf{E}|\xi_1|^4$. Верна более сильная

**Теорема** (УЗБЧ Колмогорова). *Для того, чтобы* $\frac{S_n}{n} \xrightarrow[\text{п.н.}]{} a$, *необходимо и достаточно, чтобы существовало математическое ожидание* $\mathbf{E}\xi_1 = a$.

Следовательно, выборочное среднее является сильно состоятельной оценкой для истинного среднего (математического ожидания), если только $\mathbf{E}|f(x_1)| < \infty$. Выборочная дисперсия $S^2$ — сильно состоятельная оценка для истинной дисперсии, если $\mathbf{E}x_1^2 < \infty$.

# Построение оценок неизвестных параметров.

Пусть $\theta$ — неизвестный параметр наблюдаемого распределения (скалярный или векторный). Он является некоторым функционалом от наблюдаемого распределения: $\theta = G(\mathcal{P}_{x_1})$. Распределение может быть неизвестно полностью или частично. Типичный пример

$$\theta = \mathbf{E}x_1^m = \int\limits_{\mathbb{R}} t^m \, \mathcal{P}_{x_1}(dt) = \int\limits_{\mathbb{R}} t^m \, dF_{x_1}(t).$$

**Определение.** *Оценка* $\theta_n^*(X)$ — это функция от выборки (т.е. *статистика*), которая в том или ином смысле приближает неизвестный параметр $\theta$.

Существует общий приём построения оценок — *метод подстановки*. Этот метод в качестве оценки $\theta_n^*$ предписывает взять функцию $\theta_n^* = G(\mathbf{P}_n^*)$. Одна из реализаций такого подхода называется *методом моментов*, к изучению которого мы переходим.

Рассмотрим сначала одномерное параметрическое семейство распределений $\{F_\theta\}_{\theta\in\Theta}$, где $\Theta \subseteq \mathbb{R}$. Введём функцию

$$m_g(\theta) = \int g(t) \, dF_\theta(t), \ \theta \in \Theta,$$

где $g$ — так называемая *пробная* функция. Мы хотим построить состоятельную оценку для $\theta$. Для этого метод моментов предлагает вместо функции распределения $F_\theta(t)$ использовать эмпирическую функцию распределения $F_n^*(t)$ :

$$\int g(t) \, dF_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \overline{g(x)}.$$

Согласно УЗБЧ имеем $\displaystyle\frac{1}{n}\sum_{i=1}^{n} g(x_i) \xrightarrow[\text{п.н.}]{} \mathbf{E}g(x_1) = m_g(\theta_0)$. Значит, при большом количестве наблюдений должно выполняться приближенное равенство: $m_g(\theta_0) \approx \overline{g(x)}$, где $\overline{g(x)}$ — известная функция. Если уравнение $m_g(\theta_0) = \overline{g(x)}$ разрешимо, то его решение $\theta_n^*$ называется *оценкой по методу моментов* (ОММ).

**Теорема** (одномерный случай). *Если функция* $m_g(\theta)$ *является непрерывной и строго монотонной для некоторой пробной функции* $g$, *то оценка по методу моментов является сильно состоятельной.*

Доказательство. У монотонной непрерывной функции $m_g(\theta)$ существует обратная функция $m_g^{-1}(\theta)$, которая также является непрерывной и монотонной. Тогда $\theta_n^* = m_g^{-1}(\overline{g(x)})$.

В силу УЗБЧ $\overline{g(x)} \xrightarrow[\text{п.н.}]{} \mathbf{E}g(x_1)$. Так как $m_g^{-1}$ — непрерывная функция, то

$$m_g^{-1}(\overline{g(x)}) \xrightarrow[\text{п.н.}]{} m_g^{-1}\left(\mathbf{E}g(x_1)\right) = m_g^{-1}(m_g(\theta_0)) = \theta_0.$$

Таким образом, $\theta_n^* \xrightarrow[\text{п.н.}]{} \theta_0$. □

Перейдём к $k$-мерному варианту метода моментов, когда параметр $\theta \in \mathbb{R}^k$. Вводим $k$ пробных функций $g_1, \ldots, g_k$ и по каждой из них строим соответствующие функции $m_{g_i}(\theta_1, \ldots, \theta_k)$. Получаем систему уравнений

$$m_{g_i}(\theta_1, \ldots, \theta_k) = \overline{g_i(x)}, \;\; i = 1, \ldots, k.$$

Если эта система разрешима, то векторная оценка $\theta_n^* = (\theta_n^{*(1)}, \ldots, \theta_n^{*(k)})$ называется ОММ. Заметим, что эта оценка может быть неединственной.

**Примеры.**

**1.** *Одномерный случай.* Рассматривается распределение Пуассона $\{\pi_\lambda\}, \lambda > 0$. В качестве пробной функции возьмём $g(x) = x$. Тогда $m(\lambda) = \lambda$; эмпирическим аналогом функции $g$ будет $\overline{g(x)} = \overline{x}$. Получаем уравнение $\lambda = \overline{x}$, решением которого является $\lambda_n^* = \overline{x} \underset{\text{п.н.}}{\longrightarrow} \lambda_0$.

Если в качестве пробной функции взять $g(x) = x^2$, то $m_2(\lambda) = \lambda + \lambda^2$, и соответствующее уравнение примет вид $\lambda^2 + \lambda - \overline{x^2} = 0$. Решая его и учитывая, что параметр $\lambda$ положителен, получим

$$\lambda_n^* = \frac{\sqrt{1 + 4\overline{x^2}} - 1}{2} \underset{\text{п.н.}}{\longrightarrow} \frac{\sqrt{1 + 4(\lambda^2 + \lambda)} - 1}{2} = \frac{2\lambda + 1 - 1}{2} = \lambda.$$

**Упражнение.** *Найти ОММ, взяв в качестве пробной функции* $g(x) = x^3$.

Как мы видим, один и тот же достаточно простой метод позволяет получить множество оценок одного и того же параметра, причём сильно состоятельных.

**2.** *Многомерный случай.* Рассмотрим нормальное распределение $\{N(\alpha, \sigma)\}$. Возьмём пробные функции $g_1(x) = x, \;\; g_2(x) = x^2$. Тогда $m_1(\alpha, \sigma) = \alpha, \;\; m_2(\alpha, \sigma) = \alpha^2 + \sigma^2$. Получаем систему двух (нелинейных) уравнений

$$\begin{cases} \alpha = \overline{x}, \\ \alpha^2 + \sigma^2 = \overline{x^2}, \end{cases}$$

решением которой является двумерная статистика $\theta_n^* = (\overline{x}, S)$, где $S^2 = \overline{x^2} - (\overline{x})^2$. В силу УЗБЧ имеем $\overline{x} \underset{\text{п.н.}}{\longrightarrow} \alpha_0, S^2 \underset{\text{п.н.}}{\longrightarrow} \mathbf{D}x_1 = \sigma_0^2$. Стало быть, полученная оценка является сильно состоятельной.

**Упражнение.** *Для атомарного распределения*

$$x_1 = \begin{cases} 1, & p_1, \\ \vdots & \\ m, & p_m = 1 - \sum\limits_{i=1}^{m-1} p_i, \end{cases}$$

*где* $p_1, \ldots, p_{m-1}$ *— неизвестные параметры, построить оценку по методу моментов и исследовать ее состоятельность.*

## Метод максимального правдоподобия.

Рассмотрим ещё один способ построения оценок. Как и в методе моментов, будем работать с параметрическим классом распределений, задаваемом плотностями $\{f_\theta\}_{\theta \in \Theta}$ относительной фиксированной $\sigma$-конечной меры. Ограничимся одномерными наблюдениями.

Проиллюстрируем идею метода максимального правдоподобия на следующем примере. Пусть $f_0(x)$ — плотность, график которой имеет одну вершину. Введём семейство плотностей $f_\theta(x) = f_0(x - \theta)$, где $\theta$ — параметр сдвига. Пусть $x_1$ — некоторая точка на числовой прямой.

Принимая во внимание вероятностный смысл плотности распределения, наиболее правдоподобной плотностью из данного семейства следует считать ту, которая достигает максимума в точке $x_1$, так как плотность в той или иной точке пропорциональна вероятности попадания в окрестность этой самой точки.

Формализуем описанную процедуру, переходя на язык параметра (ибо параметр и плотность находятся во взаимно однозначном соответствии). Если $\theta_0$ — истинное значение параметра, то его оценка

$$\hat{\theta}_n = \arg\max_{\theta} f_\theta(x_1),$$

где $n = 1$ — объём наблюдений.

Пусть теперь $X = (x_1, \ldots, x_n)$ — выборка объёма $n$, которая является $n$-мерным случайным вектором с независимыми одинаково распределенными координатами. Плотность этого вектора есть произведение маргинальных плотностей:

$$f(z_1, \ldots, z_n) = \prod_{i=1}^{n} f_\theta(z_i).$$

Реализуем высказанную выше идею. Введем в рассмотрение так называемую *функцию правдоподобия*

$$\Psi_X(\theta) = \prod_{i=1}^{n} f_\theta(x_i),$$

которая как раз и представляет собой суперпозицию плотности совместного распределения координат выборочного вектора и самой выборки. Если существует экстремальная точка $\hat{\theta}_n = \arg\max\limits_{\theta} f_X^{(\theta)}(X)$, то она называется *оценкой максимального правдоподобия* (ОМП) неизвестного параметра. Заметим, что в определении ОМП мы ничего не говорили об одновершинности графика плотности. Поэтому возникают вопросы о существовании и единственности ОМП. Отметим, что если экстремальных точек $\hat{\theta}_n = \arg\max\limits_{\theta} \Psi_X(\theta)$ несколько, то все они называются ОМП.

**Пример неединственности ОМП.** Построим ОМП для параметрического семейства равномерных распределений $\{U_{[\theta,\theta+1]}\}, \theta \in \mathbb{R}$. Здесь

$$f_\theta(t) = \begin{cases} 1, & t \in [\theta, \theta+1], \\ 0, & \text{иначе.} \end{cases}$$

Отсюда легко получаем выражение для функции правдоподобия:

$$\Psi_X(\theta) = \begin{cases} 1, & x_{(n)} - 1 \leqslant \theta \leqslant x_{(1)}, \\ 0, & \text{иначе,} \end{cases}$$

откуда следует, что в данном примере множество ОМП представляет собой целый отрезок $\{\theta : \; x_{(n)} - 1 \leqslant \theta \leqslant x_{(1)}\}$.

**Упражнение.** *Построить пример параметрического семейства распределений, для которого не существует ОМП.*

*Указание: Рассмотреть параметрическое семейство, представитель которого является смесью (с фиксированными весами) произвольного нормального и стандартного нормального распределений.*

**Примеры вычисления ОМП.**

**1.** *Пуассоновское параметрическое семейство* $\{\pi_\lambda\}, \lambda > 0$. Любое решетчатое распределение имеет плотность относительно считающей меры. В данном случае

$$f_\lambda(x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x \in \mathbb{Z}_+.$$

Выпишем функцию правдоподобия

$$\Psi_X(\lambda) = \frac{\lambda^{n\overline{x}}}{\prod\limits_{i=1}^{n} x_i!}e^{-\lambda n}.$$

Исследуем эту функцию на экстремум. Для этого найдем решение уравнения $\Psi'_X(\lambda) = 0$. Получаем решение $\hat{\lambda}_n = \overline{x}$. Учитывая вид функции, нам становится ясно, что это точка максимума. Отметим, что метод моментов давал тот же результат (более того, лучшей оценки для параметра $\lambda$ в данном случае не существует). В силу УЗБЧ полученная оценка является сильно состоятельной.

**2.** *Нормальное распределение с двумя неизвестными параметрами* $\{N_{\alpha,\sigma}\}$. Нам нужно исследовать функцию правдоподобия на экстремум. Так как монотонное преобразование не меняет экстремальных точек, то для удобства можно перейти к *логарифмической функции правдоподобия (ЛФП)* $L_X(\theta) = \log \Psi_X(\theta)$.

В данном случае представителем параметрического семейства плотностей является

$$f_{(\alpha,\sigma)} = \frac{e^{-(t-\alpha)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}.$$

Переходя к ЛФП, получим

$$L_X(\alpha,\sigma) = \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n - \sum_{i=1}^{n}\frac{(x_i-\alpha)^2}{2\sigma^2} = -\frac{n}{2}\log(\sigma^2\cdot 2\pi) - \frac{n\overline{x^2}}{2\sigma^2} + \frac{\alpha n\overline{x}}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2}.$$

Для поиска стационарной точки решаем систему из двух уравнений. Первое уравнение:

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\sigma^2}(n\overline{x} - n\alpha) = 0,$$

из него мы получаем ОМП для первого параметра $\hat{\alpha}_n = \overline{x}$.

Второе уравнение после подстановки вместо $\alpha$ выборочного среднего принимает вид:

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{n\overline{x^2} - 2\alpha n\overline{x} + n\alpha^2}{2\sigma^4} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}[n\overline{x^2} - n(\overline{x})^2] = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}nS^2 = 0.$$

Из этого уравнения мы получаем ОМП для второго параметра $\hat{\sigma}_n^2 = S^2$. Наши результаты совпадают с ОММ, однако процесс вычислений оказался более трудоемким.

**3.** *Равномерное распределение на* $[0,\theta]$. Оценим $\theta$ двумя способами. Сначала воспользуемся методом моментов. В качестве функции $m(\theta)$ рассматриваем первый момент наблюдаемого распределения:

$$m(\theta) = \int\limits_0^\theta t f_\theta(t)\,dt,$$

где
$$f_\theta(t) = \begin{cases} 1/\theta, & t \in [0,\theta], \\ 0, & \text{иначе}. \end{cases}$$
Тогда $m(\theta) = \theta/2$ и по методу моментов получаем уравнение $\theta/2 = \overline{x}$. Откуда $\theta_n^* = 2\overline{x}$. Если бы мы в качестве пробных функций брали $g(x) = x^k, k \in \mathbb{N}$, то получили бы оценки $\theta_{n,k}^* = ((k+1)\overline{x^k})^{1/k}$. Все эти оценки являются сильно состоятельными.

Теперь воспользуемся методом максимального правдоподобия. Имеем
$$\Psi_X(\theta) = \begin{cases} 1/\theta^n, & \theta \geqslant x_{(n)} = \max\limits_{i \leqslant n} x_i, \\ 0, & \text{иначе}. \end{cases}$$
Получаем ОМП $\hat{\theta}_n = x_{(n)}$.

Исследуем состоятельность оценки $\hat{\theta}_n = x_{(n)}$ для равномерного распределения на $[0,\theta]$. Для этого рассмотрим
$$\mathbf{P}(|x_{(n)} - \theta| > \varepsilon) = \mathbf{P}(x_{(n)} < \theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \xrightarrow[n \to \infty]{} 0.$$
Так как последовательность $\{x_{(n)}\}$ монотонна, то сходимость по вероятности совпадает со сходимостью почти наверное, т. е. в последнем примере (впрочем, как и в предыдущих) ОМП сильно состоятельная.

Заметим, что полученная в последнем примере ОМП не совпадает ни с какой из полученных ОММ. Но всё-таки эти оценки связаны.

**Упражнение.** *Доказать, что для всех выборок при фиксированном $n$ имеет место предельное соотношение*
$$\lim_{k \to \infty} \theta_{n,k}^* = x_{(n)}.$$

На качественном уровне можно отметить, что, как правило, ОМП предпочтительнее ОММ, так как являются более точными. Далее мы еще вернемся к этому вопросу.

## Состоятельность ОМП.

Пусть $\{f(\theta)\}_{\theta \in \Theta}$ — параметрическое семейство плотностей относительно $\sigma$-конечной меры $\lambda(\cdot)$, а $l_{x_i}(\theta) = \log f_\theta(x_i)$. Обозначим через $\theta_0$ истинное значение параметра наблюдаемого распределения. Введем в рассмотрение функцию $\varphi(\theta) = \mathbf{E}_{\theta_0} l_{x_1}(\theta)$, где $\theta \in \Theta$. При условии существования функции $\varphi(\theta)$ для каждого $\theta \in \Theta$ исследуем её свойства. Сформулируем условия, в рамках которых будут доказаны последующие утверждения.

**Условия** $(A)$:

1. $(\Theta, d)$ — метрический компакт с метрикой $d$.

2. Параметрическое семейство плотностей находится во взаимно-однозначном соответствии с параметрическим множеством $\Theta$, т.е.
$$f_{\theta_1}(x) = f_{\theta_2}(x) \text{ почти всюду относительно } \sigma\text{−конечной меры } \lambda \Leftrightarrow \theta_1 = \theta_2.$$

3. Носители всех распределений из параметрического семейства не зависят от $\theta$ (т. е. $\operatorname{supp} f_\theta$ один и тот же для всех распределений).

**Лемма**. *Пусть выполнены условия $(A)$. Тогда $\varphi(\theta) < \varphi(\theta_0)$ для всех $\theta \neq \theta_0$.*

ДОКАЗАТЕЛЬСТВО. Обозначим $\mathfrak{X} = \mathrm{supp} f_{\theta_0}$. Оценим сверху разность

$$\varphi(\theta) - \varphi(\theta_0) = \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z) \log f_\theta(\bar z)\, \lambda(d\bar z) - \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z) \log f_{\theta_0}(\bar z)\, \lambda(d\bar z) =$$
$$= \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z) \log \frac{f_\theta(\bar z)}{f_{\theta_0}(\bar z)}\, \lambda(d\bar z) = \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z) \log\Big(\frac{f_\theta(\bar z)}{f_{\theta_0}(\bar z)} \pm 1\Big)\, \lambda(d\bar z) =$$
$$= \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z) \log\Big(1 + \Big(\frac{f_\theta(\bar z)}{f_{\theta_0}(\bar z)} - 1\Big)\Big)\, \lambda(d\bar z).$$

Так как $\log(1+x) \leqslant x$ для всех $x \geqslant -1$ и знак равенства возможен только при $x = 0$, то

$$\varphi(\theta) - \varphi(\theta_0) \leqslant \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z)\Big(-1 + \frac{f_\theta(\bar z)}{f_{\theta_0}(\bar z)}\Big)\, \lambda(d\bar z) = \int\limits_{\mathfrak{X}} f_\theta(\bar z)\, \lambda(d\bar z) - \int\limits_{\mathfrak{X}} f_{\theta_0}(\bar z)\, \lambda(d\bar z) = 1 - 1 = 0.$$

В силу второго пункта условий $(A)$ и условий теоремы, неравенство будет строгим. □

**Замечание.** Если множество $\Theta$ является метрическим компактом, то единственный абсолютный экстремум непрерывной функции $\varphi$ будет *отделимым* (это легко доказывается от противного), т. е. при любом $\delta > 0$

$$\sup_{\theta:\, d(\theta,\theta_0)>\delta} \varphi(\theta) < \varphi(\theta_0).$$

Это обстоятельство является очень важным для доказательства состоятельности ОМП.

**Определение.** Величину $L_X^{(n)}(\theta) = \frac{1}{n} L_X(\theta)$ будем называть *нормированной логарифмической функцией правдоподобия*. Так как умножение на константу есть монотонное преобразование, то экстремальные точки этой функции совпадают с экстремальными точками логарифмической функции правдоподобия, а значит, и самой функции правдоподобия.

**Лемма**. *Пусть выполнены условия $(A)$, и $l_x(\theta) \in Lip(K(x))$, где $\mathbf{E}_\theta K(x_1) < \infty$. Тогда*

$$\sup_{\theta\in\Theta} |L_X^{(n)}(\theta) - \varphi(\theta)| \xrightarrow[\text{п.н.}]{} 0 \ \text{ при } n \to \infty.$$

**Замечание.** Процесс доказательства полностью аналогичен доказательству обобщенной теоремы Гливенко — Кантелли.

ДОКАЗАТЕЛЬСТВО. Отметим, что для каждого $\theta$ выполнено $L_X^{(n)}(\theta) \xrightarrow[\text{п.н.}]{} \varphi(\theta)$ в силу УЗБЧ. Пусть $\varepsilon > 0$. Так как параметрическое множество $\Theta$ является метрическим компактом, то для каждого $\varepsilon > 0$ в $\Theta$ существует конечная $\varepsilon$-сеть. Пусть $\{\, \theta_i \mid i \leqslant N(\varepsilon) \,\}$ — совокупность всех узлов $\varepsilon$-сети, $S_i(\varepsilon) = \{\, \theta \in \Theta \mid d(\theta, \theta_i) \leqslant \varepsilon \,\}$ — замкнутый шар радиуса $\varepsilon$ с центром $\theta_i$. Совокупность всех $S_i(\varepsilon)$ образует конечное покрытие $\Theta$. Поэтому

$$\sup_{\theta\in\Theta} |L_X^{(n)}(\theta) - \varphi(\theta)| = \max_{i\leqslant N(\varepsilon)} \sup_{\theta\in S_i(\varepsilon)} |L_X^{(n)}(\theta) - \varphi(\theta)|.$$

Заметим, что

$$\sup_{\theta\in S_i(\varepsilon)} |L_X^{(n)}(\theta) - \varphi(\theta)| = \sup_{\theta\in S_i(\varepsilon)} |L_X^{(n)}(\theta) \pm L_X^{(n)}(\theta_i) \pm \varphi(\theta_i) - \varphi(\theta)| \leqslant$$
$$\leqslant \sup_{\theta\in S_i(\varepsilon)} |L_X^{(n)}(\theta) - L_X^{(n)}(\theta_i)| + \sup_{\theta\in S_i(\varepsilon)} |\varphi(\theta_i) - \varphi(\theta)| + |L_X^{(n)}(\theta_i) - \varphi(\theta_i)|.$$

Вспомним, что $l_x(\theta) \in Lip(K(x))$, поэтому $|l_x(\theta_1) - l_x(\theta_2)| \leqslant K(x)d(\theta_1, \theta_2)$. Оценим первое слагаемое:

$$\sup_{\theta \in S_i(\varepsilon)} |L_X^{(n)}(\theta) - L_X^{(n)}(\theta_i)| = \sup_{\theta \in S_i(\varepsilon)} |\frac{1}{n}\sum_{j=1}^{n}(l_{x_j}(\theta) - l_{x_j}(\theta_i)| \leqslant \sup_{\theta \in S_i(\varepsilon)} |d(\theta, \theta_i)\frac{1}{n}\sum_{j=1}^{n} K(x_j)|.$$

В последнем выражении под знаком модуля стоит величина $\frac{1}{n}\sum\limits_{j=1}^{n} K(x_j)$ — среднее арифметическое независимых одинаково распределенных случайных величин с конечным первым моментом. Тогда

$$\limsup_{n\to\infty} \sup_{\theta \in S_i(\varepsilon)} |L_X^{(n)}(\theta) - L_X^{(n)}(\theta_i)| \le \limsup_{n\to\infty} |\frac{1}{n}\varepsilon\sum_{j=1}^{n} K(x_i)| = \varepsilon\mathbf{E}_{\theta_0}K(x_1).$$

Оценка второго слагаемого:

$$\sup_{\theta \in S_i(\varepsilon)} |\varphi(\theta_i) - \varphi(\theta)| = \sup_{\theta \in S_i(\varepsilon)} |\mathbf{E}_{\theta_0}(l_{x_1}(\theta) - l_{x_1}(\theta_i))| \leqslant \sup_{\theta \in S_i(\varepsilon)} \mathbf{E}_{\theta_0}|l_{x_1}(\theta) - l_{x_1}(\theta_i)| \leqslant$$
$$\leqslant \mathbf{E}_{\theta_0}K(x_1) \sup_{\theta \in S_i(\varepsilon)} d(\theta, \theta_i) \leqslant \varepsilon\mathbf{E}_{\theta_0}K(x_1).$$

Первое неравенство в этой цепочке — следствие применения неравенства треугольника, второе — монотонности математического ожидания. Таким образом, $\varphi \in \mathrm{Lip}(\mathbf{E}_{\theta_0}K(x_1))$ и

$$\limsup_{n\to\infty} \sup_{\theta \in S_i(\varepsilon)} |\varphi(\theta_i) - \varphi(\theta)| \leqslant \varepsilon\mathbf{E}_{\theta_0}K(x_1).$$

В силу УЗБЧ имеем $\limsup\limits_{n\to\infty} |L_X^{(n)}(\theta_i) - \varphi(\theta_i)| = 0$. В итоге получаем

$$\limsup_{n\to\infty} \sup_{\theta \in S_i(\varepsilon)} |L_X^{(n)}(\theta) - \varphi(\theta)| \leqslant 2\varepsilon\mathbf{E}_{\theta_0}K(x_1).$$

Так как левая часть неравенства не зависит от $\varepsilon$, лемма доказана. □

**Теорема** (Сильная состоятельность ОМП). *Пусть выполнены условия* $(A)$ *и* $l_x(\theta) \in \mathrm{Lip}(K(x))$*, где* $\mathbf{E}_\theta K(x_1) < \infty$*. Тогда ОМП сильно состоятельна, т. е.* $\widehat{\theta}_n \xrightarrow[\text{п.н.}]{} \theta_0$ *при* $n \to \infty$.

**Упражнение.** *Доказать теорему самостоятельно методом от противного, используя две предыдущие леммы.*

**Пример.** Рассмотрим параметрическое семейство пуассоновских распределений с параметром $\lambda$, отделенным от нуля, т. е. $\{\pi_\lambda\}_{\lambda \geq \delta > 0}$. Тогда $l_x(\lambda) = x\log\lambda - \lambda - \log x!$. Заметим, что функция $l_x(t)$ липшицева при $t \in \Theta$, если $\sup\limits_{t\in\Theta}|l'_x(t)| < \infty$. В нашем случае $|\frac{x}{\lambda} - 1| = |\frac{x}{\delta} - 1| = K(x)$. Таким образом, $\mathbf{E}_\lambda K(x)$ существует, и значит, выполняются условия теоремы.

Пункт 3 условий $(A)$ является техническим ограничением и может быть существенно ослаблен. Отметим, что параметрическое семейство $\{U[0,\theta]\}_{\theta>0}$ не удовлетворяет этому условию.

## Сравнение оценок.

Для одного и того же параметра можно построить бесконечное число оценок. Например, для семейства $\{U_{[0,\theta]}\}_{\theta>0}$ ,была построена последовательность оценок по методу моментов:

$\theta^*_{nk} = ((k+1)\overline{x^k})^{1/k},\ \ k \in \mathbb{N}$. Кроме того, методом максимального правдоподобия мы построили оценку $\widehat{\theta}_n = X_{(n)}$. Причем, все эти оценки сильно состоятельные. Какую же оценку из этого счетного набора предпочесть?

Введем в рассмотрение так называемую *функцию потерь*:

$$\delta_{\theta^*_n}(\theta) = \mathbf{E}_\theta(\theta^*_n - \theta)^2,$$

где $\mathbf{E}_\theta$ обозначает математическое ожидание, вычисленное при условии, что $\theta$ — истинное значение параметра наблюдаемого распределения. Если мы находимся в рамках параметрического семейства плотностей $\{f_\theta\}_{\theta\in\Theta}$, то можно записать функцию потерь более подробно:

$$\delta_{\theta^*_n}(\theta) = \int\limits_{\mathfrak{X}^n}\!\cdots\!\int (\theta^*_n(\bar z) - \theta)^2 \Psi_{\bar z}(\theta)\lambda^n(d\bar z).$$

Величина функции потерь и будет критерием качества (или точности) выбираемых оценок.

**Определение.** *Оценка $\theta^*_{n,1}$ «лучше» оценки $\theta^*_{n,2}$ в точке $\theta_1$*, если $\delta_{\theta^*_{n,1}}(\theta_1) < \delta_{\theta^*_{n,2}}(\theta_1)$.

**Определение.** *Оценка $\theta^*_{n,1}$ «лучше» оценки $\theta^*_{n,2}$*, если при всех $\theta$

$$\delta_{\theta^*_{n,1}}(\cdot) \le \delta_{\theta^*_{n,2}}(\cdot)$$

и хотя бы для одного $\theta$ неравенство будет строгим.

Заметим, что второе определение употребляется чаще, но не всегда применимо, поскольку две функции потерь могут быть несравнимы. Наша цель — указать по возможности более широкий класс оценок, в котором возможно отыскать оценку с минимальной функцией потерь. Другими словами, график этой функции должен быть нижней огибающей для функций потерь всех оценок из рассматриваемого класса.

**Определение.** Оценка называется *несмещенной*, если $\mathbf{E}_\theta\theta^*_n = \theta, \forall\theta \in \Theta$. Если $\theta^*_n = \theta + b_n(\theta)$, то оценка называется *смещенной*, а $b_n$ — ее *смещением*.

Для несмещенной оценки $\delta_{\theta^*_n}(\theta) = \mathbf{D}\theta^*_n$. Заметим, что понятие несмещенности и состоятельности не связаны. Например, для пуассоновского семейства распределений $\{\pi_\lambda\}$ примером несмещенной оценки является $\theta^*_n = x_1$, при этом, очевидно, здесь нет никакой состоятельности.

**Пример.** Сравним оценки $\theta^*_n = 2\overline{X}$ и $\theta^*_n = x_{(n)}$ для параметра $\theta$ равномерного распределения на отрезке $[0,\theta],\ \theta > 0$. Для первой оценки

$$\mathbf{E}_\theta\theta^*_n = 2\mathbf{E}_\theta\overline{X} = \frac{2}{n}n\mathbf{E}x_1 = 2\frac{\theta}{2} = \theta.$$

Следовательно, оценка несмещенная. Значит, функция потерь для этой оценки совпадает с дисперсией.

**Упражнение.** *Показать, что оценка $\widehat{\theta}_n = x_{(n)}$ смещенная и вычислить ее смещение.*

Вернемся к нашему примеру и найдем функции потерь для обеих оценок:

$$\delta_{\theta^*_n}(\theta) = \mathbf{D}(2\overline{X}) = \frac{4}{n^2}n\mathbf{D}x_1 = \frac{4}{n}\mathbf{D}x_1 = \frac{4}{n}\left(\mathbf{E}x_1^2 - \left(\frac{\theta}{2}\right)^2\right) = \frac{4}{n}\left(\frac{\theta^2}{3} - \frac{\theta^2}{4}\right) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Так как $F_{x_{(n)}}(t) = \left(\frac{t}{\theta}\right)^n$, то $p_{x_{(n)}}(t) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1}$, и мы получаем

$$\delta_{x_{(n)}}(\theta) = \int\limits_0^\theta (t-\theta)^2 p_{x_{(n)}}(t)\,dt = \int\limits_0^\theta (t^2 - 2\theta t + \theta^2)\frac{nt^{n-1}}{\theta^n}\,dt,$$

откуда следует

$$\delta_{x_{(n)}}(\theta) = \frac{n}{\theta^n}\left(\frac{\theta^{n+2}}{n+2} - \frac{2\theta^{n+2}}{n+1} + \frac{\theta^{n+2}}{n}\right) = \frac{2\theta^2}{(n+1)(n+2)}.$$

Сравнивая функции потерь, мы делаем вывод, что ОМП на “порядок” точнее (по $n$) для данного параметрического класса распределений. Заметим, что при $k \neq 1$ явно найти функцию потерь для оценки по методу моментов весьма затруднительно, так как интеграл

$$\int\limits_{[0,\theta]^n}\!\!\cdots\!\int \left(\sqrt[k]{(k+1)\frac{1}{n}\sum z_1^k} - \theta\right)^2 \frac{1}{\theta^n}\,d\bar{z}$$

не вычисляется в явном виде.

## Асимптотически нормальные оценки.

**Определение.** Оценка $\theta_n^*$ называется *асимптотически нормальной* (АНО), если при $n \to \infty$

$$\sqrt{n}(\theta_n^* - \theta) \Rightarrow \eta \in N(0,\sigma)$$

для некоторого $\sigma$. Эквивалентным определением будет

$$\frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma} \Rightarrow \eta_0 \in N(0,1),$$

где $\sigma(\theta)$ называется *коэффициентом рассеивания* или *коэффициентом асимптотической нормальности.*

**Упражнение.** *Показать, что любая АНО будет состоятельной.*

Напомним, что нами была доказана теорема о том, что слабая сходимость $\xi_n \Rightarrow \xi$ при условии, что $\sup_n \mathbf{E}|\xi_n| I(|\xi_n| > N) \to 0$ при $N \to \infty$ (условие равномерной интегрируемости), влечет за собой сходимость моментов

$$\mathbf{E}\xi_n \to \mathbf{E}\xi.$$

Следовательно, если последовательность $\xi_n = n(\theta_n^* - \theta)^2$ удовлетворяет условию равномерной интегрируемости и $\theta_n^*$ — АНО, то мы получим $\xi_n \Rightarrow \eta^2$, где $\eta \in N(0,\sigma)$. Тогда при выполнении указанных условий

$$n\delta_{\theta_n^*}(\theta) = \mathbf{E}_{\theta^n}(\theta_n^* - \theta)^2 \to \mathbf{E}\eta^2 = \sigma^2.$$

Это означает, что при $n \to \infty$ имеет место следующая эквивалентность:

$$\delta_{\theta_n^*}(\theta) \sim \frac{\sigma^2(\theta)}{n}.$$

Поэтому естественным представляется следующее правило сравнения оценок:

**Определение.** Пусть $\theta^*_{n,1}$ и $\theta^*_{n,2}$ — две АНО. Тогда $\theta^*_{n,1}$ *«лучше»* $\theta^*_{n,2}$, если для любого $\theta$

$$\sigma_1(\theta) \le \sigma_2(\theta).$$

Другими словами, мы сформулировали так называемый *асимптотический подход* к сравнению оценок.

**Пример.** При условии $\mathbf{E}g^2(x_1) < \infty$ статистика $\overline{g(x)}$ является асимптотически нормальной оценкой для соответствующего момента: $\theta^*_n = \overline{g(x)} \to \theta = \mathbf{E}g(x_1)$. В самом деле, в силу центральной предельной теоремы имеем

$$\sqrt{n}(\theta^*_n - \theta) = \frac{\sum\limits_{i=1}^{n} g(x_i) - \mathbf{E}g(x_1)}{\sqrt{n}} \Rightarrow \eta \in N(0, \sigma),$$

где $\sigma^2 = \mathbf{D}g(x_1)$ — коэффициент рассеивания.

**Теорема (о суперпозиции).** *Пусть $\theta^*_n$ произвольная АНО, коэффициент рассеивания которой $\sigma(\theta) > 0$ непрерывен. Тогда суперпозиция $\widetilde{\theta}^*_n = H(\theta^*_n)$, где $H \in \mathbb{C}^1(\cup_n \mathrm{supp} F_{\theta^*_n})$ и $H'(\theta) \neq 0$ при всех $\theta$, будет АНО для параметра $\widetilde{\theta} = H(\theta)$, причем новый коэффициент рассеивания вычисляется по формуле $\widetilde{\sigma} = \sigma|H'(\theta)|$.*

Доказательство. По формуле конечных приращений получаем

$$\sqrt{n}(\widetilde{\theta}^*_n - \widetilde{\theta}) = \sqrt{n}(H(\theta^*_n) - H(\theta)) = H'(\widehat{\theta}^*_n)\sqrt{n}(\theta^*_n - \theta),$$

где $\min(\theta^*_n, \theta)) \leqslant \widehat{\theta}^*_n \leqslant \max(\theta^*_n, \theta))$.

Ранее мы отметили, что всякая АНО является состоятельной, т. е. $\theta^*_n \to \theta$. В силу гладкости функции $H$ по “принципу двух милиционеров” получаем, что $H'(\theta^*_n) \to H'(\theta)$.

Воспользуемся следующим утверждением из курса теории вероятностей: если $\xi_n \Rightarrow \xi$ и $\eta_n \underset{p}{\to} c = \mathrm{const}$, то $\xi_n\eta_n \Rightarrow c\xi$. В нашем случае $\eta_n = H'(\theta^*_n)$, $c = H'(\theta)$. Тогда из приведенного утверждения следует, что

$$\sqrt{n}(\widetilde{\theta}^*_n - \widetilde{\theta}) \Rightarrow H'(\theta)\eta,$$

где $\eta \in N_{(0,\sigma)}$. Тогда по свойству нормального распределения $\widetilde{\sigma}^2 = \mathbf{D}H'(\theta)\eta = (H'(\theta))^2\sigma^2$. Следовательно, $\widetilde{\sigma} = \sigma|H'(\theta)|$. □

**Следствие.** *Все оценки $\theta^*_{n,k} = ((k+1)\overline{x^k})^{1/k}, k \in \mathbb{N}$, являются АНО для параметра $\theta$ равномерного распределения на отрезке $[0, \theta]$.*

В этом случае $H(t) = ((k+1)t)^{1/k}, t > 0$ и $\theta^*_n = \overline{x^k}$ — это АНО.

Вычислим коэффициент рассеивания $\sigma$ внутренней аддитивной оценки $\overline{x^k}$:

$$\sigma^2 = \mathbf{D}x_1^k = \mathbf{E}x_1^{2k} - (\mathbf{E}x_1^k)^2 = \frac{\theta^{2k}}{2k+1} - \left(\frac{\theta^k}{k+1}\right)^2 = \frac{\theta^{2k}\cdot k^2}{(2k+1)(k+1)^2},$$

Кроме того, $H'(t) = (k+1)^{1/k}t^{1/k-1}k^{-1}$.

Поскольку $\theta^*_n \xrightarrow[\text{п.н.}]{} \frac{\theta^k}{k+1}$, то

$$H'\left(\frac{\theta^k}{k+1}\right) = \frac{(k+1)^{1/k}}{k}\left(\frac{\theta^k}{k+1}\right)^{1/k-1} \cdot \frac{\theta^k \cdot k}{(k+1)\sqrt{2k+1}} = \frac{\theta}{\sqrt{2k+1}} = \tilde{\sigma}.$$

Вывод: чем больше $k$, тем точнее получается оценка. Напомним, что в пределе мы имеем $\lim\limits_{k\to\infty}((k+1)\overline{x^k})^{1/k} = x_{(n)}$. □

Ранее было показано, что при некоторых минимальных ограничениях справедливо следующее асимптотическое представление для функции потерь АНО:

$$\delta_{\theta_n^*}(\theta) \sim \frac{\sigma_k^2}{n}.$$

Однако функция потерь для $x_{(n)}$ имеет иной порядок по $n$:

$$\delta_{x_{(n)}}(\theta) = \frac{2\theta^2}{(n+1)(n+2)}.$$

Разумеется, это не является доказательством того, что $x_{(n)}$ не АНО. Докажем это факт напрямую. Заметим, что для любого $n$ с вероятностью 1 выполняется неравенство $\sqrt{n}(x_{(n)} - \theta) < 0$. Но если бы $x_{(n)}$ была АНО, то предельная случайная величина $\eta$ имела бы распределение $N(0, \sigma)$, которое симметрично. Иными словами, тогда бы имело место соотношение

$$\mathbf{P}(\sqrt{n}(x_{(n)} - \theta) < 0) \to \mathbf{P}(\eta < 0) = 1/2,$$

чего не может быть.

## Достаточные статистики.

**Определение.** Статистика $S = S(X)$ называется *достаточной для параметра $\theta$*, если условное распределение выборки $\mathbf{P}_\theta(X \in A|S) \equiv g(A, S)$ при фиксации $S$ как сопутствующего наблюдения (возможно, многомерного) *не зависит от параметра $\theta$*.

Достаточность означает, что в статистике $S$ содержится вся информация о выборке, достаточная для построения оптимальных оценок. Дальнейшая конкретизация положения выборочных значений бессмысленна. Если выборка из дискретного распределения, то в вышеприведенном определении вместо ортопроекции можно использовать классическое условное распределение $\mathbf{P}_\theta(X \in A|S = s_i)$, которое можно интерпретировать как распределение на поверхности $S = s_i$ в выборочном пространстве. Если $S$ — достаточная статистика, то это распределение не зависит от параметра $\theta$ при всевозможных значениях $s_i$.

**Замечание.** Статистика $S = X$ является достаточной всегда.

**Пример.** Покажем, что статистика $S = n\overline{x}$ является достаточной для параметра $\lambda$ класса пуассоновских распределений $\{\pi_\lambda\}_{\lambda>0}$. Областью значений исследуемой статистики является множество $\mathbb{Z}_+$.

Пусть $k \in \mathbb{Z}_+$. Тогда

$$\mathbf{P}_\lambda(X = \vec{z}|S(X) = k) = \frac{\mathbf{P}_\lambda(X = \vec{z}, S(X) = k)}{\mathbf{P}_\lambda(S(X) = k)} = \begin{cases} \dfrac{\mathbf{P}_\lambda(X = \vec{z})}{\mathbf{P}_\lambda(S(X) = k)}, & k = S(\vec{z}), \\ 0, & k \neq S(\vec{z}). \end{cases}$$

Распределение Пуассона устойчиво, поэтому $S = \sum\limits_{i=1}^n x_i \in \Pi_{n\lambda}$. Таким образом, в случае $k = S(\vec{z}) = \sum\limits_{i=1}^n z_i$ выполняется соотношение:

$$\mathbf{P}_\lambda(S(X) = k) = \frac{(n\lambda)^k}{k!} e^{-n\lambda}.$$

Следовательно,

$$\mathbf{P}_\lambda(X=\vec{z}|S(X)=k)=\frac{\prod\limits_{i=1}^{n}\mathbf{P}_\lambda(x_i=z_i)}{\mathbf{P}_\lambda(S(X)=k)}=\frac{e^{-\lambda n}k!\lambda^k}{e^{-\lambda n}(n\lambda)^k\prod\limits_{i=1}^{n}z_i!}=\frac{k!}{n^k\prod\limits_{i=1}^{n}z_i!}.$$

Последнее выражение не зависит от $\lambda$ при всех $k$ и $\vec{z}$. Значит, статистика $S$ достаточная.

**Упражнение.** *Дать вероятностную интерпретацию следующего выражения*:

$$\frac{k!}{n^k\prod\limits_{i=1}^{n}z_i!},$$

*где* $k=\sum\limits_{i=1}^{n}z_i$.

**Теорема** (Факторизационная теорема Неймана–Фишера). *Статистика $S$ является достаточной для параметра $\theta$ тогда и только тогда, когда функция правдоподобия $\Psi_X(\theta)$ допускает факторизацию*

$$\Psi_X(\theta)=\varphi(S,\theta)h(X)\quad \textit{п.н.},$$

*где каждая из функций $\varphi$ и $h$ зависит только от указанных аргументов.*

Доказательство. Проведем доказательство теоремы только для параметрического класса дискретных распределений.

Достаточность. Наши рассуждения вполне аналогичны только что проведенным в рассмотренном выше примере. Имеем

$$\mathbf{P}_\theta(X=\vec{z}|S=s_i)=\frac{\mathbf{P}_\theta(X=\vec{z},S=s_i)}{\mathbf{P}_\theta(S=s_i)}=\begin{cases}\dfrac{\prod\limits_{i=1}^{n}\mathbf{P}_\theta(x_i=z_i)}{\mathbf{P}_\theta(S=s_i)}, & s_i=S(\vec{z}),\\[2ex] 0, & s_i\neq S(\vec{z}).\end{cases}$$

Запишем функцию правдоподобия для дискретных распределений: $\Psi_{\vec{z}}(\theta)=\prod\limits_{i=1}^{n}\mathbf{P}_\theta(x_i=z_i)$. Исходя из условий теоремы,

$$\Psi_{\vec{z}}(\theta)=\varphi(S(\vec{z}),\theta)h(\vec{z})\quad \text{п.н.}$$

Таким образом, в случае $s_i=S(\vec{z})$ имеем

$$\mathbf{P}_\theta(X=\vec{z}|S=s_i)=\frac{\varphi(S(\vec{z}),\theta)h(\vec{z})}{\sum\limits_{\vec{y}:S(\vec{y})=s_i}\mathbf{P}_\theta(X=\vec{y})}=\frac{\varphi(S(\vec{z}),\theta)h(\vec{z})}{\sum\limits_{\vec{y}:S(\vec{y})=s_i}\varphi(S(\vec{y}),\theta)h(\vec{y})}=\frac{h(\vec{z})}{\sum\limits_{\vec{y}:S(\vec{y})=s_i}h(\vec{y})}.$$

Последнее выражение не зависит от параметра $\theta$, что и требовалось показать. □

**Упражнение.** *Доказать необходимость приведенной теоремы.*

**Пример.** Функция правдоподобия $\Psi_X(\lambda)$ для пуассоновского семейства распределений $\{\pi_\lambda\}_{\lambda>0}$ имеет вид

$$\Psi_X(\lambda)=\frac{\lambda^{n\overline{x}}}{\prod\limits_{i=1}^{n}x_i!}e^{-\lambda n}.$$

Положим $\varphi(\overline{x},\lambda) = \lambda^{n\overline{x}}e^{-\lambda n}$ и $h(X) = 1/\prod\limits_{i=1}^{n} x_i!$. Тогда статистика $S = n\overline{x}$ является достаточной.

**Следствие.** *Пусть $S : \mathfrak{X}^n \to \mathbb{R}^k$ – достаточная статистика. Если отображение $g : \mathbb{R}^k \to \mathbb{R}^m$ – взаимно-однозначное, то $S' = g(S)$ также будет достаточной статистикой.*

Доказательство. Действительно, $\Psi_X(\theta) = \varphi(g^{-1}(S'),\theta)h(X) = \widetilde{\varphi}(S',\theta)h(X)$ п.н. □

**Следствие.** *Оценка максимального правдоподобия является функцией от достаточной статистики.*

Доказательство. Действительно, $\hat{\theta}_n = \arg\max_\theta \varphi(S,\theta)$. Так что указанная экстремальная точка будет зависеть от выборочных значений только через статистику $S$. □

**Пример.** Вариационный ряд будет достаточной статистикой.

**Пример.** Рассмотрим параметрическое семейство нормальных распределений $\{N(\alpha,\sigma)\}$, где $\alpha \in \mathbb{R}$ и $\sigma > 0$. Построим двумерную достаточную статистику для параметра $\theta = (\alpha,\sigma)$. Вычислим функцию правдоподобия

$$\Psi_X(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum\limits_{i=1}^{n}(x_i-\alpha)^2}{2\sigma^2}\right\} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(n\overline{x^2} - n\overline{x}\alpha) - \frac{n\alpha^2}{2\sigma^2}\right\}.$$

Положим $h(X) = 1$ и

$$\varphi(S,\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(n\overline{x^2} - n\overline{x}\alpha + \frac{n\alpha^2}{2\sigma^2}\right)\right\},$$

где $S = (\overline{x^2},\overline{x})$. Согласно теореме Неймана – Фишера векторная статистика $S$ является достаточной для параметра $\theta$.

Напомним ОМП двумерного параметра нормального распределения:

$$\begin{cases} \widehat{\sigma^2} = \overline{x^2} - (\overline{x})^2, \\ \widehat{\alpha} = \overline{x}. \end{cases}$$

Статистика $S_1 = (\widehat{\alpha},\widehat{\sigma^2})$ тоже будет достаточной поскольку является взаимно-однозначным отображением уже построенной достаточной статистики $S$.

**Упражнение.** *Построить скалярную достаточную статистику для параметра $(\alpha,\sigma)$ нормального распределения.*

## Улучшение оценок с помощью достаточных статистик.

В соответствии со среднеквадратичным подходом мы ищем нижнюю огибающую для функций потерь $\delta_{\theta_n^*}(\theta) = \mathbf{E}_\theta(\theta_n^* - \theta)^2$, оценок из того или иного класса. Возникает закономерный вопрос: в каком именно классе оценок искать наилучшую? Рассмотрим класс всех оценок с конечным вторым моментом. Пусть $\theta_n^* = \mathrm{const} = \theta_0 \in \Theta$. Тогда $\delta_{\theta_n^*}(\theta) = 0$ при $\theta = \theta_0$. Так как $\theta_0$ произвольно, то это означает, что искомая нижняя огибающая должна быть тождественным нулем. В свою очередь, это означает, что искомая наилучшая оценка тождественно совпадает с параметром, чего не может быть (любая оценка не зависит от параметра!). Этот пример показывает, что необходимо сузить класс рассматриваемых оценок.

Прежде всего отметим одно замечательное свойство достаточных статистик.

**Теорема**. *Пусть $S$ — достаточная статистика. Тогда для произвольной оценки $\theta_n^*$ с конечным вторым моментом функция $\theta_S^* = \mathbf{E}(\theta_n^*|S)$ зависит только от выборки, т. е. является оценкой, причем $\delta_{\theta_n^*}(\theta) \geqslant \delta_{\theta_S^*}(\theta)$ при всех $\theta$.*

Доказательство. Заметим, что если $S$ — достаточная статистика, то условное распределение $\mathbf{P}_\theta(\theta_n^* \in A|S)$ при фиксированном $S$ не зависит от параметра $\theta$. А так как условное среднее есть известный интеграл от соответствующего условного распределения, то и оно не будет зависеть от параметра.

Далее, представим функцию потерь для $\theta_n^*$ следующим образом:

$$\delta_{\theta_n^*}(\theta) = \mathbf{E}_\theta(\theta_n^* - \theta)^2 = \mathbf{E}_\theta(\theta_n^* - \theta \pm \theta_S^*)^2 = \mathbf{E}_\theta(\theta_S^* - \theta_n^*)^2 + \delta_{\theta_S^*}(\theta) + 2\mathbf{E}_\theta(\theta_n^* - \theta_S^*)(\theta_S^* - \theta).$$

Вспомним, что для скалярной случайной величины $\xi$ и вектора сопутствующих наблюдений $\eta$ имеют место формулы: $\mathbf{E}\xi g(\eta) = \mathbf{E}(\mathbf{E}(\xi g(\eta)|\eta)) = g(\eta)\mathbf{E}(\xi|\eta)$. Положим $\xi = \theta_n^* - \theta_S^*$ и $g(\eta) = g_\theta(S) = \theta_S^* - \theta$. В результате имеем

$$\mathbf{E}_\theta(\theta_n^* - \theta_S^*)(\theta_S^* - \theta) = \mathbf{E}_\theta(g_\theta(S)\mathbf{E}(\theta_n^* - \theta_S^*|S)) = \mathbf{E}_\theta(g_\theta(S)(\mathbf{E}_\theta(\theta_n^*|S) - \theta_S^*) = 0.$$

Таким образом, $\delta_{\theta_n^*}(\theta) = \mathbf{E}_\theta(\theta_n^* - \theta_S^*)^2 + \delta_{\theta_S^*}(\theta)$. Так как $\mathbf{E}_\theta(\theta_n^* - \theta_S^*)^2 \geqslant 0$, то $\delta_{\theta_n^*}(\theta) \geqslant \delta_{\theta_S^*}(\theta)$. $\square$

## Понятие полноты статистики.

**Определение.** Статистика $S$ называется *полной относительно параметрического семейства* $\{F_\theta\}_{\theta\in\Theta}$, если для любой измеримой функции $g$ такой, что $\mathbf{E}_\theta g^2(S) < \infty$, тождество (по параметру $\theta$) $\mathbf{E}_\theta g(S) \equiv 0$ влечет за собой $g(\cdot) \equiv 0$ почти наверное относительно каждого распределения из рассматриваемого параметрического класса.

**Пример.** Рассмотрим пуассоновское параметрическое семейство $\{\pi_\lambda\}_{\lambda>0}$ и статистику $S = n\overline{x}$. Как мы уже знаем, она является достаточной для параметра $\lambda$. Исследуем ее полноту. Распределение Пуассона устойчиво относительно свертки, поэтому $n\overline{x} = \sum\limits_{i=1}^n x_i \sim \pi_{n\lambda}$. Тогда $\mathbf{E}_\lambda g(n\overline{x}) = \sum\limits_{k=0}^\infty g(k)\frac{(n\lambda)^k}{k!}e^{-n\lambda}$. Рассмотрим тождество $\mathbf{E}_\lambda g(n\overline{x}) = 0$, откуда $\sum\limits_{k=0}^\infty g(k)\frac{(n\lambda)^k}{k!} = 0$ для всех $\lambda > 0$. Если ввести обозначение $z = n\lambda$, то последнее выражение перепишется в виде эквивалентного тождества $\sum\limits_{k=0}^\infty g(k)\frac{z^k}{k!} = 0$ при всех $z > 0$. Это влечет за собой $g(k) \equiv 0$, так как степенной ряд является аналитической функцией, а из теоремы единственности следует, что все коэффициенты этого ряда равны нулю.

**Пример.** Рассмотрим параметрическое семейство $\{U[0,\theta]\}_{\theta>0}$ и ОМП $\hat{\theta}_n = x_{(n)}$. Докажем полноту этой статистики. Имеем $\mathbf{E}_\theta g(x_{(n)}) = \int\limits_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}\,dt = 0$. Следовательно, для каждого $\theta > 0$ должно быть выполнено соотношение $\int\limits_0^\theta g(t)t^{n-1}\,dt = 0$, из которого $\int\limits_{\theta_1}^{\theta_2} g(t)t^{n-1}\,dt = 0$ для всех $\theta_1 < \theta_2$. Поскольку интеграл как функция множества (области интегрирования) обладает свойством счетной аддитивности, то отсюда следует, что для любого борелевского множества $A \in \mathcal{B}$ верно аналогичное утверждение: $\int\limits_A g(t)t^{n-1}\,dt = 0$. Допустим, что $g(t) \neq 0$ почти всюду относительно меры Лебега. Тем самым, одно из множеств, $A^+ = \{\, t \mid g(t) > 0 \,\}$ или $A^- = \{\, t \mid g(t) < 0 \,\}$, имеет положительную меру. Пусть для определенности $\Lambda(A^+) > 0$. Тогда $\int\limits_{A^+} g(t)t^{n-1}\,dt > 0$. Противоречие с равенством интеграла нулю для каждого $A \in \mathcal{B}$.

**Основное свойство**. *При любом измеримом отображении полнота статистик сохраняется.*

В самом деле, пусть $S_1 = f(S)$ для некоторой измеримой функции $f$, где статистика $S$ полная. Тогда для любой интегрируемой относительно распределения $S_1$ функции $g$ получаем из тождества

$$\mathbf{E}g(S_1) = \mathbf{E}g \circ f(S) \equiv 0,$$

что $g \circ f = 0$ почти наверное относительно распределения статистики $S$ для любого значения параметра, что и означает тождественное равенство нулю функции $g$ почти наверное относительно распределения $S_1$.

## Построение эффективных оценок.

**Определение.** Пусть задан класс $K$ оценок с конечными вторыми моментами, т. е. $K \subseteq L_2(\Omega, \mathbf{P})$. Оценка $\breve{\theta}_n$ называется *эффективной в классе* $K$, если для любой оценки $\theta_n^* \in K$ выполняется неравенство $\delta_{\breve{\theta}_n}(\cdot) \leqslant \delta_{\theta_n^*}(\cdot)$ для всех $\theta$.

Обозначим через $K_b = \{\, \theta_n^* \mid \mathbf{E}_\theta \theta_n^* = \theta + b_n(\theta) \,\}$ класс всех оценок с фиксированным смещением $b_n(\theta)$.

**Теорема**. *Пусть $S$ — полная достаточная статистика. Тогда $\theta_S^*$ является единственной эффективной оценкой в классе $K_b$.*

Доказательство. В предыдущем пункте уже было установлено, что $\theta_S^*$ является оценкой (т.е. не зависит от параметра $\theta$). Используя формулу полной вероятности для условного математического ожидания, проверим, что $\theta_S^* \in K_b$:

$$\mathbf{E}_\theta \theta_S^* = \mathbf{E}_\theta(\mathbf{E}(\theta_n^*|S)) = \mathbf{E}_\theta \theta_n^* = \theta + b_n(\theta).$$

Покажем, что в классе $K_b$ существует единственная оценка, представимая в виде: $\tilde{\theta}_n = \tilde{g}(S)$. Допустим, что это не так, т. е. существует еще и другая оценка $\widehat{g}(S) \in K_b$. Обозначим $g^0 = \tilde{g} - \widehat{g}$. Так как статистика $S$ — полная, то $\mathbf{E}_\theta g^0(S) = \mathbf{E}_\theta \tilde{g}(S) - \mathbf{E}_\theta \widehat{g}(S) \equiv 0$, и значит, функции $\tilde{g}$ и $\widehat{g}$ совпадают почти всюду относительно каждого распределения рассматриваемого параметрического класса.

Далее, предположим, что имеется оценка $\theta_n' \in K_b$ такая, что существует $\theta_0 \in \Theta$: $\delta_{\theta_n'}(\theta_0) < \delta_{\theta_S^*}(\theta_0)$. Построим для $\theta_n'$ оценку $g'(S) = \theta_S' = \mathbf{E}(\theta_n'|S)$. В силу теоремы предыдущего пункта, для любого $\theta \in \Theta$ выполнено: $\delta_{\theta_S'}(\theta) \leqslant \delta_{\theta_n'}(\theta)$. В частности, в точке $\theta_0$:

$$\delta_{\theta_S'}(\theta_0) \leqslant \delta_{\theta_n'}(\theta_0) < \delta_{\theta_S^*}(\theta_0).$$

Последнее неравенство противоречиво, так как по доказанному выше, оценки $\theta_S^*$ и $\theta_S'$ должны совпадать почти всюду для всевозможных значений параметра $\theta$, а следовательно, должны совпадать их функции потерь. Таким образом, $\delta_{\theta_n'}(\theta) = \delta_{\theta_S^*}(\theta)$ для всех $\theta \in \Theta$. Осталось показать, что оценки $\theta_n'$ и $\theta_S^*$ совпадают как отображения. Действительно,

$$\begin{aligned}
\mathbf{E}_\theta(\theta_n' - \theta_S^*)^2 = \mathbf{E}_\theta(\theta_n' \pm \theta - \theta_S^*)^2 &= \delta_{\theta_n'}(\theta) + \delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_n' - \theta)(\theta_S^* - \theta) = \\
&= \delta_{\theta_n'}(\theta) + \delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_S^* - \theta)\mathbf{E}((\theta_n' - \theta)|S) = \\
&= \delta_{\theta_n'}(\theta) + \delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_S^* - \theta)(\mathbf{E}(\theta_n'|S) - \theta) = \\
&= \delta_{\theta_n'}(\theta) + \delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_S^* - \theta)(\theta_S' - \theta) = \\
&= \delta_{\theta_n'}(\theta) + \delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_S^* - \theta)(\theta_S^* - \theta) = 2\delta_{\theta_S^*}(\theta) - 2\mathbf{E}_\theta(\theta_S^* - \theta)^2 = 0. \ \square
\end{aligned}$$

**Пример.** Рассмотрим параметрическое семейство $\{\pi_\lambda\}_{\lambda>0}$. Статистика $S = n\overline{x}$ является полной и достаточной. Так как $\mathbf{E}_\lambda(\overline{x}|n\overline{x}) = \overline{x}$, то $\overline{x}$ будет единственной эффективной оценкой в классе $K_0$.

**Следствие.** *Пусть $\check{\theta}_n$ — полная достаточная оценка. Тогда $\check{\theta}_n$ является единственной эффективной оценкой в классе $K_b$.*

Доказательство. $\mathbf{E}_\theta(\check{\theta}_n|\check{\theta}_n) = \check{\theta}_n$. $\square$

**Следствие.** *Пусть для некоторого параметрического семейства распределений существуют полная достаточная статистика $S$, а также ОМП, которая имеет конечный второй момент для любого распределения из рассматриваемого параметрического класса. Тогда указанная ОМП будет единственной эффективной оценкой в соответствующем классе $K_b$.*

Доказательство. В качестве следствия факторизационной теоремы Неймана — Фишера мы уже отмечали, что $\hat{\theta}_n = g(S)$. Но в силу известных свойств условного математического ожидания $\mathbf{E}_\theta(\hat{\theta}_n|S) = g(S)$ п.н. $\square$

**Пример.** Рассмотрим параметрическое семейство $\{U[0,\theta]\}_{\theta>0}$. Согласно следствию, ОМП $x_{(n)}$ — единственная эффективная оценка в классе $K_b$, где $b_n(\theta) = -\frac{\theta}{n+1}$. Очевидно также, что оценка $\theta_n^* = \frac{n+1}{n}x_{(n)}$ является единственной эффективной в классе $K_0$.

## Понятие $R$-эффективности.

В этом разделе мы получим нижнюю оценку для функции потерь $\delta_{\theta_n^*}(\theta) = \mathbf{E}_\theta(\theta_n^* - \theta)^2$ в параметрическом семействе плотностей при некоторых условиях регулярности $(R)$, которые уточним позже.

**Теорема (неравенство Рао—Крамера).** *Пусть выполнены условия регулярности $(R)$. Тогда для любой оценки $\theta_n^* \in K_b = \{\theta_n^* \mid \boldsymbol{E}\theta^* = \theta + b_n(\theta)\}$ справедливо неравенство*

$$\delta_{\theta_n^*}(\theta) \geqslant \frac{(1 + b_n'(\theta))^2}{nI(\theta)} + b_n^2(\theta),$$

*где $I(\theta) = \mathbf{E}_\theta(l'_{x_1}(\theta))^2$ — информация Фишера (или информация о неизвестном параметре, содержащаяся в одном наблюдении; здесь $l'_x(\theta) = \frac{d}{d\theta}\log f_\theta(x)$).*

Доказательство. Имеем

$$\mathbf{E}_\theta\theta_n^*(X) = \int\limits_{\mathcal{X}^n}\cdots\int \theta_n^*(\overrightarrow{z})\Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) = G(\theta),$$

где $\Psi_{\overrightarrow{z}}(\theta)$ — функция правдоподобия, а $\lambda^n$ — $n$-кратная продакт-мера. Продифференцируем $G$ по параметру. Но для этого нам необходимы дополнительные ограничения. Например, для возможности дифференцирования можно потребовать, чтобы параметрическое семейство плотностей было бы непрерывно дифференцируемым по параметру, и равномерно по $\theta' \in (\theta - \varepsilon, \theta + \varepsilon)$ должно выполняться неравенство

$$|L'_{\overrightarrow{z}}(\theta')| \leqslant g_{\theta,\varepsilon}(\overrightarrow{z})$$

для каждого $\theta \in \Theta$ и некоторого положительного $\varepsilon = \varepsilon(\theta)$, причем

$$\mathbf{E}g^2_{\theta,\varepsilon}(X) = \int\cdots\int g^2_{\theta,\varepsilon}(\overrightarrow{z})\Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) < \infty.$$

Вот эти условия и составляют $(R)$. При этих условиях можно внести производную под знак интеграла:

$$\frac{d}{d\theta}G(\theta) = \int\limits_{\mathcal{X}^n}\cdots\int \theta_n^*(\overrightarrow{z})\Psi'_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) = \int\limits_{\mathcal{X}^n}\cdots\int \theta_n^*(\overrightarrow{z})\frac{\Psi'_{\overrightarrow{z}}(\theta)}{\Psi_{\overrightarrow{z}}(\theta)}\Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) =$$

$$= \int\limits_{\mathcal{X}^n} \cdots \int \theta_n^*(\overrightarrow{z}) L'_{\overrightarrow{z}}(\theta) \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z}) = \mathbf{E}_\theta(\theta_n^* L'_X(\theta)).$$

Обоснование последнего равенства. Применяя неравенство Коши — Буняковского и формулу конечных приращений, мы с помощью условий $(R)$ получим для любого $\Delta \in (\theta - \varepsilon, \theta + \varepsilon)$

$$\int \cdots \int \theta_n^* \frac{L_{\overrightarrow{z}}(\theta + \Delta) - L_{\overrightarrow{z}}(\theta)}{\Delta} \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z})$$

$$\leqslant \sqrt{\int \cdots \int (\theta_n^*(\overrightarrow{z}))^2 \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z})} \sqrt{\int \cdots \int g^2_{\theta,\varepsilon}(\overrightarrow{z}) \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z})} < \infty,$$

так как мы рассматриваем только оценки с конечным вторым моментом:

$$\mathbf{E}_\theta(\theta_n^*)^2 = \int \cdots \int (\theta_n^*(\overrightarrow{z}))^2 \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z}) < \infty.$$

Тем самым, мы построили не зависящую от $\Delta$ интегрируемую мажоранту для разностного аналога производной. Остается воспользоваться классической теоремой Лебега о мажорируемой сходимости, которая позволяет нам внести оператор дифференцирования (т. е. предел по $\Delta$) под знак интеграла.

Далее, поскольку $G(\theta) = \theta + b_n(\theta)$, то

$$\frac{d}{d\theta} G(\theta) = \mathbf{E}_\theta(\theta_n^* L'_{\overrightarrow{z}}(\theta)) = 1 + b'_n(\theta).$$

Заметим, что для вырожденной случайной величины $\xi \equiv 1$

$$\mathbf{E}1 = \theta - \theta + 1, \; b = 1 - \theta, \; b' = -1,$$

т. е. $\mathbf{E}_\theta \theta_n^* L'_{\overrightarrow{z}}(\theta) = 1 + b'_n(\theta) = \mathbf{E}_\theta L'_{\overrightarrow{z}}(\theta) = 0$. Тогда получаем пару тождеств

$$\begin{cases} \mathbf{E}_\theta L'_X(\theta) = \mathbf{E}_\theta \mathbf{E} \theta_n^* L'_X(\theta) = 0, \\ \mathbf{E}_\theta \theta_n^* L'_X(\theta) = 1 + b'_n(\theta), \end{cases}$$

откуда $(\mathbf{E}_\theta(\theta_n^* - \mathbf{E}\theta_n^*) L'_X(\theta))^2 = (1 + b'_n(\theta))^2$. Применив к левой части этого тождества неравенство Коши — Буняковского $(\mathbf{E}\xi\eta)^2 \leqslant \mathbf{E}\xi^2 \mathbf{E}\eta^2$, получим

$$(\mathbf{E}_\theta(\theta_n^* - \mathbf{E}\theta_n^*) L'_X(\theta))^2 \leqslant \mathbf{D}\theta_n^* \cdot \mathbf{E}(L'_X(\theta))^2.$$

А так как $\mathbf{E}L'_X(\theta) = 0$ и $\mathbf{E}L'_X(\theta) = \sum\limits_{i=1}^n \mathbf{E} l'_{x_i}(\theta)$, то у случайной величины $L'_X(\theta)$ второй момент совпадает с дисперсией. Дисперсия в классе независимых случайных величин аддитивна:

$$\mathbf{D}L'_X(\theta) = \sum_{i=1}^n \mathbf{D} l'_{x_i}(\theta) = n \mathbf{D} l'_{x_1}(\theta) = n I(\theta).$$

В итоге получаем неравенство

$$\mathbf{D}_\theta \theta_n^* \geqslant \frac{(1 + b'_n(\theta))^2}{n I(\theta)}.$$

Найдём связь между дисперсией и функцией потерь:

$$\delta_{\theta_n^*} = \mathbf{E}_\theta(\theta_n^* - \theta \pm \mathbf{E}\theta_n^*)^2 = \mathbf{D}\theta_n^* + b_n^2(\theta),$$

так как соответствующий второй смешанный момент равен 0. □

Заметим, что единственное неравенство в приведенном доказательстве появляется при использовании неравенства Коши — Буняковского. Равенство в нём возможно, когда либо $\xi$ и $\eta$ линейно связаны, либо по крайней мере одна из этих величин тождественно равна нулю ($\theta_n^* \equiv const$).

**Определение.** Оценка $\theta_n^*$ из класса $K_b$ называется *R-эффективной,* если в неравенстве Рао — Крамера достигается равенство.

**Следствие**. *Для того, чтобы в неравенстве Рао — Крамера достигалось равенство, необходимо и достаточно, чтобы либо $\theta_n^* \equiv const$ п. н., либо*

$$L'_X(\theta) = c_n(\theta)(\theta_n^* - \mathbf{E}\theta_n^*) \quad \text{п. н.},$$

*где $c_n(\theta)$ — константа.*

Малосодержательный случай $\theta_n^* \equiv const$ п. н. мы исключаем из рассмотрения. Тогда

$$L_X(\theta) = A_n(\theta)\theta_n^* + B_n(\theta) + S(X),$$

где $A_n, B_n$ — новые константы, а $S$ — некоторая статистика (т. е. не зависящая от параметра функция от выборки). Потенциируя последнее равенство, получаем, что с необходимостью функция правдоподобия должна допускать следующую факторизацию:

$$\Psi_X(\theta) = \exp\{A_n(\theta)\theta_n^* + B_n(\theta)\}h(X) \quad \text{п. н.},$$

что является *критерием R-эффективности.*

Отметим, что

1. $R$-эффективная оценка $\theta_n^*$ всегда будет достаточной.

2. Правая часть в неравенстве Рао — Крамера зависит только от смещения $b$ и параметрического семейства.

3. Для регулярных параметрических семейств из $R$-эффективности следует эффективность (обратное, вообще говоря, неверно).

**Пример.** Рассмотрим параметрическое семейство распределений Пуассона $\{\Pi_\lambda\}$. Мы знаем, что $\overline{x}$ — эффективная оценка в $K_0$. Функция правдоподобия

$$\Psi_X(\lambda) = \frac{\lambda^{n\overline{x}}}{\prod x_i!}e^{-n\lambda} = \exp n\overline{x}\log\lambda - n\lambda h(x).$$

Тогда $\overline{x}$ является $R$-эффективной по критерию. Проверим $R$-эффективность по определению. Имеем:

$$\theta_n^* = \overline{x} \Rightarrow \delta_{\overline{x}}(\theta) = \mathbf{D}\overline{x} = \frac{\lambda}{n}, \quad I(\lambda) = \mathbf{D}l'_{x_1}(\lambda),$$
$$f_\lambda(x) = \frac{\lambda^x}{x!}e^{-\lambda} \Rightarrow l'_{x_1}(\lambda) = \frac{x}{\lambda} - 1 \Rightarrow \mathbf{D}l'_{x_1}(\lambda) = \frac{1}{\lambda}.$$

Таким образом, в неравенстве Рао — Крамера достигается равенство.

**Пример.** Рассмотрим нормальное распределение $\{N_{(\alpha,1)}\}$. Оценка $\overline{x}$ является эффективной в классе $K_0$. Функция правдоподобия

$$\Psi_X(\alpha) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum(x_i-\alpha)^2\right\} = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(-2\overline{x}\alpha + n\alpha^2)\right\}h(\overline{x}).$$

Итак, и для этого параметрического семейства $\overline{x}$ — $R$-эффективная оценка.

**Контрпример.** Рассмотрим показательное распределение $\{E_\alpha\}$

$$f_\alpha(x) = \begin{cases} \alpha e^{-\alpha x},\ x \geqslant 0, \\ 0,\ x < 0. \end{cases}$$

функция правдоподобия

$$\Psi_X(\alpha) = \alpha^n e^{-\alpha n \overline{x}} = \exp\{-\alpha n \overline{x} + n \log \alpha\} = \exp\left\{-\alpha n \frac{1}{(1/\overline{x})} + n \log \alpha\right\}.$$

Значит, оценка $\hat{\alpha}_n = 1/\overline{x}$ не является $R$-эффективной, хотя является эффективной в соответствующем классе $K_b$.

**Упражнение.** *Проверить по определению, что оценка $\hat{\alpha}_n$ не является $R$-эффективной (напомним, что если $x_i$ имеет распределение $E_\alpha$, то $n\overline{x}$ имеет распределение $\Gamma_{n,\alpha}$).*

## Связь $R$-эффективных оценок с ОМП.

**Теорема.** *При выполнении условий регулярности $(R)$ если $\theta_n^*$ — $R$-эффективная оценка в классе $K_0$, то она совпадает с ОМП: $\theta_n^* = \hat{\theta}_n$.*

Доказательство. Так как $\theta_n^* \in K_0$, то $\mathbf{D}_\theta \theta_n^* = \mathbf{E}_\theta(\theta_n^* - \theta)^2 = \frac{1}{nI(\theta)}$. При этом согласно следствию имеет место представление

$$L_X'(\theta) = c(\theta)(\theta_n^* - \theta). \tag{1}$$

Так как $b_n(\theta) = 0$, то $b_n'(\theta) = 0$. Тогда из доказательства неравенства Рао — Крамера имеем

$$\mathbf{E}_\theta(\theta_n^* - \theta) L_X'(\theta) = 1. \tag{2}$$

Подставляя (1) в (2), получаем $c(\theta)\mathbf{E}(\theta_n^* - \theta)^2 = 1$. Стало быть, коэффициент $c(\theta)$ положителен, т. е. $L_X'$ меняет знак с плюса на минус при переходе через точку $\theta = \theta_n^*$. Таким образом, точка $\theta_n^*$ доставляет максимум логарифмической функции правдоподобия. Значит, $\theta_n^* = \hat{\theta}_n$. $\square$

## Байесовский подход к построению оценок. Байесовские оценки.

До сих пор в качестве критерия точности оценки мы рассматривали функцию потерь

$$\delta_{\theta_n^*}(\theta) = \mathbf{E}_\theta(\theta_n^* - \theta)^2.$$

При достаточно широких условиях мы построили оценки, для которой функция потерь является нижней огибающей для всех функций потерь из некоторого класса. Но для построения этой оценки необходимым условием является полнота, а это условие трудно проверить.

Идея Байеса состоит в сравнении *усредненных* (или *взвешенных*) функций потерь.

**Определение.** Пусть на параметрическом множестве $(\Theta, \mathcal{F})$ задана конечная мера $Q(\cdot)$. Тогда *усредненная функция потерь* есть интеграл

$$\delta_Q(\theta_n^*) = \int_\Theta \delta_{\theta_n^*}(\theta)\, Q(d\theta).$$

**Определение.** Оценка $\theta_{n,1}^*$ *«лучше»* $\theta_{n,2}^*$ ($\theta_{n,1}^* \prec \theta_{n,2}^*$) относительно меры $Q$, если $\delta_Q(\theta_{n,1}^*) < \delta_Q(\theta_{n,2}^*)$.

Достоинством данного подхода является то, что в классе всех оценок с двумя конечными моментами (это самый широкий из классов, которые мы можем рассматривать) существует оптимальная оценка, минимизирующая усредненную функцию потерь. И эта *байесовская оценка* определяется параметрическим семейством и усредняющей мерой.

Заметим, что без ограничения общности можно предполагать, что $Q$ — вероятностная мера (так как всегда можно ввести в рассмотрение меру $Q(\cdot)/Q(\Theta)$). Предполагаем в дальнейшем, что у этой меры существует плотность $q(\theta)$ относительно $\sigma$-конечной меры $\mu(d\theta)$ при условии, что $\Theta$ — это область в $\mathbb{R}$ и что задано параметрическое семейство плотностей $\{f_\theta\}_{\theta\in\Theta}$ относительно $\sigma$-конечной меры $\lambda(\cdot)$ на $\mathfrak{X}$.

**Теорема.** *Для любого $Q$ существует байесовская оценка $\theta^*_Q : \delta_Q(\theta^*_Q) = \inf\limits_{\theta^*_n} \delta_Q(\theta^*_n)$.*

Доказательство. По определению усредненной функции потерь для любой оценки $\theta^*_n$

$$\delta_\theta(\theta^*_n) = \int\limits_{\Theta} \left( \int \cdots \int\limits_{\mathfrak{X}^n} (\theta^*_n(\overrightarrow{z}) - \theta)^2 \Psi_{\overrightarrow{z}}(\theta) \lambda^n(d\overrightarrow{z}) \right) q(\theta)\mu(d\theta).$$

Так как интегрант этого повторного интеграла неотрицательная функция, то по теореме Фубини данный интеграл можно рассматривать как кратный на множестве $\Theta \times \mathfrak{X}^n$. Далее заметим, что $q(\theta)\Psi_{\overrightarrow{z}}(\theta))$ будет плотностью на расширенном пространстве $\Theta \times \mathfrak{X}^n$ относительно новой продакт-меры $\lambda^n(\cdot)\mu(\cdot)$, так как $\int \cdots \int\limits_{\mathfrak{X}^n} \Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) = 1$. Следовательно, по теореме Фубини

$$\int \cdots \int\limits_{\Theta\times\mathfrak{X}^n} \Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z})q(\theta)\mu(d\theta) = \int\limits_{\Theta} q(\theta) \left( \int \cdots \int\limits_{\mathfrak{X}^n} \Psi_{\overrightarrow{z}}(\theta)\lambda^n(d\overrightarrow{z}) \right) \mu(d\theta) = \int\limits_{\Theta} q(\theta)\mu(d\theta) = 1,$$

так как $Q$ есть вероятностная мера. Тогда $\delta_Q(\theta^*_n) = \widetilde{\mathbf{E}}(\theta^*_n - \theta)^2$, где $\theta^*_n$ и $\theta$ — *случайные величины* на введенном расширенном вероятностном пространстве, а мера $Q$ будет распределением (так называемым *априорным*) параметра $\theta$ . Мы ищем оценку $\widetilde{\theta^*}$, минимизирующую выражение $\widetilde{\mathbf{E}}(\theta^*_n - \theta)^2$.

Вспомним, что если $\xi$ — скалярная случайная величина, $\eta$ — вектор сопутствующих наблюдений, то на ортопроекции $\widehat{g}(\eta) = \mathbf{E}(\xi|\eta) = \widehat{\xi}|_{L(\eta)}$, где $L(\eta) = \{g(\eta) \mid \mathbf{E}g^2(\eta) < \infty\}$ реализуется расстояние от элемента $\xi$ до подпространства $L(\eta)$, т. е.

$$\inf_{g(\eta)\in L(\eta)} \mathbf{E}(\xi - g(\eta))^2 = \mathbf{E}(\xi - \mathbf{E}(\xi|\eta))^2.$$

И это соотношение верно на любом вероятностном пространстве.

В нашем случае мы ищем $\inf\limits_{\theta^*_n} \widetilde{\mathbf{E}}(\theta^*_n - \theta)^2$. Пусть $\theta^*_n(X) = g(X) \in L(X)$, т. е. считаем выборку вектором сопутствующих наблюдений. Попадаем в условия вышеописанной модели. Тогда существует единственная в $L_2$ оценка $\widetilde{\theta^*_n} = \widetilde{\mathbf{E}}(\theta|X)$. Осталось понять, как считать $\widetilde{\mathbf{E}}(\theta|X)$.

Напомним, что если пара $(\xi, \eta)$ имеет абсолютно непрерывное распределение с плотностью совместного распределения $p_{\xi,\eta}(t, \overrightarrow{z})$, то $\widehat{g}(\eta) = \mathbf{E}(\xi|\eta) = \int\limits_{\mathbb{R}} tp_\xi(t|\eta)\lambda(dt)$, где условная плотность распределения вычисляется по формуле Байеса

$$p_\xi(t|\eta) = \frac{p_{\xi,\eta}(t,\eta)}{p_\eta(\eta)} \text{ и } p_\eta(\overrightarrow{z}) = \int p_{\xi,\eta}(t, \overrightarrow{z})\lambda(dt).$$

В нашем случае $\widetilde{\theta_n^*} = \widetilde{\mathbf{E}}(\theta|X) = \int\limits_{\Theta} tq(t|X)\mu(dt)$, где $q(t|X)$ — *апостериорная* плотность распределения параметра при фиксации выборки:

$$q(t|X) = \frac{q(t)\Psi_X(t)}{\int\limits_{\Theta} q(s)\Psi_X(s)\mu(ds)}.$$

Таким образом, задача решена полностью и дан алгоритм построения байесовских оценок.

Байесовский подход не лишен субъективизма, так как выбор $Q$ в наших рассуждениях никак не был обоснован. Меру $Q$ иногда называют мерой уверенности исследователя.

**Пример.** Рассмотрим семейство нормальных распределений $\{N_{(\alpha,1)}\}$ с $\alpha \in N_{(0,\sigma)}$, т. е. считаем, что параметр $\alpha$ находится в окрестности нуля, являясь положительным и отрицательным с одинаковыми шансами (с точки зрения исследователя).

Выпишем функцию правдоподобия

$$\Psi_x(t) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum(x_i - t)^2\right) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{n\overline{x^2}}{2} + tn\overline{x} - \frac{nt^2}{2}\right).$$

Посчитаем апостериорную плотность. Так как $q(t) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{t^2}{2\sigma^2}\right)$, то выделяя полный квадрат по переменной $t$, получим

$$q(t)\Psi_X(t) = \widetilde{C_x} \exp\left(-\frac{1}{2}\left(t - \frac{\overline{x}n}{n + 1/\sigma^2}\right)^2\right).$$

Мы получили плотность нормального распределения с параметром сноса $\dfrac{\overline{x}n}{n + 1/\sigma^2}$. Тогда выражение для математического ожидания есть

$$\widetilde{\mathbf{E}}(\theta|x) = \int\limits_{\Theta} tq(t|x)\mu(dt) = \frac{\overline{x}n}{n + 1/\sigma^2} = \frac{\overline{x}}{1 + 1/n\sigma^2}.$$

При больших значениях величины $n\sigma^2$ получаем близость оценки к ОМП $\overline{x}$.

## Интервальное оценивание. Метод доверительных интервалов.

До сих пор мы имели дело с параметрическим семейством $\{F_\theta\}$ (возможно, представленным плотностями) и строили функцию от выборки $\theta_n^*(X)$ той же размерности, что и параметр — т. е. точку, в том или ином смысле приближающую истинное значение параметра наблюдаемого распределения. Была задана мера точности — функция потерь $\delta_{\theta_n^*}(\theta)$. При интервальном оценивании в качестве оценки неизвестного параметра предлагается не одна точка, а целое множество.

**Определение.** Пара упорядоченных статистик $(\theta_n^-(X), \theta_n^+(X))$ называется *доверительным интервалом уровня доверия* $1 - \varepsilon$ (иногда говорят *уровня* $\varepsilon$), если при всех объемах наблюдений справедлива следующая оценка вероятности покрытия случайным интервалом неизвестного параметра

$$P(\theta_n^- < \theta < \theta_n^+) \geqslant 1 - \varepsilon.$$

При знаке равенства получаем *минимальный доверительный интервал.*

Это определение означает, что при малых $\varepsilon$ (типичные порядки $\varepsilon = 10^{-2} \div 10^{-4}$) событие под знаком вероятности практически достоверно, т. е. с очень высокой вероятностью реализуется двойное неравенство $\theta_n^- < \theta < \theta_n^+$ (т. е. по существу можно убрать знак вероятности!). Конструкция оправдана, если доверительный интервал достаточно мал и его длина $\underset{\text{п.н.}}{\longrightarrow} 0$ при $n \to \infty$. В данном методе мы получаем не усредненную характеристику погрешности оценки, а практически детерминированную.

**Пример** (Иллюстрация основного приема построения доверительных интервалов). Рассматриваем однопараметрическое семейство $\{N_{\alpha,1}\}$ с неизвестным параметром сдвига. Задача: построить доверительный интервал заданного уровня для $\alpha$.

1-й шаг построения. Рассмотрим функционал $G(\alpha, X) = \sqrt{n}(\overline{X} - \alpha) \in N_{(0,1)}$ – стандартный нормальный закон. Хотя $G(\alpha, X)$ явно зависит от $\alpha$, но его *распределение от $\alpha$ не зависит*.

2-й шаг. Отметим монотонность и непрерывность по $\alpha$ этого функционала (при этом знак монотонности должен быть одинаков для всех $X$; в данном случае – убывание). Далее, по таблицам $N_{0,1}$ построим симметричный двойной барьер (в данном случае распределение симметрично)

$$P(-t_\varepsilon < G(\alpha, X) < t_\varepsilon) = 1 - \varepsilon,$$

$$P(-t_\varepsilon < G(\alpha, X) < t_\varepsilon) = \Phi(t_\varepsilon) - \Phi(-t_\varepsilon) = 1 - 2\Phi(-t_\varepsilon) = 1 - \varepsilon.$$

Таким образом, построение доверительного интервала в данном примере сводится к решению уравнения $\Phi(-t_\varepsilon) = \dfrac{\varepsilon}{2}$, где $\Phi(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ – функция Лапласа (здесь мы использовали свойство симметрии этой функции: $\Phi(x) = 1 - \Phi(-x)$).

3-й шаг. Неравенство $-t_\varepsilon < G(\alpha, X) < t_\varepsilon$ разрешаем относительно $\alpha$ в силу монотонности $G(\alpha, \cdot)$:

$$\overline{x} - \frac{t_\varepsilon}{\sqrt{n}} < \alpha < \overline{x} + \frac{t_\varepsilon}{\sqrt{n}}.$$

А поскольку $P(-t_\varepsilon < G(\alpha, X) < t_\varepsilon) = 1 - \varepsilon$, то с той же вероятностью будет выполнено и вышеприведенное двойное неравенство. Таким образом, мы построили доверительный интервал уровня доверия $1 - \varepsilon$ для $\alpha$ с границами $\alpha_n^- = \overline{x} - \frac{t_\varepsilon}{\sqrt{n}}$, $\alpha_n^+ = \overline{x} + \frac{t_\varepsilon}{\sqrt{n}}$.

**Замечание.** Если рассматривать не $\{N_{(\alpha,1)}\}$, а $\{N_{(\alpha,\sigma_0)}\}$, где $\sigma_0$ известно, то используем следующий функционал $G(\alpha, X) = \dfrac{\sqrt{n}(\overline{x} - \alpha)}{\sigma_0} \in N_{(0,1)}$. В итоге получим

$$P\left(\overline{x} - \frac{\sigma_0 t_\varepsilon}{\sqrt{n}} < \alpha < \overline{x} + \frac{\sigma_0 t_\varepsilon}{\sqrt{n}}\right) = 1 - \varepsilon.$$

Ширина доверительного интервала в данном случае $2\sigma_0 t_\varepsilon/\sqrt{n}$, т. е. погрешность приближения параметра имеет порядок $1/\sqrt{n}$.

Теперь приведем теорему, описывающую общую схему построения доверительных интервалов, все основные элементы которой содержатся в вышеприведенном примере.

**Теорема.** *Пусть для данного параметрического семейства $F_\theta$ существует функционал $G(\theta, X)$, распределение которого не зависит от $\theta$. Причем функционал $G(\theta, X)$ имеет строгую монотонность одного знака по $\theta$ для всевозможных значениях выборки и непрерывен по параметру. Тогда для наперед заданного $\varepsilon > 0$ пара (неупорядоченная) чисел $\{G^{-1}(t_\varepsilon^{(1)}, X), G^{-1}(t_\varepsilon^{(2)}, X)\}$ с условием $P(t_\varepsilon^{(1)} < G(\theta, X) < t_\varepsilon^{(2)}) \geqslant 1 - \varepsilon$ образует доверительный интервал уровня $1 - \varepsilon$, где $G^{-1}(t, X)$ – обратная функция по первому аргументу.*

Доказательство. Проведя эквивалентное преобразование неравенств при монотонно возрастающей функции $G$ мы получим

$$P(t_\varepsilon^{(1)} < G(\theta, X) < t_\varepsilon^{(2)}) = P(G^{-1}(t_\varepsilon^{(1)}, X) < \theta < G^{-1}(t_\varepsilon^{(2)}, X)) \geqslant 1 - \varepsilon.$$

Для монотонно убывающей функции $G$ индексы (1) и (2) нужно поменять местами. $\square$

**Пример.** Рассматриваем нормальное распределение $\{N_{0,\sigma}\}$, где $\sigma$ — неизвестно. Построим доверительный интервал для $\sigma^2$. Строим функционал $G(\sigma, X) = \dfrac{n\mathbf{S}_0^2}{\sigma^2}$, где $\mathbf{S}_0^2 = \frac{1}{n}\sum\limits_{i=1}^n x_i^2 \equiv \overline{x^2}$ — второй выборочный момент. Следовательно, $G(\sigma, X) = \sum\limits_{i=1}^n \dfrac{x_i^2}{\sigma^2} = \sum\limits_{i=1}^n \xi_i^2$, где $\xi_i \in N_{(0,1)}$, $\xi_i$ — независимы.

**Определение.** Сумма $m$ квадратов независимых одинаково распределенных величин $\xi_i$, имеющих стандартное нормальное распределение, имеет распределение $\chi_m^2$ — *Хи - квадрат с $m$ степенями свободы*.

Тогда $G(\sigma, X) \in \chi_n^2$. По таблицам находим точки $t_\varepsilon^{(1)}$ и $t_\varepsilon^{(2)}$ такие, что $P(t_\varepsilon^{(1)} < G(\sigma, X) < t_\varepsilon^{(2)}) = 1 - \varepsilon$. Проведя эквивалентные преобразования неравенств получим следующий доверительный интервал:

$$\left(\frac{n\mathbf{S}_0^2}{t_\varepsilon^{(2)}}, \frac{n\mathbf{S}_0^2}{t_\varepsilon^{(1)}}\right).$$

# Построение доверительных интервалов для параметров нормального распределения. Случай двух неизвестных параметров.

**Лемма** (Р. Фишер). *Пусть $X$ — стандартный нормальный вектор, а $C = \|c_{ij}\|_{n\times n}$ — ортогональная матрица. Тогда $Y = XC$ является стандартным нормальным вектором.*

Доказательство. Пусть $Y = (y_1, \ldots, y_n)$. Заметим, что $y_k = \sum\limits_{i=1}^n x_i c_{ik}$ для $k = 1, \ldots, n$. В силу устойчивости нормального распределения все $y_k$ нормальны с параметрами $\alpha$ и $\sigma^2$. Определим эти параметры, вспомнив их вероятностный смысл — математического ожидания и дисперсии:

$$\alpha = \mathbf{E}y_k = \mathbf{E}\Big(\sum_{i=1}^n x_i c_{ik}\Big) = \sum_{i=1}^n c_{ik}\mathbf{E}x_i = 0;$$

$$\sigma^2 = \mathbf{D}y_k = \mathbf{E}y_k^2 = \mathbf{E}\Big(\sum_{i=1}^n x_i c_{ik}\Big)^2 = \sum_{i,j=1}^n \mathbf{E}x_i x_j c_{ik} c_{jk} = \sum_{i,j=1}^n \delta_{ij} c_{ik} c_{jk} = \sum_{i=1}^n c_{ik}^2 = 1.$$

Предпоследний знак равенства написан в силу того, что матрица $C$ ортогональна, и значит, ее столбцы образуют ортонормированный базис.

Осталось доказать независимость всех $y_k$ в совокупности. Согласно лемме, доказанной в курсе теории вероятностей, координаты вектора, имеющего многомерное нормальное распределение, независимы тогда и только тогда, когда они некоррелированы. Случайный вектор $Y$ нормальный, поскольку он представляет собой невырожденное линейное преобразование нормального случайного вектора $X$ (это утверждение также было доказано в курсе теории вероятностей). Вычислим матрицу ковариаций случайного вектора $Y$:

$$\mathbf{Cov}(y_k, y_m) = \mathbf{E} y_k y_m = \sum_{i,j=1}^{n} \mathbf{E} x_i x_j c_{ik} c_{jm} = \sum_{i=1}^{n} c_{ik} c_{im} = \delta_{km}.$$

Таким образом, случайный вектор $Y$ является стандартным нормальным. □

**Следствие 1.** *Пусть выполнены условия леммы Фишера и $r < n$. Тогда квадратичная форма $Q(X) = \sum_{i=1}^{n} x_i^2 - y_1^2 - \ldots - y_r^2$ не зависит от случайного вектора $(y_1, \ldots, y_r)$ и имеет распределение $\chi^2$ с $n - r$ степенями свободы.*

Доказательство. Заметим, что расстояние между любыми двумя точками в пространстве $\mathbb{R}^n$ сохраняется при ортогональных преобразованиях. Значит, длины векторов $X$ и $Y$ совпадают. Тогда $Q(X) = \sum_{i=1}^{n} x_i^2 - y_1^2 - \ldots - y_r^2 = \sum_{i=1}^{n} y_i^2 - y_1^2 - \ldots - y_r^2 = \sum_{i=r+1}^{n} y_i^2$. В силу этого представления для $Q(X)$ оба утверждения следствия очевидны. □

**Следствие 2.** *Пусть $X$ — выборочный вектор из распределения $N(\alpha, \sigma)$. Тогда*

1) $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$;

2) *случайные величины $S^2$ и $\overline{x}$ независимы.*

Доказательство. Преобразуем выборочную дисперсию:

$$\frac{S^2}{\sigma^2} = \frac{1}{n\sigma^2} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n} \sum_{i=1}^{n} (\frac{x_i - \overline{x} \pm \alpha}{\sigma})^2.$$

Рассмотрим перенормированные величины $\widetilde{x}_i = \frac{x_i - \alpha}{\sigma} \in N(0,1)$. Тогда $\frac{S^2}{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (\widetilde{x}_i - \overline{\widetilde{x}})^2 = \widetilde{S}^2$. Следовательно,

$$\frac{nS^2}{\sigma^2} = n\overline{\widetilde{x}^2} - \left(\overline{\widetilde{x}}\sqrt{n}\right)^2 = \sum_{i=1}^{n} \widetilde{x}_i^2 - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{x}_i\right)^2 = \sum_{i=1}^{n} \widetilde{x}_i^2 - y_1^2.$$

Вектор $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ можно достроить до ортонормированного базиса, а значит, можно восстановить матрицу $C$ из леммы Фишера. В силу следствия 1 получаем, что случайная величина $\frac{nS^2}{\sigma^2}$ распределена по закону $\chi^2_{n-1}$.

Квадратичная форма

$$Q(X) = \frac{nS^2}{\sigma^2} = \sum_{i=1}^{n} \widetilde{x}_i^2 - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{x}_i\right)^2 = \sum_{i=1}^{n} \widetilde{x}_i^2 - y_1^2$$

не зависит от $y_1$ опять же в силу следствия 1. Осталось вспомнить, что любые борелевские преобразования независимых случайных величин будут также независимыми, и заметить, что

$$y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i}{\sigma} - \frac{\alpha}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\overline{x} - \frac{\alpha}{\sigma\sqrt{n}}.$$ □

**Построение доверительного интервала для дисперсии.** Вышеприведенное следствие 2 указывает алгоритм построения доверительного интервала для параметра $\sigma$ при неизвестном параметре $\alpha$. Рассмотрим функционал $G(\sigma) = \frac{nS^2}{\sigma^2}$. Тогда согласно теореме о построении доверительных интервалов необходимо найти числа $t_\varepsilon^{(1)}$ и $t_\varepsilon^{(2)}$ такие, что $\mathbf{P}(t_\varepsilon^{(1)} < G(\sigma) < t_\varepsilon^{(2)}) = 1-\varepsilon$. Обычно границы интервала ищут так, чтобы выполнялись соотношения

$$\mathbf{P}(G(\sigma) < t_\varepsilon^{(1)}) = \frac{\varepsilon}{2},$$

$$\mathbf{P}(G(\sigma) < t_\varepsilon^{(2)}) = 1 - \frac{\varepsilon}{2}.$$

Для нахождения значений $t_\varepsilon^{(1)}$ и $t_\varepsilon^{(2)}$ используются таблицы распределения $\chi^2_{n-1}$.

**Построение доверительных границ для среднего.**

**Определение.** Пусть $\xi_0, ..., \xi_m$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$\frac{\xi_0}{\sqrt{\frac{1}{m}\sum\limits_{i=1}^{m}\xi_i^2}}$$

называется *распределением Стьюдента с $m$ степенями свободы* и обозначается $T_m$.

Заметим, что для независимых случайных величин $\xi_0 \in N(0,1)$ и $\eta_m \in \chi^2_m$ имеем

$$\frac{\xi_0}{\sqrt{\frac{1}{m}\sum\limits_{i=1}^{m}\xi_i^2}} \stackrel{d}{=} \frac{\xi_0}{\sqrt{\frac{1}{m}\eta_m}} \sim T_m.$$

В силу УЗБЧ имеем $\dfrac{1}{m}\sum\limits_{i=1}^{m}\xi_i^2 \xrightarrow[\text{п.н.}]{} 1$ при $m \to \infty$. Используя известные теоремы непрерывности, получаем, что распределение Стьюдента $T_m$ слабо сходится к стандартному нормальному распределению при $m \to \infty$.

Отметим, что распределение Стьюдента симметрично.

Для построения доверительного интервала введем в рассмотрение функционал

$$G(\alpha, X) = \frac{\sqrt{n}(\overline{x} - \alpha)}{S_0},$$

где $S_0^2 = \frac{n}{n-1}S^2$ — несмещенная оценка для дисперсии. Докажем, что $G(\alpha, X) \in T_{n-1}$. Действительно,

$$G(\alpha, X) = \frac{\sqrt{n}(\overline{x} - \alpha)/\sigma}{S_0/\sigma}.$$

Согласно следствию 2 леммы Фишера числитель и знаменатель последней дроби являются независимыми случайными величинами. Окончательно,

$$G(\alpha, X) = \frac{\xi_0}{\sqrt{\frac{nS^2}{(n-1)\sigma^2}}} \stackrel{d}{=} \frac{\xi_0}{\sqrt{\frac{1}{n-1}\eta_{n-1}}} \in T_{n-1}.$$

Пусть $\tau_\varepsilon$ — квантиль уровня $1 - \varepsilon/2$ (при $\varepsilon \in (0,1)$) распределения $T_{n-1}$. Тогда границы доверительного интервала для неизвестного параметра $\alpha$ выглядят следующим образом:

$$\alpha_n^- = \overline{x} - \frac{S_0\tau_\varepsilon}{\sqrt{n}},$$

$$\alpha_n^+ = \overline{x} + \frac{S_0 \tau_\varepsilon}{\sqrt{n}}.$$

## Асимптотические доверительные интервалы.

Если распределение выборки не является нормальным, то вышеописанная конструкция доверительных интервалов не работает. Однако при значительных объемах наблюдений можно использовать асимптотический подход, который по сути сводит рассматриваемую задачу к нормальным выборкам.

**Определение.** Упорядоченная пара статистик $(\theta_n^-, \theta_n^+)$ называется *асимптотическим доверительным интервалом уровня доверия* $1-\varepsilon$, если

$$\lim_{n\to\infty} P(\theta_n^- < \theta < \theta_n^+) \geqslant 1 - \varepsilon.$$

## Построение асимптотических доверительных интервалов с помощью АНО.

**Теорема.** *Пусть $\theta_n^+$ — произвольная асимптотически нормальная оценка для параметра $\theta$, т. е. $\sqrt{n}(\theta_n^+ - \theta) \Rightarrow \eta \in N_{(0,\sigma)}$, где коэффициент рассеивания $\sigma(\theta)$ непрерывен. Тогда асимптотические доверительные границы определяются по формулам*

$$\theta_n^\pm = \theta_n^* \pm \frac{t_\varepsilon \sigma(\theta_n^*)}{\sqrt{n}},$$

*где $t_\varepsilon$ — квантиль уровня $1-\varepsilon/2$ для стандартного нормального распределения, т. е.*

$$\Phi(-t_\varepsilon) = \varepsilon/2.$$

Доказательство. Поскольку любая асимптотически нормальная оценка состоятельна и функция $\sigma(\theta)$ непрерывна, то

$$\frac{\sigma(\theta)}{\sigma(\theta_n^*)} \underset{p}{\to} 1, \qquad \frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma(\theta)} \Rightarrow \eta \in N_{(0,1)}.$$

Вспомним лемму из курса теории вероятностей: если $\xi_n \Rightarrow \xi,\ \eta_n \underset{p}{\to} c$, то $\xi_n\eta_n \Rightarrow c\xi$. Используя это утверждение, мы получаем

$$\frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma(\theta_n^*)} = \frac{\sigma(\theta)}{\sigma(\theta_n^*)} \frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma(\theta)} \Rightarrow \eta \in N_{(0,1)}.$$

Построение асимптотических доверительных интервалов на этом заканчивается:

$$\lim_{n\to\infty} P\left(-t_\varepsilon < \frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma(\theta_n^*)} < t_\varepsilon\right) = \Phi(t_\varepsilon) - \Phi(-t_\varepsilon) = 1 - 2\Phi(-t_\varepsilon) = 1 - \varepsilon$$

(используя свойство $\Phi(x) = 1 - \Phi(-x)$). $\square$

**Пример.** Пусть $X \in \{\pi_\lambda\}$, $\theta_n^* = \overline{x}$ — асимптотически нормальная оценка, $\sigma(\lambda) = \sqrt{\lambda}$. Тогда при больших $n$ доверительные границы вычисляются по формулам $\lambda_n^\pm = \overline{x} \pm \dfrac{t_\varepsilon\sqrt{\overline{x}}}{\sqrt{n}}$.

**Упражнение.** *Несимметричная монетка бросается* 10000 *раз. Построить асимптотические доверительные интервалы для параметра* $p$ *бернуллиевского распределения* $B_p$.

## Задачи проверки статистических гипотез.

**Определение.** *Гипотезой* $H_k$ будем называть любое суждение о неизвестном распределении. Гипотеза называется *простой*, если она однозначно восстанавливает неизвестное распределение: $H_k = \{F_\theta = F_{\theta_0}\}$ или, в терминах параметров, $H_k = \{\theta = \theta_0\}$.

В противном случае гипотеза $H_k$ называется *сложной*.

**Пример 1.** Наблюдаемое распределение имеет стандартный нормальный закон, т. е. $\theta_0$ есть вектор $(0, 1)$ — двумерный параметр. Это простая гипотеза.

**Пример 2.** Наблюдаемое распределение пуассоновское $H_k = \{F \in \{\pi_\lambda\}\}$ — сложная гипотеза.

Будем рассматривать случай, когда проверка гипотез сводится к конечному числу гипотез (наиболее часто — двух). Проверка означает выбор наиболее правдоподобной гипотезы.

**Определение.** *Статистический критерий для проверки конечного числа гипотез* — измеримое отображение выборочного пространства в конечный отрезок натурального ряда:

$$\delta : \mathfrak{X}^n \to \{1, \ldots, m\}.$$

Если $\delta(X) = i$, то номер $i$ соответствует гипотезе $H_i$.

Что же определяет качество статистического критерия? Начнем с проверки простых гипотез, т. е. будем считать $H_1, \ldots, H_m$ простыми. Символом $P_i$ будем обозначать вероятность тех или иных событий на выборочном пространстве, если выборка распределена в соответствии с $i$-ой гипотезой.

*Ошибкой $i$-го рода* называется событие $\delta(X) \neq i$ при том, что гипотеза $H_i$ верна. *Вероятность ошибки $i$-го рода* есть величина

$$\alpha_i(\delta) = P_i(\delta(x) \neq i), \quad i = \overline{1, m}.$$

**Пример.** Пусть $m = 2$. Врач принимает пациента. Обозначим гипотезы: $H_1 = \{\text{здоров}\}$, $H_2 = \{\text{болен}\}$. Выборкой являются анализы. Пациент может быть здоров, но анализы покажут, что он болен — ошибка первого рода. Может быть наоборот — ошибка второго рода. Оба случая не очень хорошие, но нельзя сказать, что они симметричны.

Качество критерия характеризуется величинами ошибок $i$-го рода (желательно, чтобы все они были небольшими, но это не всегда реализуемо).

### Существование оптимального критерия в случае двух простых гипотез. Теорема Неймана - Пирсона.

Рассмотрим параметрическое семейство плотностей $\{f_\theta\}$ относительно некоторой $\sigma$-конечной меры. Проверяются две простые гипотезы $H_1 = \{\theta = \theta_1\}$ и $H_2 = \{\theta = \theta_2\}$, так что параметрическое семейство состоит из двух точек.

Введем в рассмотрении класс $K_\varepsilon = \{\delta \mid \alpha_1(\delta) \leqslant \varepsilon\}$ — класс всех статистических критериев, у которых вероятность ошибки первого рода не превосходит $\varepsilon$.

**Теорема** (Нейман — Пирсон) *В классе $K_\varepsilon$ существует критерий $\delta^*$ с минимальной вероятностью ошибки второго рода*:

$$\alpha_2(\delta^*) = \inf\{\alpha_2(\delta) \mid \delta \in K_\varepsilon\}.$$

*Оптимальный критерий определяется следующим образом*:

$$\delta^*(x) = \begin{cases} 1, & \frac{\Psi_X(2)}{\Psi_X(1)} \leqslant c_\varepsilon, \\ 2, & \text{иначе,} \end{cases}$$

*где* $\Psi_x(i)$ *— функция правдоподобия* $i$*-го распределения. Критический уровень* $c_\varepsilon$ *вычисляется из определения вероятности ошибки первого рода*:

$$P_1(\delta^*(X) = 2) = P_1\left(\frac{\Psi_X(2)}{\Psi_X(1)} > c_\varepsilon\right) = \varepsilon.$$

Предполагается, что для данного $\varepsilon$ такое $c_\varepsilon$ найдется, хотя далеко не всегда оно существует, скажем, когда рассматривается выборка из дискретного распределения. В этом случае проблема решается введением так называемого *рандомизированного* критерия.

Обычно одну из гипотез называют *основной*, а другую — *конкурирующей или альтернативной*. Оптимальный критерий $\delta^*$ традиционно называют *критерием отношения правдоподобия*, поскольку он построен с помощью статистики, задаваемой отношением функций правдоподобия.

ДОКАЗАТЕЛЬСТВО. Для любого $\delta \in K_\varepsilon$ пусть $S_1 \subset \mathfrak{X}^n$ — область приема первой (основной) гипотезы, $S_2 = \overline{S_1}$ — так называемая *критическая область* приема конкурирующей гипотезы. Для критерия $\delta^*$ соответствующие области приема гипотез обозначим через $S_1^*$ и $S_2^*$. Рассмотрим вероятность ошибки второго рода критерия $\delta$:

$$\alpha_2(\delta) = P_2(\delta = 1) = P_2(X \in S_1) =$$

[так как у вектора $X$ есть плотность распределения $\Psi_{\overrightarrow{z}}(2)$, то]

$$= \int\limits_{S_1} \Psi_{\overrightarrow{z}}(2)\lambda^n(d\overrightarrow{z}) = \alpha_2(\delta^*) + \int\limits_{S_1} \Psi_{\overrightarrow{z}}(2)\lambda^n(d\overrightarrow{z}) - \int\limits_{S_1^*} \Psi_{\overrightarrow{z}}(2)\lambda^n(d\overrightarrow{z}).$$

Обозначим

$$\Delta = \int\limits_{S_1} \Psi_{\overrightarrow{z}}(2)\lambda^n(d\overrightarrow{z}) - \int\limits_{S_1^*} \Psi_{\overrightarrow{z}}(2)\lambda^n(d\overrightarrow{z}).$$

Покажем, что $\Delta \geqslant 0$. В дальнейшем для кратности будем использовать сокращенную запись интегралов. Имеем

$$\int\limits_{S_1} \Psi_{\overrightarrow{z}}(2) = \int\limits_{S_1 \cap S_2^*} \Psi_{\overrightarrow{z}}(2) + \int\limits_{S_1 \cap \overline{S_2^*}} \Psi_{\overrightarrow{z}}(2) \geqslant$$

[$\overline{S_2^*} = S_1^*$ и на множестве $S_1 \cap S_2^*$ выполнено $\Psi_{\overrightarrow{z}}(2) \geqslant c_\varepsilon \Psi_{\overrightarrow{z}}(1)$ — подставляем и получаем оценку снизу]

$$\geqslant c_\varepsilon \int\limits_{S_1 \cap S_2^*} \Psi_{\overrightarrow{z}}(1) + \int\limits_{S_1 \cap S_1^*} \Psi_{\overrightarrow{z}}(2).$$

Для второго интеграла в определении $\Delta$ аналогично получаем оценку сверху:

$$\int\limits_{S_1^*} \Psi_{\overrightarrow{z}}(2) = \int\limits_{S_1^* \cap S_2} \Psi_{\overrightarrow{z}}(2) + \int\limits_{S_1^* \cap S_1} \Psi_{\overrightarrow{z}}(2) \leqslant c_\varepsilon \int\limits_{S_1^* \cap S_2} \Psi_{\overrightarrow{z}}(1) + \int\limits_{S_1^* \cap S_1} \Psi_{\overrightarrow{z}}(2).$$

Таким образом,

$$\Delta \geqslant c_\varepsilon \left( \int\limits_{S_1\cap S_2^*} \Psi_{\overrightarrow{z}}(1) - \int\limits_{S_1^*\cap S_2} \Psi_{\overrightarrow{z}}(1) \right)$$

$$= c_\varepsilon \left( \int\limits_{S_1\cap S_2^*} \Psi_{\overrightarrow{z}}(1) - \int\limits_{S_1^*\cap S_2} \Psi_{\overrightarrow{z}}(1) \pm \int\limits_{S_2\cap S_2^*} \Psi_{\overrightarrow{z}}(1) \right) = c_\varepsilon \int\limits_{S_2^*} \Psi_{\overrightarrow{z}}(1) - c_\varepsilon \int\limits_{S_2} \Psi_{\overrightarrow{z}}(1) =$$

$$= c_\varepsilon P_1(\delta^* = 2) - c_\varepsilon \alpha_1(\delta).$$

Следовательно, $\Delta \geqslant c_\varepsilon(\varepsilon - \alpha_1(\delta))$. Так как $\alpha_1(\delta) \leqslant \varepsilon$, то $\Delta \geqslant 0$, что и требовалось доказать. □

## Байесовский подход для проверки конечного числа простых гипотез.

Мы уже отметили, что качество статистического критерия для проверки конечного числа простых гипотез определяется малостью вероятностей $i$-го рода

$$\alpha_i(\delta) = P_i(\delta(x) \neq i), \quad i = \overline{1, m}.$$

Как правило, попытки уменьшить одну или несколько из указанных вероятностей приводит к возрастанию оставшихся. Случай двух простых гипотез, описанный в теореме Неймана – Пирсона, стоит особняком.

*Байесовский подход* предлагает определять достоинство того или иного критерия не малостью вероятностей ошибок $i$-го рода в отдельности, а с помощью так называемой *взвешенной вероятности ошибки*:

$$\alpha_Q(\delta) = \sum_{i=1}^{m} \alpha_i(\delta) q_i.$$

Здесь без ограничения общности предполагается, что набор положительных весов $\{q_i;\ i = 1, ..., m\}$ является распределением, т. е. что $\sum q_i = 1$. Оно называется *априорным распределением гипотез*.

Оказывается, что при таком подходе существует оптимальный критерий, называемый *байесовским*, который минимизирует взвешенную вероятность ошибки при любом наперед заданном априорном распределении гипотез.

**Теорема** (Байес). *Для любого априорного распределения $Q = \{q_i;\ i = 1, ..., m\}$ гипотез байесовский критерий определяется по формуле*

$$\delta_Q(X) = i, \quad \text{\textit{если}} \quad q(i|X) = \max_{j \leqslant m} q(j|X),$$

*где*

$$q(j|X) = \frac{\Psi_X(j) q_j}{\sum_{k \leqslant m} \Psi_X(k) q_k}.$$

**Замечание**. Набор вероятностей $\{q(i|X);\ i = 1, ..., m\}$ называется *апостериорным распределением гипотез*. Так что смысл приведенного утверждения достаточно прозрачен – мы выбираем наиболее правдоподобную гипотезу с позиции апостериорного распределения. При этом экстремальных точек у апостериорной плотности $q(i|X)$ может быть несколько. В этом случае приведенный выше критерий выбирает любую из них.

Доказательство.Сначала мы получим нижнюю оценку для взвешенной вероятности ошибки произвольного критерия $\delta$. Обозначим через $S_i$ область приема $i$-ой гипотезы критерием $\delta$. Ясно, что $\{S_i\}$ образуют конечное разбиение выборочного пространства $\mathfrak{X}^n$. Имеем

$$\alpha_Q(\delta) = \sum_{i=1}^{m}(1 - P_i(\delta = i))q_i = 1 - \sum_{i=1}^{m}\int\limits_{S_i}\Psi_{\overrightarrow{z}}(i)q_i\lambda^n(d\overrightarrow{z}) \geqslant 1 - \sum_{i=1}^{m}\int\limits_{S_i}\max_{j\leqslant m}(\Psi_{\overrightarrow{z}}(j)q_j)\lambda^n(d\overrightarrow{z})$$

$$= 1 - \int\limits_{\mathfrak{X}^n}\max_{j\leqslant m}(\Psi_{\overrightarrow{z}}(j)q_j)\lambda^n(d\overrightarrow{z}) = 1 - \int\limits_{\mathfrak{X}^n}\max_{j\leqslant m} q(j|\overrightarrow{z})\tilde{\Psi}(\overrightarrow{z})\lambda^n(d\overrightarrow{z}),$$

где

$$\tilde{\Psi}(\overrightarrow{z}) = \sum_{k\leqslant m}\Psi_{\overrightarrow{z}}(k)q_k.$$

Отметим, во-первых, что правая часть приведенного неравенства *не зависит ни от какого критерия*. Во-вторых, это неравенство превратится в равенство, если в качестве $S_i$ взять область приема $i$-ой гипотезы для критерия, определенного в теореме. □

## Проверка простой гипотезы против сложной альтернативы.

Пусть $H_1 = \{F = F_0\}$ – простая гипотеза, $H_2 = \overline{H_1} = \{F \neq F_0\}$ – классическая альтернатива к $H_1$ (т. е. все распределения, отличающиеся от $F_0$).

В параметрическом случае проверки простой гипотезы против сложной альтернативы корректно определена вероятность ошибки первого рода. Аналогом вероятности ошибки второго рода будет функция от параметра $\theta$, задаваемая соотношением $\alpha_\theta(\delta) = \mathbf{P}_\theta(\delta \neq \theta)$. Здесь мы как-будто проверяем две простые гипотезы: значение неизвестного параметра принимает только два возможных значения $\theta_0$ (основная гипотеза) и $\theta$ (конкурирующая гипотеза).

**Определение**. Функция $\beta_\delta(\theta) = 1-\alpha_\theta(\delta)$ называется *мощностью* критерия $\delta$ для альтернативного значения параметра $\theta$. Если в некотором классе критериев существует критерий $\delta^*$, имеющий максимальную мощность при всевозможных значениях параметра, то $\delta^*$ называется *равномерно наиболее мощным* в указанном классе.

Рассмотрим один специальный случай, когда в параметрической постановке основная гипотеза имеет вид $H_1 = \{\theta = \theta_0\}$, а конкурирующая – $H_2 = \{\theta > \theta_0\}$или $H_2 = \{\theta < \theta_0\}$. В этом случае конкурирующие гипотезы называются *односторонними альтернативами*.

**Теорема** (существование равномерно наиболее мощного критерия). *Пусть $S$ – некоторая достаточная статистика для параметрического семейства плотностей $\{f_\theta\}$ со скалярным параметром $\theta$. При этом первая компонента факторизации Неймана – Фишера имеет вид*

$$\varphi(S,\theta) = \exp\{A_n(\theta)S + B_n(\theta)\},$$

*где функция $A_n(\theta)$ монотонна. Тогда существует равномерно наиболее мощный критерий для проверки простой гипотезы против односторонней альтернативы. Например, если альтернатива имеет вид $H_2 = \{\theta > \theta_0\}$ и функция $A_n(\theta)$ возрастает, то критическая область этого критерия задается соотношением $S > c_\varepsilon$, где критический уровень $c_\varepsilon$ определяется из уравнения*

$$\mathbf{P}_1(S > c_\varepsilon) = \varepsilon.$$

Доказательство. Ограничимся случаем, описанным в формулировке теоремы. Другие возможные комбинации односторонних альтернатив и знака монотонности функции $A_n(\theta)$ рассматриваются аналогично. Если мы покажем, что при любом фиксированном $\theta_1 > \theta_0$ наиболее мощный критерий для проверки двух простых гипотез $H_1 = \{\theta = \theta_0\}$ и $H_2 = \{\theta = \theta_1\}$ не зависит от $\theta_1$, т. е. будет универсальным, то утверждение будет доказано. В самом деле, по теореме Неймана – Пирсона критическая область для проверки двух упомянутых простых гипотез в условиях нашей теоремы выглядит как

$$\frac{\Psi_X(\theta_1)}{\Psi_X(\theta_0)} = \exp\{(A_n(\theta_1) - A_n(\theta_0))S + (B_n(\theta_1) - B_n(\theta_0))\} > \tilde{c}_\varepsilon$$

или в эквивалентной форме: $S > c_\varepsilon$, где $c_\varepsilon = (A_n(\theta_1) - A_n(\theta_0))^{-1}(\log \tilde{c}_\varepsilon + B_n(\theta_0) - B_n(\theta_1))$. Ясно, что $c_\varepsilon$ зависит только от $\theta_0$ и $\varepsilon$. □

**Упражнение**. *Построить равномерно наиболее мощный критерий для проверки простой гипотезы $H_1 = \{\alpha = 0\}$ против односторонней альтернативы $H_2 = \{\alpha > 0\}$ для выборки из нормального распределения $N(\alpha, 1)$.*

**Упражнение**. *Построить равномерно наиболее мощный критерий для проверки простой гипотезы $H_1 = \{\lambda = 1\}$ против односторонней альтернативы $H_2 = \{\lambda > 1\}$ для выборки из пуассоновского распределения $\Pi_\lambda$.*

**Принцип минимального расстояния.** Пусть на множестве всех функций распределения $\{F\}$ задана функция двух переменных $d(F_1, F_2)$, называемая *статистическим расстоянием*. Для любых функций распределения $F_1$ и $F_2$ предполагается выполненным следующее:

**1.** $d(F_1, F_1) = 0$,

**2.** $d(F_1, F_2) \geqslant 0$.

Это минимальные требования (мы не требуем симметричности и выполнения неравенства треугольника). Напомним, что по теореме Гливенко – Кантелли эмпирическая функция распределения $F_n^*$ с вероятностью 1 стремится в равномерной топологии к истинной функции распределения $F_0$. Рассмотрим поведение функционала $d(F_n^*, F_0)$. Если он обладает свойством непрерывности относительно указанной равномерной топологии, то $d(F_n^*, F_0) \to 0$ с вероятностью 1 при $n \to \infty$. Если это не так, то $F_0$ определенно не является истинным распределением. Мы пришли к конструкции "асимптотического" критерия (которую, впрочем, на практике реализовать невозможно):

$$\delta(X) = \begin{cases} 1, & d(F_n^*, F_0) \to 0, \\ 2, & d(F_n^*, F_0) \nrightarrow 0. \end{cases}$$

Это и есть описание "на пальцах" так называемого *принципа минимального расстояния*: если мы угадали истинное распределение (или выбрали близкое к нему), то статистическое расстояние должно быть мало́. Критерии, основанные на таких расстояниях, называются *критериями согласия* (основной гипотезы с экспериментальными данными). Но для того чтобы сделать такой критерий конструктивным, нужно формализовать понятие "мало́".

**Критерий Колмогорова.** Рассмотрим равномерную метрику в пространстве функций распределения $d_k(F_1, F_2) = \sup\limits_{t \in \mathbb{R}} |F_1(t) - F_2(t)|$. Этот функционал является не только статистическим, но и метрическим расстоянием.

**Лемма 1.** *Если истинное распределение $F_0$ непрерывно, то выполняется свойство непараметричности критерия, т. е. функция распределения $P(d_k(F_n^*, F_0) < t)$ не зависит от $F_0$.*

Доказательство. В качестве упражнения. *Указание*: с помощью квантильных преобразований сделать замену переменных под знаком супремума.

Это замечательное свойство позволяет применять универсальные таблицы для вычисления так называемых *критических уровней* $c_\varepsilon$ — разделительных барьеров между упомянутыми выше понятиями "мало́"и "не мало́". В результате получим *критерий Колмогорова*

$$\delta(x) = \begin{cases} 1, & d(F_n^*, F_0) \leqslant c_\varepsilon, \\ 2, & d(F_n^*, F_0) > c_\varepsilon, \end{cases}$$

где $c_\varepsilon$ находится из уравнения $P_1(d(F_n^*, F_0) > c_\varepsilon) = \varepsilon$ (это есть вероятность ошибки первого рода) и соответствующих таблиц. Однако эти таблицы созданы только для $n \leq 100$. Если же объем выборки существенно больше, то для вычисления критических уровней используют следующее утверждение.

**Теорема** (А. Н. Колмогоров). *Если* $F_0 \in \mathbb{C}(\mathbb{R})$*, то имеет место следующее предельное соотношение:*

$$\lim_{n\to\infty} P(\sqrt{n} d_k(F_n^*, F_0) < t) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2t^2} \equiv K(t).$$

Функция $K(t)$ называется *функцией Колмогорова*. Она является функцией распределения из класса $\mathbb{C}^\infty$, но не аналитическая. Приведенный результат неверен для разрывных функций $F_0$.

Стало быть, при больших $n$ справедливо приближенное равенство

$$P(\sqrt{n} d_k(F_n^*, F_0) < t) \cong \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2t^2} = K(t),$$

которое и используется для вычисления критических уровней $c_\varepsilon$.

**Критерий** $\chi^2$**.** Речь пойдет о проверке двух гипотез в двух случаях, когда проверяется простая гипотеза против сложной альтернативы, а также сложная против сложной.

Здесь мы имеем дело с наблюдениями произвольной природы. Поэтому будем рассматривать расстояния вида $d(\mathcal{P}_1, \mathcal{P}_2)$ на *пространстве всех распределений*, которые заданы на всех измеримых подмножествах измеримого множества $(\mathfrak{X}, \mathcal{B})$.

Рассмотрим произвольное конечное разбиение $\{\Delta_i, i \leqslant m\}$ множества $\mathfrak{X}$. По этому разбиению определим так называемое расстояние $\chi^2$:

$$d_{\chi^2}(\mathcal{P}_1, \mathcal{P}_2) = \sum_{i=1}^{m} \frac{(\mathcal{P}_1(\Delta_i) - \mathcal{P}_2(\Delta_i))^2}{\mathcal{P}_2(\Delta_i)}$$

(по умолчанию предполагается $\mathcal{P}_2(\Delta_i) \neq 0$), которое не является метрикой в отличии от расстояния Колмогорова, так как для него не выполнено неравенство треугольника и нет симметрии относительно перестановки аргументов. При этом выполнены упомянутые выше свойства статистического расстояния (с очевидной заменой функций распределения на сами распределения).

Следуя принципам построения критериев согласия, рассмотрим $d(P_n^*, \mathcal{P}_0)$, где $P_n^*$ — эмпирическое распределение, построенное по выборке $x_1, \ldots, x_n$. (для случая простой основной гипотезы $H_1 = \{\mathcal{P} = \mathcal{P}_0\}$ и альтернативы $H_2 = \overline{H_1}$).

Обозначим через $\nu_i = \#\{x_j \mid x_j \in \Delta_i\}$ число выборочных точек, попавших в множество $\Delta_i$; положим $p_i = \mathcal{P}_0(\Delta_i) \neq 0$. Тогда

$$d(P_n^*, \mathcal{P}_0) = \sum_{i=1}^{m} \frac{\left(\frac{\nu_i}{n} - p_i\right)^2}{p_i}.$$

Критерий строится следующим образом:

$$\delta(x) = \begin{cases} 1, & d(P_n^*, \mathcal{P}_0) \leqslant c_\varepsilon, \\ 2, & \text{иначе.} \end{cases}$$

Вероятность ошибки первого рода для этого критерия: $\varepsilon = P_1(\delta = 2) = P_1(d(P_n^*, \mathcal{P}_0) > c_\varepsilon)$. Тем самым мы получили уравнение для нахождения $c_\varepsilon$, где $\varepsilon$ называют *уровнем критерия*. По аналогии с критерием Колмогорова имеет место свойство *асимптотической непараметричности* критерия $\chi^2$ в классе всех распределений.

Для данного расстояния используем только асимптотический результат. Приведем без доказательства *теорему Пирсона*:

**Теорема**. *Если $\mathcal{P}_0$ — истинное распределение $x_1$, то при $n \to \infty$*

$$n d_{\chi^2}(P_n^*, \mathcal{P}_0) \Rightarrow \eta \in \chi^2_{m-1}.$$

Схема доказательства. После умножения на $n$ исследуем асимптотическое поведение отдельного слагаемого из представления критерия. В силу теоремы Муавра — Лапласа имеем

$$\frac{(\nu_i - np_i)^2}{np_i} = \left(\frac{\nu_i - np_i}{\sqrt{np_i}}\right)^2 = (1 - p_i)\left(\frac{\nu_i - np_i}{\sqrt{np_i(1-p_i)}}\right)^2 \Rightarrow (1 - p_i)\xi_i^2,$$

где $\xi_i \in N_{(0,1)}$; здесь мы учли, что $\nu_i = \sum\limits_{j=1}^{n} I(x_j \in \Delta_i)$ — сумма независимых бернуллиевских случайных величин с вероятностью успеха $p_i = P(x_1 \in \Delta_i) = \mathcal{P}_0(\Delta_i)$.

Трудность заключается в том, что упомянутые слагаемые в представлении статистики $\chi^2$ зависимы между собой (так как $\nu_1 + \ldots + \nu_m = n$), а стало быть, зависимы и определенные выше гауссовские случайные величины $\xi_i$. Отметим, что доказательство этого утверждения проводят с помощью *многомерной центральной предельной теоремы*. При этом предельная квадратичная форма будет иметь распределение $\chi^2$ с числом степеней свободы $m - 1$.

Следовательно, для нахождения критических точек мы должны использовать таблицы распределения $\chi^2$ с тем или иным числом степеней свободы.

**Замечание.** Следует иметь в виду, что это теорема предельная. Но может быть ситуация, когда некоторые величины $p_i$ близки к 0. Тогда при небольших $n$ в силу малости знаменателя мы можем существенно превзойти критический уровень (даже если основная гипотеза была верна!), что исказит наши выводы. В случае, если $\nu_i$ близки к 0, а $p_i$ значимы, имеем $\dfrac{(\nu_i - np_i)^2}{np_i} \approx np_i$ — тоже можем получить большой выброс, который выравнивается лишь при $n \to \infty$, если гипотеза была справедливой. Поэтому при небольших объемах наблюдений на выбор разбиения желательно накладывать следующие ограничения во избежание появления такого сорта особенностей:

**1.** $\nu_i \sim np_i$,

**2.** $np_i \geq 8 \div 10$.

Критерий $\chi^2$ также называют *критерием сгруппированных данных*.

**Обобщенная теорема Пирсона.** Рассматриваем проверку двух сложных гипотез, когда основная имеет вид $H_1 = \{\mathcal{P} \in \{F_\theta\}\}$, где $\{F_\theta\}$ — некоторый параметрический класс, а конкурирующая есть классическая альтернатива $H_2 = \overline{H_1}$. От параметрического класса требуются некоторые свойства регулярности $R^*$, близкие к условиям $(R)$ в неравенстве Рао - Крамера. Отметим, что параметрические семейства нормальных, пуассоновских и некоторых других распределений удовлетворяют условиям $R^*$. Отличие от предыдущей схемы состоит в том, что не понятно, как искать $p_i = p_i(\theta)$ при неизвестном параметре. Предлагается $\theta$ заменить на ОМП $\widehat{\theta}_n$.

**Теорема**. *При сделанных предположениях и $n \to \infty$*

$$\frac{\sum\limits_{i=1}^{m}(\nu_i - np_i(\widehat{\theta}_n))^2}{np_i(\widehat{\theta}_n)} \Rightarrow \eta \in \chi^2_{m-1-\dim(\theta)}.$$

**Пример.** Рассмотрим историческую задачу на применение критерия $\chi^2$. Как известно, во время второй мировой войны южную часть Лондона обстреливалась ракетами "ФАУ". К моменту, когда возникла необходимость статистического анализа, ситуация была следующей: в указанной части города было зафиксировано 535 попаданий реактивных снарядов. Немецкая пропаганда утверждала, что их ракеты могут поразить заранее выбранный объект размером с дом. Началась паника. Решили обратиться к статистикам с вопросом: можно ли по имеющейся информации говорить о детерминированном выборе цели? Иначе говоря, осуществляется ли на самом деле прицельное бомбометание или "точки" бросаются наудачу.

Пострадавшую часть Лондона разбили на 576 участков ($n = 576$) площадью в четверть квадратной мили каждый. Далее подсчитали, сколько "пробоин" было на каждом участке. Получили таблицу сгруппированных данных:

| i | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $\nu_i$ | 229 | 211 | 93 | 35 | 7 | 1 |

где $i$ — число "пробоин" в отдельно взятом квадрате, $\nu_i$ — число квадратов с $i$ попаданиями. Возникает вопрос о том, какую гипотезу нужно проверять. Если мы предполагаем, что бросание точек происходит наудачу (основная гипотеза), то в наперед заданный квадрат попадаем с вероятностью $p = \frac{1}{576}$, при этом суммарное число испытаний ("пробоин") равно $N = 535$. По теореме Пуассона о редких событиях распределение числа успехов будет идеально описываться распределением Пуассона с параметром $\lambda = Np \sim 0,9$. Как следствие основной гипотезы возникает новая: *наблюдаемая случайная величина имеет распределение Пуассона с неизвестным параметром*. ОМП для параметра $\lambda$ хорошо известна:

$$\widehat{\lambda}_n = \overline{x} = \frac{1}{576}\sum_{i=1}^{576} x_i = \frac{535}{576} \approx 0.9,$$

где $x_i$ — количество "точек" в $i$-ой группе. Принимая во внимание замечание, сделанное выше, объединим группы $i = 4$ и $i = 5$ в одну. По таблицам для пуассоновского распределения с параметром $0,9$ находим

$$p_0 = 0.41,$$
$$p_1 = 0.37,$$
$$p_2 = 0.16,$$
$$p_3 = 0.05,$$
$$p_4 = 0.01,$$
$$\dim(\theta) = 1.$$

Используя вышеприведенную таблицы сгруппированных данных, получаем

$$\sum_{i=1}^{m} \frac{(\nu_i - \widetilde{n} p_i(\widehat{\theta}_n))^2}{\widetilde{n} p_i(\widehat{\theta}_n)} \approx 2.5,$$

где $\widetilde{n} = 576$. Значение 2.5 — это $\chi^2$-расстояние между эмпирическим и гипотетическим распределениями. По таблицам для распределения $\chi_3^2$ при $\varepsilon = 0.05$ имеем $c_\varepsilon = 7.81$. Так как $2.5 < 7.81$, то основная гипотеза о пуассоновости принимается. Следовательно, принимается и гипотеза о равномерности бросания "точек".

# Двувыборочные критерии согласия.

Проводится $n$ испытаний в одних условиях, а $m$ — в других. Возникают две выборки $x_1 \dots x_n$ и $y_1 \dots y_m$ ($x_i$ и $y_i$ независимы в совокупности). Требуется определить, изменилось ли распределение наблюдаемых величин после изменения условий проведения испытаний.

**Критерии Стьюдента и Фишера**. Постановка задачи: пусть $x_i \in N_{(\alpha,\sigma^2)}$, $y_i \in N_{(\alpha_1,\sigma^2)}$, где $\alpha, \alpha_1$ — неизвестны. Критерий Стьюдента проверяет сложную гипотезу $H = \{\alpha = \alpha_1\}$ против конкурирующей альтернативы. Этот критерий основан на статистике

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{nS_x^2 + mS_y^2}} \sqrt{\frac{m+n-2}{\frac{1}{m} + \frac{1}{n}}},$$

которая имеет распределение Стьюдента с $n+m-2$ степенями свободы, где $S_\cdot^2$ — выборочная дисперсия, построенная по соответствующей выборке.

Поясним, почему распределение точное, а не предельное. По следствию из леммы Фишера $\overline{x}$ и $S_x^2$, а также $\overline{y}$ и $S_y^2$ — независимы (ведут себя как причинно несвязные, хотя строятся по одним выборкам). По построению эти две пары не зависят друг от друга. Стало быть, $\overline{x}, S_x^2, \overline{y}, S_y^2$ независимы в совокупности. Так как борелевские преобразования независимости не портят, то числитель дроби в выражении для $T$ не зависит от знаменателя. При этом при выполнении гипотезы $H$

$$\frac{\overline{x} - \overline{y}}{\sigma} \sqrt{\frac{nm}{n+m}} \in N_{(0,1)}, \quad \frac{nS_x^2 + mS_y^2}{\sigma^2} \in \chi^2_{m+n-2}.$$

Получаем вероятностную модель закона Стьюдента: $\xi_0 / \sqrt{\frac{1}{r}\chi_r^2}$, т. е. $T \in T_{n+m-2}$.

Для проверки простой гипотезы о равенстве дисперсий двух независимых нормальных выборок с одинаковыми средними используется критерий Фишера, основанный на статистике со структурой отношения $\dfrac{\xi_1^2 + \ldots + \xi_n^2}{\widetilde{\xi}_1^2 + \ldots + \widetilde{\xi}_m^2}$, где $\xi_i, \widetilde{\xi}_j \in N_{(0,1)}$ — независимы в совокупности. Эта случайная величина распределена по закону Фишера $F_{n,m}$ с параметрами $n, m$. В силу леммы Фишера и уже проделанных элементарных преобразований $\dfrac{nS_x^2}{mS_y^2} \in F_{n-1,m-1}$. Дальнейшее построение критерия согласия проходит по стандартной схеме.