Почти наверное достаточное учебное пособие

2.5-е издание

Теория вероятностей и математическая статистика

CMC MSU

@MandaloreUltimate

Данное учебное пособие составлено на основе экзаменационных билетов 2020 года по курсу «Теория вероятностей и математическая статистика» факультета ВМК МГУ (первый поток) и *не является* конспектом лекций, а также *не проверялось* преподавателями курса. При составлении преимущественно использовались материалы, указанные в разделе **Литература**.

Авторский коллектив:

- Главный редактор (теория вероятностей) — Рожков И.С.
- Главный редактор (мат. статистика), верстальщик — Рыгин А.С.
- Рецензент, главный редактор (2-е издание) — Селезнёв М.В.
- Иллюстратор графиков — Васильев Р.Л.

Последняя версия учебного пособия доступна по адресу: `https://mandaloreultimate.github.io/stat-book/main.pdf`.

Вы можете внести исправления в репозитории проекта: `https://github.com/MandaloreUltimate/stat-book`.

Авторы *запрещают* Российскому авторскому обществу и любым другим организациям производить любого рода лицензирование данного произведения и осуществлять в интересах авторов какую бы то ни было иную связанную с авторскими правами деятельность без их письменного разрешения.

Оглавление

Теория вероятностей

1.1 Вероятностное пространство. Операции над событиями. Свойства вероятности

Определение. *Пространство элементарных исходов* Ω — любое непустое множество, содержащее все возможные результаты случайного эксперимента. Элементы $\omega \in \Omega$ — *элементарные исходы.*

Определение. *Алгебра* $\mathcal{A}$ — множество подмножеств Ω, обладающее следующими свойствами:

1. $\Omega \in \mathcal{A}$;
2. $A \in \mathcal{A} \implies \overline{A} \in \mathcal{A}$; [1]
3. $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$ (по индукции: $A_1, A_2, \ldots, A_n \in \mathcal{A} \implies$ $\implies \bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$).

Замечание. Если $A, B \in \mathcal{A}$, то $A \cap B \equiv \overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$.

Определение. *σ-алгебра* $\mathcal{F}$ — множество подмножеств Ω, обладающее следующими свойствами:

1. $\Omega \in \mathcal{F}$;
2. $A \in \mathcal{F} \implies \overline{A} \in \mathcal{F}$;
3. $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \implies \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

Пример. Множество всех подмножеств 2^Ω и множество $\{\varnothing, \Omega\}$ — σ-алгебры над Ω.

Замечание. Любая σ-алгебра является алгеброй. Первые два пункта определений идентичны, рассмотрим третий. Для любой конечной последовательности $A_1, A_2, \ldots, A_n \in \mathcal{A}$ составим соответствующую счётную последовательность $A_1, A_2, \ldots, A_n, A_{n+1} = \varnothing, A_{n+2} = \varnothing, \ldots \in \mathcal{A}$. Пустое множество $\varnothing$ принадлежит σ-алгебре (т.к. $\varnothing = \overline{\Omega}$). По определению σ-алгебры: $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F} \implies \bigcup\limits_{i=1}^{n} A_i \in \mathcal{F}$, следовательно, выполнен третий пункт определения алгебры.

[1] Здесь и в дальнейшем $\overline{A} \equiv \Omega \setminus A$ — *дополнение* к A.

Определение. *Случайное событие* A — элемент σ-алгебры $\mathcal{F}$, т.е. некоторое подмножество элементарных исходов. $A = \varnothing$ —*невозможное событие*, $A = \Omega$ — *достоверное событие*. Событие $\overline{A}$ — *противоположное* A, т.е. происходит тогда и только тогда, когда не происходит A.

Операции над событиями:

- *Объединение* $A \cup B$ — происходит или A, или B, или оба вместе.
- *Пересечение* $A \cap B$ (или AB) — происходят и A и B вместе. Если $AB = \varnothing$, то события A и B называются *несовместными*.
- *Разность* $A \setminus B$ — происходит A и не происходит B.
- *Симметрическая разность* $A \triangle B$ — либо происходит A и не происходит B, либо происходит B и не происходит A.

Определение. σ-алгебра *порождена классом* K, если она является пересечением всех σ-алгебр, содержащих K, т.е. является *минимальной σ-алгеброй*, содержащей K.

Пример. Пусть $K = \{A\}$, тогда $\sigma(K) = \{\varnothing, A, \overline{A}, \Omega\}$.

Определение. *Вероятностная мера* или *вероятность* — функция $\mathbb{P}\colon \mathcal{F} \mapsto \mathbb{R}$, обладающая следующими свойствами:

1. $\mathbb{P}(A) \geqslant 0 \quad \forall A \in \mathcal{F}$ (*неотрицательность*);
2. $\mathbb{P}(\Omega) = 1$ (*нормировка*);
3. $\forall A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}\ A_iA_j = \varnothing\ (i \neq j)\colon\ \mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mathbb{P}(A_i)$ (*счётная аддитивность*).

Замечание. Из счётной аддитивности, очевидно, следует и конечная аддитивность, достаточно рассмотреть последовательность событий $A_1, A_2, \ldots, A_n, A_{n+1} = \varnothing, A_{n+2} = \varnothing, \ldots \in \mathcal{F}$.

Свойства вероятности.

1. *$\mathbb{P}(\varnothing) = 0$;*
2. *$A, B \in \mathcal{F},\ B \subset A \implies \mathbb{P}(A) \geqslant \mathbb{P}(B)$ (монотонность);*
3. *$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(AB)$;*
4. *$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$;*
5. *$\forall A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots,\ \bigcap\limits_{n=1}^{\infty} A_n = A\colon\ \lim\limits_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ (непрерывность).*

Доказательство.

1. Рассмотрим последовательность событий $A_1 = \Omega, A_2 = \varnothing, \ldots,$ $A_n = \varnothing, \ldots$:

$$\bigcup_{i=1}^{\infty} A_i = \Omega \implies \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}(\Omega) = 1.$$

   При этом $A_i A_j = \varnothing$ $(i \neq j)$, следовательно, по пункту 3 определения вероятности: $\sum\limits_{i=2}^{\infty} \mathbb{P}(\varnothing) = 0 \implies \mathbb{P}(\varnothing) = 0$.

2. $B \subset A \implies A = (A \setminus B) \cup B$. Из неотрицательности вероятности и того, что $(A \setminus B) \cap B = \varnothing$, следует, что $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(B) \geqslant \mathbb{P}(B)$. Кроме того, в этом случае $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$.

3. Доказательство аналогично пункту 2 при представлении A в виде $A = (A \setminus B) \cup AB$.

4. Представим объединение событий A и B в виде $A \cup B = (A \setminus AB) \cup B$. Очевидно, что $(A \setminus AB) \cap B = \varnothing$, откуда по пункту 3 определения вероятности следует:

$$\mathbb{P}(A \cup B) = \mathbb{P}((A \setminus AB) \cup B) = \mathbb{P}(A \setminus AB) + \mathbb{P}(B),$$
$$\mathbb{P}(A \setminus AB) = \mathbb{P}(A) - \mathbb{P}(A \cap AB) = \mathbb{P}(A) - \mathbb{P}(AB).$$

   Следовательно, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$.

5. Рассмотрим множества $C_n = A_n \setminus A_{n+1}$, $n \in \mathbb{N}$. Они несовместны (пусть $l < m$, тогда $C_m = (A_m \setminus A_{m+1}) \subset A_m \subset A_{m-1} \subset \ldots \subset A_{l+1}$, но $C_l = A_l \setminus A_{l+1} \implies C_l \cap C_m = \varnothing$). Тогда

$$A_1 = \bigcup_{n=1}^{\infty} C_n \cup A, \quad \mathbb{P}(A_1) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} C_n \cup A\right) = \sum_{n=1}^{\infty} \mathbb{P}(C_n) + \mathbb{P}(A),$$
$$\sum_{n=1}^{\infty} \mathbb{P}(C_n) = \mathbb{P}(A_1) - \mathbb{P}(A).$$

   Таким образом, ряд из вероятностей событий C_n сходится. Это равносильно тому, что его остаток $\sum\limits_{n=k}^{\infty} \mathbb{P}(C_n)$ стремится к нулю при $k \to \infty$.

Но при этом

$$\bigcup_{n=k}^{\infty} C_n \cup A = A_k, \quad \mathbb{P}(A_k) = \mathbb{P}\left(\bigcup_{n=k}^{\infty} C_n \cup A\right) = \sum_{n=k}^{\infty} \mathbb{P}(C_n) + \mathbb{P}(A).$$

Перейдя в последнем равенстве к пределу при $k \to \infty$, получим

$$\lim_{k\to\infty} \mathbb{P}(A_k) = \mathbb{P}(A).$$

$\blacksquare$

Замечание. Можно показать, что счётная аддитивность равносильна одновременному наличию конечной аддитивности и непрерывности. Иными словами, третью аксиому в определении вероятности можно заменить на пару утверждений:

- $\forall A, B,\ A \cap B = \varnothing\colon\ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (по индукции можно показать аддитивность для любого конечного n);
- $\forall A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots;\ \bigcap\limits_{n=1}^{\infty} A_n = A\colon\ \lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Определение. *Вероятностное пространство* — тройка $(\Omega, \mathcal{F}\, \mathbb{P})$, где Ω — множество элементарных исходов, $\mathcal{F}$ — σ-алгебра над Ω, вероятность $\mathbb{P}$ определена на $\mathcal{F}$.

Замечание. Вероятностное пространство не является пространством в функциональном смысле.

Пример. Тройка $([0,1], \mathfrak{B}_{[0,1]}, \lambda_{[0,1]})$, где λ — мера Лебега, $\mathfrak{B}$ — борелевская σ-алгебра, является вероятностным пространством. В самом деле:

- $\Omega = [0,1]$ — непустое множество;
- $\mathfrak{B}_{[0,1]}$ — σ-алгебра над $\Omega = [0,1]$;
- $\lambda(\Omega) = \lambda\big([0,1]\big) = 1,\ \forall A \in \mathfrak{B}_{[0,1]}\ \lambda(A) \geqslant 0$ и мера счётного объединения непересекающихся множеств есть счётная сумма их мер, т.е. выполняются три аксиомы вероятности.

В то же время тройка $([-1,1], \mathfrak{B}_{[-1,1]}, \lambda_{[-1,1]})$ не будет вероятностым пространством, т.к. нарушается свойство нормировки вероятностной меры: $\lambda(\Omega) = 2$, следовательно, вероятность не задана.

Определение. Пусть $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ — конечное непустое множество, $\mathcal{F}$ — множество всех подмножеств Ω. Положим $\mathbb{P}\big(\{\omega_i\}\big) = p_i$. Вероятностное пространство, определённое таким образом, — *дискретное вероятностное* *пространство.* При этом для любого события $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$ его вероятность $\mathbb{P}(A) = \sum\limits_{j=1}^{k} p_{i_j}$. Таким образом вводится *классическое определение вероятности* на дискретном вероятностном пространстве:

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n}, \quad \mathbb{P}(A) = \frac{k}{n},\ k = |A|,$$

т.е. все элементарные исходы считаются *равновозможными.*

1.2 Условная вероятность. Независимость событий. Критерий независимости. Формула полной вероятности. Формула Байеса

Определение. Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, события $A, B \in \mathcal{F}$, $\mathbb{P}(B) > 0$. *Условная вероятность события A при событии B*:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

Утверждение. *Условная вероятность $\mathbb{P}(A|B)$ — вероятность, заданная на σ-алгебре $\mathcal{F}$.*

Доказательство. Проверим три аксиомы из определения вероятности.

1. $\forall A \in \mathcal{F} \quad \mathbb{P}(A|B) \geqslant 0, \text{т.к. } \mathbb{P}(AB) \geqslant 0,\ \mathbb{P}(B) > 0$;
2. $\mathbb{P}(\Omega|B) = \dfrac{\mathbb{P}(B \cap \Omega)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$;
3. Пусть дана некоторая последовательность событий $A_1, A_2, \ldots A_n, \ldots$; $A_i A_j = \varnothing$ $(i \neq j)$. Тогда:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{\mathbb{P}\left(\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)} =$$

$$= \frac{\sum\limits_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B).$$

■

Замечание. Некоторые свойства условной вероятности:

1. Если $A \cap B = \varnothing$, то $\mathbb{P}(A|B) = 0$.

2. Если $B \subset A$, то $\mathbb{P}(A|B) = 1$. Например, $\mathbb{P}(B|B) = 1$.

Независимость событий

Определение. Пусть есть вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$. События $A_1, \ldots, A_n \in \mathcal{F}$ называются *независимыми в совокупности*, если $\forall k = \overline{2, n}$ $\forall i_1, \ldots, i_k \colon 1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ выполняется:

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Иными словами, события независимы в совокупности, если вероятность одновременного наступления *любого набора* из этих событий равна *произведению вероятностей* событий, входящих в этот набор. В частности, при $n = 2$ события A и B независимы, если $\mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Свойства независимых событий.

1. *Если $A = \varnothing$ или $\mathbb{P}(A) = 0$, то $\forall B\colon\ \mathbb{P}(B) > 0$ события A и B независимы.*

2. *Пусть A и B независимы. Тогда события $\overline{A}$ и B, A и $\overline{B}$, $\overline{A}$ и $\overline{B}$ также независимы.*

3. *Пусть $A \subset B$ и $\mathbb{P}(A) > 0$, $\mathbb{P}(B) < 1$. Тогда A и B зависимы.*

4. *Если события A и B независимы и $\mathbb{P}(B) > 0$, то $\mathbb{P}(A|B) = \mathbb{P}(A)$.*

Доказательство.

1. Если $A = \varnothing$, то $AB = \varnothing \implies \mathbb{P}(AB) = 0$. Но $\mathbb{P}(A)\,\mathbb{P}(B) = 0 \cdot \mathbb{P}(B) = 0 \implies \mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B)$.

   Если же $\mathbb{P}(A) = 0$, то $AB \subset A \implies \mathbb{P}(AB) \leqslant \mathbb{P}(A) = 0$. В то же время $0 = \mathbb{P}(AB) = 0 \cdot \mathbb{P}(B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

2. Докажем независимость $\overline{A}$ и B, представив последнее в виде $B = AB \cup \overline{A}B$. Тогда

$$\mathbb{P}(B) = \mathbb{P}(AB) + \mathbb{P}(\overline{A)B} = \mathbb{P}(A)\,\mathbb{P}(B) + \mathbb{P}(\overline{A}B) \implies$$
$$\implies \mathbb{P}(\overline{A}B) = \mathbb{P}(B) - \mathbb{P}(A)\,\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(\overline{A)}\,\mathbb{P}(B).$$

Независимость $\overline{A}$ и B доказана. Аналогично доказываются остальные утверждения.

3. Предположим, что события независимы. Тогда $\mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B)$, но в силу вложенности $A \subset B$: $\mathbb{P}(AB) = \mathbb{P}(A) > 0$, следовательно, $\mathbb{P}(B) = 1$, что противоречит условию.

4. $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(A)\,\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$

■

Замечание. В общем случае из попарной независимости событий $A_1, \ldots, A_n$ не следует их независимость в совокупности.

Пример. Рассмотрим правильный тетраэдр, три грани которого окрашены соответственно в красный, синий, зелёный цвета, а четвёртая грань содержит все три цвета. Событие R (соответственно, G, B) означает, что выпала грань, содержащая красный (соответственно, зелёный, синий) цвета.

Т.к. каждый цвет есть на двух гранях из четырёх, то

$$\mathbb{P}(R) = \mathbb{P}(G) = \mathbb{P}(B) = \frac{1}{2}.$$

Вероятность пересечения, соответственно:

$$\mathbb{P}(RG) = \mathbb{P}(GB) = \mathbb{P}(RB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2},$$

следовательно, все события попарно независимы. Однако вероятность пересечения всех трёх:

$$\mathbb{P}(RGB) = \frac{1}{4} \neq \mathbb{P}(R)\,\mathbb{P}(G)\,\mathbb{P}(B),$$

т.е. события не являются независимыми в совокупности.

Обозначение.

$$A_i^{(\delta)} = \begin{cases} A_i, & \delta = 1; \\ \overline{A_i}, & \delta = 0. \end{cases}$$

Критерий независимости. *События $A_1, \ldots, A_n$ независимы в совокупности $\iff \forall \delta_1, \delta_2, \ldots \delta_n \in \{0, 1\}$ выполнено равенство*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i^{(\delta_i)}\right) = \prod_{i=1}^{n} \mathbb{P}\left(A_i^{(\delta_i)}\right).$$

Формула полной вероятности. *Пусть даны события* $A, B_1, \ldots, B_n, \ldots$; $\mathbb{P}(B_i) > 0$, *причём* $B_i B_j = \varnothing$ $(i \neq j)$ *и* $\bigcup\limits_{i=1}^{\infty} B_i \supset A$ *(например,* $\bigcup\limits_{i=1}^{\infty} B_i = \Omega$*). Тогда справедлива формула:*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}\left(B_i\right) \mathbb{P}\left(A|B_i\right).$$

Доказательство. Достаточно заметить, что при вышеперечисленных условиях $A = \bigcup\limits_{i=1}^{\infty}(AB_i)$, и $AB_i \cap AB_j = \varnothing$ $(i \neq j)$. Тогда, учитывая $\mathbb{P}(B_i) > 0$, получаем

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}\left(AB_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(B_i\right) \frac{\mathbb{P}\left(AB_i\right)}{\mathbb{P}\left(B_i\right)} = \sum_{i=1}^{\infty} \mathbb{P}\left(B_i\right) \mathbb{P}\left(A|B_i\right).$$

■

Формулы Байеса. *Пусть даны события* $A, H_1, \ldots, H_n, \ldots$; $\mathbb{P}(A) > 0$, $\mathbb{P}(H_i) > 0$, *причём* $H_i H_j = \varnothing$ $(i \neq j)$ *и* $\bigcup\limits_{i=1}^{\infty} H_i \supset A$ *(например,* $\bigcup\limits_{i=1}^{\infty} H_i = \Omega$*). Тогда справедливы формулы Байеса:*

$$\mathbb{P}\left(H_i|A\right) = \frac{\mathbb{P}\left(H_i\right) \mathbb{P}\left(A|H_i\right)}{\sum\limits_{j=1}^{\infty} \mathbb{P}\left(H_j\right) \mathbb{P}\left(A|H_j\right)}, \quad i = \overline{1, n}.$$

Доказательство. Согласно формуле полной вероятности, в знаменателе дроби стоит вероятность события A. Тогда

$$\frac{\mathbb{P}\left(H_i\right) \mathbb{P}\left(A|H_i\right)}{\mathbb{P}(A)} = \frac{\mathbb{P}\left(H_i\right) \mathbb{P}\left(AH_i\right)}{\mathbb{P}(A)\, \mathbb{P}\left(H_i\right)} = \frac{\mathbb{P}\left(AH_i\right)}{\mathbb{P}(A)} = \mathbb{P}\left(H_i|A\right).$$

■

Вероятности $P(H_i)$, вычисленные заранее, до проведения эксперимента, называют *априорными вероятностями* (a'priori — «до опыта»). Условные вероятности $\mathbb{P}(H_i|A)$ называют *апостериорными вероятностями* (a'posteriori — «после опыта»). Формула Байеса позволяет переоценить заранее известные вероятности после того, как получено знание о результате эксперимента.

Пример. Тест на рак имеет надёжность 99% (т.е. вероятность как положительной, так и отрицательной ошибки равна 0,01), рак появляется у 1% населения. Какова вероятность того, что человек болен раком, если у него позитивный результат теста?

Составим таблицу для вероятностей всех возможных событий:

Результат теста	Пациент реально болен	
	Да	Нет
Положительный	$0{,}99 \cdot 0{,}01$	$0{,}01 \cdot 0{,}99$
Отрицательный	$0{,}01 \cdot 0{,}01$	$0{,}99 \cdot 0{,}99$

Введём следующие обозначения для событий: $H_+ = \{\text{пациент болен}\}$, $H_- = \{\text{пациент здоров}\}$, $R_+ = \{\text{положительный результат теста}\}$, $R_- = \{\text{отрицательный результат теста}\}$. Найдём вероятность события H_+ при условии R_+ по формуле Байеса:

$$\mathbb{P}(H_+|R_+) = \frac{\mathbb{P}(H_+)\,\mathbb{P}(R_+|H_+)}{\mathbb{P}(H_+)\,\mathbb{P}(R_+|H_+) + \mathbb{P}(H_-)\,\mathbb{P}(R_+|H_-)} = \\ = \frac{0{,}99 \cdot 0{,}01}{(0{,}99 \cdot 0{,}01) + (0{,}01 \cdot 0{,}99)} = 0{,}5$$

Иными словами, вероятность того, что пациент болен, равна отношению вероятности правильного положительного результата теста к вероятности любого положительного результата.

Рассмотрим более общий случай. Пусть q — вероятность неправильного результата теста, p — вероятность заболеть раком, тогда

$$\mathbb{P}(H_+|R_+) = \frac{(1-q)p}{(1-q)p + q(1-p)} = \frac{p - qp}{p + q - 2qp}$$

Эта функция принимает значение 0,5 на диагонали $p = q$; ниже диагонали — вероятность выше 0,5, т.е. чтобы верить результатам теста, вероятность болезни должна превышать вероятность его ошибки.

1.3 Случайная величина. Порождённое и индуцированное вероятностные пространства. Функция распределения, ее свойства

Определение. *Борелевская σ-алгебра* $\mathfrak{B}$ — σ-алгебра, порождённая множеством всех открытых интервалов на $\mathbb{R}$ (иными словами, минимальная σ-алгебра, содержащая все открытые интервалы). Элемент $B \in \mathfrak{B}$ — *борелевское множество.*

Определение. *Борелевская функция* — функция $f\colon \mathbb{R} \mapsto \mathbb{R}$:

$$\forall B \in \mathfrak{B} \quad f^{-1}(B) \in \mathfrak{B}.$$

Т.е. борелевская функция — это функция, для которой прообраз (множество $f^{-1}(B) = \{x\colon f(x) \in B\}$) любого борелевского множества также является борелевским множеством.

Пример. Функция Дирихле $D\colon \mathbb{R} \to \{0, 1\}$

$$D(x) = \begin{cases} 1, & x \in \mathbb{Q}; \\ 0, & x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases}$$

является борелевской.

В самом деле, прообразом любого борелевского множества A, такого что $1 \in A$, $0 \notin A$ является множество рациональных чисел; прообразом борелевского множества $B\colon 0 \in B$, $1 \notin B$ — множество иррациональных чисел; прооборазом борелевского множества $C\colon 0 \in C$, $1 \in C$ — вся вещественная прямая, наконец, прообразом борелевского множества $E\colon 0 \notin E$, $1 \notin E$ — пустое множество. Но $\mathbb{Q}$, $\mathbb{I}$, $\mathbb{R}$, $\varnothing$ — борелевские множества, а значит, выполняется определение борелевской функции.

Определение. Пусть $\mathcal{F}$ — σ-алгебра на некотором непустом множестве Ω. Функция ξ: $\Omega \mapsto \mathbb{R}$ называется *измеримой относительно σ-алгебры $\mathcal{F}$*, если полный прообраз борелевского множества B лежит в $\mathcal{F}$, т.е.

$$\xi^{-1}(B) = \{\omega\colon \xi(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathfrak{B}.$$

Замечание. Если в определении измеримой функции положить $\Omega = \mathbb{R}$ и выбрать борелевскую σ-алгебру $\mathfrak{B}$, то мы получим в точности определение борелевской функции. Т.е. борелевские функции — это подмножество измеримых функций.

Случайные величины

Определение. Пара $(X, \mathcal{U})$, где X – произвольное множество, а $\mathcal{U}$ – σ-алгебра над ним — *измеримое пространство.* Например, $(\Omega, \mathcal{F})$ и $(\mathbb{R}, \mathfrak{B})$ — измеримые пространства. Элементы σ-алгебры $\mathcal{U}$ называются *измеримыми множествами.*

Определение. Пусть даны измеримые пространства $(\Omega, \mathcal{F})$ и $(\mathbb{R}, \mathfrak{B})$. Тогда измеримая относительно $\mathcal{F}$ функция $\xi\colon \Omega \to \mathbb{R}$ называется *случайной величиной.*

Замечание. Если мы вспомним, что элементы σ-алгебры $\mathcal{F}$ называются событиями, то определение можно переформулировать следующим образом:

Пусть даны измеримые пространства $(\Omega, \mathcal{F})$ и $(\mathbb{R}, \mathfrak{B})$. Функция $\xi\colon \Omega \mapsto \mathbb{R}$ называется *случайной величиной*, если прообраз любого борелевского множества $B \in \mathfrak{B}$ является событием.

Пример. Пусть дана функция ξ:

$$\xi(\omega) = \begin{cases} 1, & \omega \in \left[0, \frac{1}{2}\right]; \\ 0, & \omega \in \left(\frac{1}{2}, 1\right], \end{cases}$$

$\Omega = [0, 1], \mathcal{F} = \{\varnothing, \Omega\}$ — минимальная σ-алгебра.

Докажем неизмеримость функции ξ; для этого достаточно найти такое борелевское множество, прообраз которого не будет принадлежать σ-алгебре. В данном случае $\mathcal{F}$ состоит всего лишь из двух множеств — $\{[0, 1], \varnothing\}$.

Как и в примере с функцией Дирихле, попробуем перебрать борелевские множества, содержащие значения $\xi(\omega)$. Тогда мы увидим, что для любого борелевского множества $A\colon 0 \in A$, $1 \notin A$ — например, множества $A_1 = \left(-\infty, \frac{1}{3}\right)$ — его прообразом является множество $\left(\frac{1}{2}, 1\right]$. Но это множество не входит в $\mathcal{F}$, а значит, $\xi(\omega)$ неизмерима относительно $\mathcal{F}$.

Отсюда можно сделать несколько выводов. Во-первых, измеримость функции зависит от выбора σ-алгебры (как и было подчёркнуто в определении). Например, если мы рассмотрим ту же функцию $\xi(\omega)$ на том же $\Omega = [0, 1]$, но с другой σ-алгеброй $\widehat{\mathcal{F}} = \left\{[0, 1], \left[0; \frac{1}{2}\right], \left(\frac{1}{2}, 1\right], \varnothing\right\}$, то наша функция будет измеримой, а следовательно — случайной величиной (проверьте!).

Во-вторых (забегая немного вперёд), именно из-за неизмеримости $\xi(\omega)$ относительно $\mathcal{F}$ мы не можем посчитать вероятность попадания значений этой функции в некоторые интервалы, к примеру, $\mathbb{P}\left(\xi < \frac{1}{3}\right)$. Ведь $\mathbb{P}\left(\xi < \frac{1}{3}\right) = \mathbb{P}\left(\xi \in A_1\right) = \mathbb{P}\left(\omega \in \left(\frac{1}{2}, 1\right]\right)$, но множество $\left(\frac{1}{2}, 1\right] \notin \mathcal{F}$, а вероятность — это отображение $\mathbb{P} : \mathcal{F} \mapsto \mathbb{R}$, и она не определена для этого множества.

Утверждение. *Пусть ξ — случайная величина, $g : \mathbb{R} \to \mathbb{R}$ — борелевская функция. Тогда $g(\xi)$ — случайная величина.*

Доказательство. Напомним, что функция является случайной величиной, если прообраз любого борелевского множества принадлежит σ-алгебре, то есть

$$\xi^{-1}(B) = \{\omega\colon \xi(\omega) \in B\} \in \mathcal{F} \qquad \forall B \in \mathfrak{B}.$$

Рассмотрим прообраз произвольного борелевского множества для $\eta = g(\xi)$:

$$\eta^{-1}(B) = \{\omega\colon g(\xi(\omega)) \in B)\} = \{\omega\colon \xi(\omega) \in g^{-1}(B)\}.$$

Функция g по предположению борелевская, следовательно, прообраз борелевского множества тоже будет борелевским: $g^{-1}(B) = C \in \mathfrak{B}$. В свою очередь,

ξ — случайная величина, и прообраз борелевского множества C лежит в σ-алгебре $\mathcal{F}$. Таким образом,

$$\eta^{-1}(B) = \{\omega\colon g(\xi(\omega)) \in B)\} = \{\omega\colon \xi(\omega) \in C)\} \in \mathcal{F}.$$

Мы получили, что прообораз произвольного борелевского множества принадлежит σ-алгебре. Значит, $\eta = g(\xi)$ — случайная величина. ■

Утверждение. *Пусть $\mathcal{E}$ — класс подмножеств $\mathbb{R}$, $\sigma(\mathcal{E}) = \mathfrak{B}$ (например, $\mathcal{E}$ — класс интервалов).*

Тогда ξ — случайная величина $\iff \forall E \in \mathcal{E}\colon\ \xi^{-1}(E) \in \mathcal{F}$.

Доказательство.

$\Longleftarrow$ Пусть $\mathcal{D} = \{D\colon D \in \mathfrak{B},\, \xi^{-1}(D) \in \mathcal{F}\}$. Тогда $\mathcal{E} \subseteq \mathcal{D}$. Далее, в силу свойств прообразов[1] и случайной величины ξ:

$$\xi^{-1}\left(\bigcup_i A_i\right) = \bigcup_i \xi^{-1}(A_i), \quad \xi^{-1}(\overline{A}) = \overline{\xi^{-1}(A)},$$
$$\xi^{-1}\left(\bigcap_i A_i\right) = \bigcap_i \xi^{-1}(A_i) \quad \forall A_i \in \mathcal{D}.$$

Следовательно, $\mathcal{D}$ — σ-алгебра. $\mathfrak{B} = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{D}) = \mathcal{D} \subseteq \mathfrak{B} \implies$
$\implies \mathfrak{B} = \mathcal{D}$.

$\Longrightarrow$ Следует непосредственно из определения случайной величины, т.к. $\mathcal{E} \subseteq \mathfrak{B}$.

■

Следствие. ξ *— случайная величина* $\iff \forall x \in \mathbb{R}\colon \{\omega\colon \xi(\omega) < x\} \in \mathcal{F}$. *Причем вместо знака $<$ может стоять любой другой знак неравенства, как строгого, так и нестрогого.*

Порождённое и индуцированное вероятностные пространства

Определение. σ-алгебра, порожденная случайной величиной ξ:

$$\mathcal{F}_\xi = \{\xi^{-1}(B),\, B \in \mathfrak{B}\}.$$

Отметим следующие факты:

1. $\mathcal{F}_\xi \subset \mathcal{F}$;

[1] Свойства нетрудно проверить исходя из определения прообраза: $\xi^{-1}(B) = \{\omega \in \Omega\colon \xi(\omega) \in B\}$.

2. $\mathcal{F}_\xi$ — σ-алгебра. Действительно:

$$\xi^{-1}(\overline{B}) = \overline{\xi^{-1}(B)}, \quad \xi^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} \xi^{-1}(B_i),$$

если B_i попарно не пересекаются.

Определение. Вероятностное пространство $(\Omega, \mathcal{F}_\xi, \mathbb{P})$ называется *порожденным случайной величиной* ξ.

Определение. *Распределение случайной величины* ξ — функция $P_\xi : \mathfrak{B} \mapsto \mathbb{R}$:

$$P_\xi(B) = \mathbb{P}(\xi^{-1}(B)) = \mathbb{P}(\xi \in B)$$

Замечание. Распределение — это композиция отображений. Если $\xi\colon \Omega \mapsto \mathbb{R}$, то полный прообраз — это отображение $\xi^{-1}\colon \mathfrak{B} \mapsto \mathcal{F}$. В свою очередь, $\mathbb{P}\colon \mathcal{F} \mapsto \mathbb{R}$. Тогда

$$P_\xi = \mathbb{P} \circ \xi^{-1}, \quad P_\xi\colon \mathfrak{B} \mapsto \mathbb{R}.$$

Определение. Вероятностное пространство $(\mathbb{R}, \mathfrak{B}, P_\xi)$ называется *индуцированным случайной величиной* ξ. [1]

Функция распределения, её свойства

Определение. *Функция распределения* $F_\xi(x)$ случайной величины ξ — функция $F_\xi\colon \mathbb{R} \mapsto \mathbb{R}$:

$$F_\xi(x) = P_\xi\left((-\infty, x)\right) = \mathbb{P}(\xi < x)$$

Утверждение. *$F_\xi(x)$ однозначно определяет $P_\xi(B)$.*

Доказательство. Действительно, любое борелевское множество может быть представлено в виде разности числовой оси, одной или двух полупрямых и не более чем счётного объединения отрезков. В силу однозначности определения $P_\xi([a;b]) = F_\xi(b+0) - F_\xi(a)$ утверждение теоремы справедливо. ■

Свойства функции распределения.

1. *$\forall x \quad 0 \leqslant F_\xi(x) \leqslant 1$;*
2. *$x_1 < x_2 \implies F_\xi(x_1) \leqslant F_\xi(x_2) \quad \forall x_1, x_2$ (монотонно неубывает);*
3. *$\lim\limits_{x\to+\infty} F_\xi(x) = 1$, $\lim\limits_{x\to-\infty} F_\xi(x) = 0$;*

[1] Нетрудно убедиться в том, что тройки $(\Omega, \mathcal{F}_\xi, \mathbb{P})$ и $(\mathbb{R}, \mathfrak{B}, P_\xi)$ удовлетворяют определению вероятностного пространства.

4. $F_\xi(x_0 - 0) = \lim\limits_{x \to x_0 - 0} F_\xi(x) = F_\xi(x_0)$ *(непрерывна слева).*

Доказательство.

1. Следует из свойств вероятности.

2. $x_1 < x_2 \implies \{\xi < x_1\} \subseteq \{\xi < x_2\}$. Из монотонности вероятности следует:
$$F_\xi(x_1) = \mathbb{P}(\xi < x_1) \leqslant \mathbb{P}(\xi < x_2) = F_\xi(x_2).$$

3. Пределы существуют в силу монотонности и ограниченности $F_\xi(x)$. Докажем, что $F_\xi(-n) \xrightarrow[n\to+\infty]{} 0$.

   Рассмотрим последовательность вложенных событий $B_n = \{\xi < -n\}$, $B_{n+1} = \{\xi < -(n+1)\} \subseteq B_n = \{\xi < -n\}\ \forall n\ \geqslant 1$:
$$\bigcap_{j=1}^{\infty} B_j = \{\omega\colon \xi(\omega) < x, \forall x \in \mathbb{R}\} \implies \bigcap_{j=1}^{\infty} B_j = \varnothing.$$

   $F_\xi(-n) = \mathbb{P}(B_n) \xrightarrow[n\to+\infty]{} \mathbb{P}(B) = 0$ (в силу непрерывности вероятностной меры). Отсюда следует:
$$F_\xi(n) \xrightarrow[n\to+\infty]{} 1 \iff 1 - F_\xi(n) = \mathbb{P}(\xi \geqslant n) \xrightarrow[n\to+\infty]{} 0.$$

4. Достаточно показать, что $F_\xi(x_0 - 1/n) \xrightarrow[n\to+\infty]{} F_\xi(x_0)$. Это равносильно
$$F_\xi(x_0) - F_\xi\left(x_0 - \frac{1}{n}\right) = \mathbb{P}(\xi < x_0) - \mathbb{P}\left(\xi < x_0 - \frac{1}{n}\right) =$$
$$= \mathbb{P}\left(x_0 - \frac{1}{n} \leqslant \xi < x_0\right) \xrightarrow[n\to+\infty]{} 0.$$

   Рассмотрим последовательность множеств $B_n = \{x_0 - \frac{1}{n} \leqslant \xi < x_0\}$. $B_1 \supset B_2 \supset \ldots \supset B_n \supset \ldots$; $\bigcap\limits_{k=1}^{n} B_k = B_n$; $\bigcap\limits_{k=1}^{\infty} B_k = \varnothing$. В силу непрерывности вероятностной меры $\lim\limits_{n\to\infty} \mathbb{P}(B_n) = \lim\limits_{n\to\infty} \mathbb{P}\left(\bigcap\limits_{k=1}^{n} B_k\right) = \mathbb{P}(\varnothing) = 0$.

■

Задача. Пусть есть не более чем счётное множество элементарных исходов Ω. Рассмотрим функции $f(\omega)\colon \Omega \mapsto \mathbb{R}$.

Какая функция всегда измерима (т.е. измерима относительно любой σ-алгебры)? Относительно какой σ-алгебры измерима любая функция $f(\omega)\colon \Omega \mapsto \mathbb{R}$?

1.4 Дискретные, сингулярные и абсолютно непрерывные функции распределения и случайные величины. Плотность распределения. Теорема Лебега о разложении функции распределения

Определение. Распределение ξ называется *дискретным*, если существует не более чем счётное множество B, т.ч. $P_\xi(B) = 1$. *Дискретная функция распределения* имеет вид:

$$F_\xi(x) = \mathbb{P}(\xi < x) = \sum_{x_i < x} p_i = \sum_{x_i < x} \mathbb{P}(\xi = x_i).$$

Замечание. Для любой дискретной функции распределения $F_\xi(x)$ число скачков — не более чем счётное.

Действительно, можно перенумеровать все скачки следующим образом:

$$\Delta_n = \left\{t\colon F_x(t+0) - F_x(t) > \frac{1}{n}\right\},\ |\Delta_n| \leqslant n.$$

Т.е. на каждом шаге мы считаем все скачки величины более $1/n$, а таких скачков не больше, чем n, так как функция распределения ограничена снизу нулём, сверху единицей, и, кроме того, монотонна. Множество точек разрыва представимо в виде $\bigcup\limits_{n=1}^{\infty} \Delta_n$, т.е. не более чем счётно.

Определение. Распределение ξ называется *абсолютно непрерывным*, если существует $f(x) \overset{\text{п.н.}}{\geqslant} 0$ такая, что для любого борелевского множества B справедливо

$$P_\xi(B) = \int\limits_B f(x)\lambda(dx),$$

где $f(x)$ — *плотность распределения*, λ — мера Лебега. *Абсолютно непрерывная функция распределения* имеет вид:

$$F_\xi(x) = \mathbb{P}(\xi < x) = \int\limits_{-\infty}^{x} f(t)dt.$$

Замечание.

1. $f(x) \overset{\text{п.н.}}{\geqslant} 0$ (почти наверное неотрицательна), если множество точек, где это неравенство не выполняется, имеет меру нуль по Лебегу, т.е. $\mathbb{P}(f(x) < 0) = 0$.

2. В определении абсолютно непрерывного распределения стоит не интеграл Римана, а его обобщение, *интеграл Лебега.*

3. В некоторых вариантах определения от плотности требуется неотрицательность не почти наверное, а всюду на $\mathbb{R}$. Это вопрос соглашения, так как интеграл Лебега по множеству меры нуль в любом случае равен нулю.

4. В случае абсолютно непрерывного распределения вероятность попасть в конкретную точку равна нулю. Действительно,

$$\mathbb{P}(\xi = x) = \int\limits_x^x f(t)dt = 0.$$

Свойства плотности.

1. *$f_\xi(x) = \frac{d}{dx}F(x)$ почти всюду (кроме, может быть, множества меры нуль по Лебегу — например, функция равномерного распределения $\mathrm{U}_{[0,1]}$ в точках 0 и 1 не дифференцируема);*

2. *$\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$ (нормировка).*

Доказательство. Первое свойство очевидно из свойств интегралов с переменным верхним пределом, рассмотрим второе. Если в определении абсолютно непрерывного распределения в качестве борелевского множества взять всю числовую прямую, получим:

$$\mathbb{P}(\xi \in \mathbb{R}) = 1 = \int\limits_{\mathbb{R}} f_\xi(x)dx.$$

■

Определение. *Точка роста* функции распределения $F_\xi(x)$ — точка x_0, для которой справедливо:

$$\forall \varepsilon > 0 \quad F_\xi(x_0 + \varepsilon) - F_\xi(x_0 - \varepsilon) > 0$$

Замечание. Возможен случай, когда точка роста является точкой разрыва:

$$\lim_{\varepsilon\to 0}\left(F_\xi\left(x_0+\varepsilon\right)-F_\xi\left(x_0-\varepsilon\right)\right)=F_\xi\left(x_0+0\right)-F_\xi\left(x_0\right)>0 \iff$$
$$\iff \lim_{\varepsilon\to 0}\mathbb{P}\left(x_0-\varepsilon\leqslant\xi<x_0+\varepsilon\right)=\mathbb{P}\left(\xi=x_0\right)>0$$

Определение. Функция распределения $F_\xi(x)$ называется *сингулярной*, если она непрерывна и множество точек её роста имеет нулевую меру Лебега.

Пример. Сингулярной функцией является *лестница Кантора*
$c\colon [0,1]\mapsto[0,1]$, которая строится следующим образом:

$c(0)=0, c(1)=1$. Далее интервал $(0,1)$ разбивается на три равные части $(0,1/3)$, $(1/3,2/3)$, $(2/3,1)$. На среднем интервале полагаем $c(x)=1/2$, оставшиеся два интервала снова разбиваются на три равные части каждый, и на соответствующих средних интервалах полагаем $c(x)=1/4$ и $c(x)=3/4$. Каждый из оставшихся интервалов снова делится на три части, и на внутренних интервалах $c(x)$ определяется как постоянная, равная среднему арифметическому между соседними, уже определенными значениями $c(x)$. На остальных точках единичного отрезка определяется по непрерывности.

Замечание. Так как любая функция распределения дифференцируема почти всюду, возможность дифференцировать функцию распределения никакого отношения к существованию плотности не имеет. Даже если мы дополнительно потребуем непрерывности функции распределения, этого не будет достаточно для абсолютной непрерывности распределения. К примеру, функция распределения сингулярного распределения непрерывна и дифференцируема почти всюду, однако плотности у этого распределения нет, так как производная функции распределения почти всюду равна нулю.

Теорема Лебега о разложении функции распределения. *Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$. Тогда существуют и определены единственным образом три функции распределения $F_{ac}(x), F_s(x), F_d(x)$, абсолютно непрерывная, сингулярная и дискретная соответственно, а также три числа $p_1,p_2,p_3\geqslant 0$, $p_1+p_2+p_3=1$ такие, что*

$$F_\xi(x)=p_1F_{ac}(x)+p_2F_s(x)+p_3F_d(x).$$

1.5 Числовые характеристики случайных величин: моменты, математическое ожидание, дисперсия. Их свойства

Математическое ожидание случайной величины

Определение. Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайная величина $\xi\colon \Omega \mapsto \mathbb{R}$. Если существует интеграл Лебега от ξ по мере $\mathbb{P}$ по множеству Ω, то он называется *математическим ожиданием* случайной величины ξ и обозначается как $\mathbb{E}\,\xi$ или $\mathrm{M}\xi$.

$$\mathbb{E}\,\xi = \int\limits_{\Omega} \xi(\omega)\,\mathbb{P}(d\omega).$$

Определение. *Математическое ожидание (среднее значение, первый момент)* случайной величины ξ, имеющей дискретное распределение со значениями $a_1, a_2, \ldots$ — сумма абсолютно[1] сходящегося ряда.

$$\mathbb{E}\,\xi = \sum_i a_i p_i = \sum_i a_i\,\mathbb{P}(\xi = a_i).$$

Определение. *Математическое ожидание* случайной величины ξ, имеющей абсолютно непрерывное распределение с плотностью распределения $f(x)$ — значение абсолютно сходящегося интеграла

$$\mathbb{E}\,\xi = \int\limits_{\mathbb{R}} x\,f(x)\,dx.$$

Матожидание имеет простой физический смысл: если на прямой разместить единичную массу, поместив в точки a_i массу p_i (для дискретного распределения) или «размазав» её с плотностью $f_\xi(x)$ (для абсолютно непрерывного), то точка $\mathbb{E}\,\xi$ будет координатой «центра тяжести» прямой.

Свойства математического ожидания. *Везде далее предполагается, что рассматриваемые математические ожидания существуют.*

[1] Согласно теореме Римана, члены условно сходящегося ряда можно переставить так, что он будет сходиться к любому наперёд заданному числу. Таким образом, если допустить условную сходимость, то математическое ожидание зависит от способа вычисления.

1. *Для произвольной борелевской функции $g(x)$ со значениями в $\mathbb{R}$:*

$$\mathbb{E}\, g(\xi) = \begin{cases} \sum\limits_k g\left(a_k\right) \mathbb{P}\left(\xi = a_k\right), & \text{если } P_\xi \text{ дискретно;} \\ \int\limits_{-\infty}^{+\infty} g(x) f_\xi(x)\, dx, & \text{если } P_\xi \text{ абсолютно непрерывно.} \end{cases}$$

Такое же свойство верно и для числовых функций нескольких аргументов $g(x_1, \ldots, x_n)$, если ξ — вектор из n случайных величин, а в сумме и в интеграле участвует их совместное распределение. Например, для $g(x, y) = x + y$ и для случайных величин ξ и η с плотностью совместного распределения $f(x, y)$ верно:

$$\mathbb{E}(\xi + \eta) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} (x + y) f(x, y) dx dy; \tag{1.1}$$

2. *Математическое ожидание линейно:*

$$\mathbb{E}(a\xi + b) = a\,\mathbb{E}\,\xi + b \quad \forall a, b \in \mathbb{R};$$

3. $\mathbb{E}(\xi + \eta) = \mathbb{E}\,\xi + \mathbb{E}\,\eta$;

4. *Если* $\xi \overset{\text{п.н.}}{\geqslant} 0$, *то* $\mathbb{E}\,\xi \geqslant 0$;

   Следствие.
   - *Если* $\xi \overset{\text{п.н.}}{\leqslant} \eta$, *то* $\mathbb{E}\,\xi \leqslant \mathbb{E}\,\eta$;
   - *Если* $a \overset{\text{п.н.}}{\leqslant} \xi \overset{\text{п.н.}}{\leqslant} b$, *то* $a \leqslant \mathbb{E}\,\xi \leqslant b$.

5. *Если ξ и η — независимые случайные величины, то* $\mathbb{E}(\xi\eta) = \mathbb{E}\,\xi\; \mathbb{E}\,\eta$;

   Замечание. *Обратное, вообще говоря, неверно.*

6. $|\mathbb{E}\,\xi| \leqslant \mathbb{E}\,|\xi|$;

7. $\xi \overset{\text{п.н.}}{\geqslant} 0,\ \mathbb{E}\,\xi = 0 \implies \xi \overset{\text{п.н.}}{=} 0$;

8. $\mathbb{P}(A) = \mathbb{E}\big(\mathrm{I}(\omega \in A)\big)$, *где* $\mathrm{I}(\omega \in A)$ — *индикатор:*

$$\mathrm{I}(\omega \in A) = \begin{cases} 1, & \omega \in A \\ 0, & \text{иначе;} \end{cases}$$

9. *Если функция $g(x)$ выпукла, то* $\mathbb{E}\, g(\xi) \geqslant g(\mathbb{E}\,\xi)$ *(неравенство Йенсена).*

Доказательство.

1. Достаточно рассмотреть случайную величину $\eta = g(\xi)$ на том же вероятностном пространстве и заметить, что $\forall \omega \in \Omega\colon \eta(\omega) = g(\xi(\omega))$. Тогда
$$\mathbb{E}\,\eta = \int\limits_{\Omega} \eta(\omega)\,\mathbb{P}(d\omega) = \int\limits_{\Omega} g(\xi(\omega))\,\mathbb{P}(d\omega).$$
Отсюда и вытекают формулы для дискретного и абсолютного случая.

2. Рассмотрим функцию $g(x) \equiv ax+b$ и произвольную случайную величину ξ. Тогда
$$\mathbb{E}(a\xi + b) = \mathbb{E}\,g(\xi) = \int\limits_{\Omega} g(\xi(\omega))\,\mathbb{P}(d\omega) = a\int\limits_{\Omega} \xi(\omega)\,\mathbb{P}(d\omega) + b\int\limits_{\Omega} \mathbb{P}(d\omega) =$$
$$= a\,\mathbb{E}\,\xi + b\,\mathbb{P}(\Omega) = a\,\mathbb{E}\,\xi + b.$$

3. Воспользуемся равенством (1.1) и теоремой о совместном распределении:
$$\begin{aligned}\mathbb{E}(\xi+\eta) &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x+y)f(x,y)\,dxdy = \\ &= \int\limits_{-\infty}^{\infty} x\,dx \int\limits_{-\infty}^{\infty} f(x,y)\,dy + \int\limits_{-\infty}^{\infty} y\,dy \int\limits_{-\infty}^{\infty} f(x,y)\,dx = \\ &= \int\limits_{-\infty}^{\infty} xf_\xi(x)\,dx + \int\limits_{-\infty}^{\infty} yf_\eta(y)\,dy = \mathbb{E}\,\xi + \mathbb{E}\,\eta.\end{aligned}$$

4. Неотрицательность ξ означает, что $a_i \geqslant 0$ при всех $i\colon p_i > 0$ в случае дискретного распределения, либо $f_\xi(x) = 0$ при $x < 0$ (кроме, может быть, множества меры нуль) - для абсолютно непрерывного распределения. И в том, и в другом случае имеем:
$$\mathbb{E}\,\xi = \sum a_i p_i \geqslant 0 \quad \text{или} \quad \mathbb{E}\,\xi = \int\limits_{0}^{\infty} xf(x)\,dx \geqslant 0.$$

5. В равенстве (1.1) заменим сложение умножением и плотность совместного распределения произведением плотностей (это возможно в силу

независимости случайных величин):

$$\mathbb{E}(\xi\eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xy f_\xi(x) f_\eta(y)\, dxdy =$$
$$= \int\limits_{-\infty}^{\infty} x f_\xi(x)\, dx \int\limits_{-\infty}^{\infty} y f_\eta(y)\, dy = \mathbb{E}\,\xi\,\mathbb{E}\,\eta.$$

6. Это верно в силу неравенства треугольника (для дискретного случая) и аналогичного неравенства для интегралов (для непрерывного случая).

7. • Дискретный случай:
   Учитывая то, что $a_k \geqslant 0$, равенство $\mathbb{E}\,\xi = \sum a_k p_k = 0$ означает, что все слагаемые в этой сумме равны нулю, т. е. все вероятности p_k нулевые, кроме вероятности, соответствующей значению $a_k = 0$.
   • Абсолютно непрерывный случай:
   Условие $\xi \overset{\text{п.н.}}{\geqslant} 0$ говорит о том, что при $x < 0$ плотность равна нулю всюду, кроме, может быть, множества меры нуль по Лебегу. В самом деле, если существует борелевское множество $B \subset \{x \mid x < 0\}$ ненулевой меры и при этом $f_\xi(x) > 0\ \forall x \in B$, то $\mathbb{P}\,(\xi \in B) = \int\limits_B f_\xi(x)\, dx > 0$, т.е. случайная величина принимает отрицательные значения с ненулевой вероятностью, что противоречит ограничению. Учитывая это, можем написать

$$\mathbb{E}\,\xi = \int\limits_{-\infty}^{+\infty} x f_\xi(x)\, dx = \int\limits_{0}^{+\infty} x f_\xi(x)\, dx \geqslant 0,$$

   так как на промежутке интегрирования $x \geqslant 0$, а плотность по определению неотрицательна.

8. Следует непосредственно из определений индикатора и матожидания.

9. Начнём со следующего утверждения: если функция g выпукла, то для любого $y \in \mathbb{R}\ \exists c = c(y)\colon \forall x \in \mathbb{R}\ g(x) \geqslant g(y) + c(y)(x - y)$. Это вытекает из того, что график выпуклой функции лежит не ниже любой из касательной к нему. [1] Положим в этом неравенстве $y = \mathbb{E}\,\xi$. Тогда

$$g(\xi) \geqslant g(\mathbb{E}\,\xi) + c(\mathbb{E}\,\xi)(\xi - \mathbb{E}\,\xi), \quad \mathbb{E}\,g(\xi) \geqslant \mathbb{E}\,g(\mathbb{E}\,\xi) + \mathbb{E}\,c(\mathbb{E}\,\xi)(\xi - \mathbb{E}\,\xi)$$

[1] Вообще говоря, выпуклая функция может не иметь первой производной и, следовательно, касательной на не более чем счётном множестве точек, но тогда её можно заменить на опорную гиперплоскость.

Здесь $g(\mathbb{E}\,\xi), c(\mathbb{E}\,\xi)$ — константы, а $\mathbb{E}(\xi - \mathbb{E}\,\xi) = 0$, а значит, $\mathbb{E}\,g(\xi) \geqslant g(\mathbb{E}\,\xi)$.

■

Дисперсия и моменты старших порядков

Определение. Пусть $\mathbb{E}\,|\xi|^k < \infty$.

1. $\mathbb{E}\,\xi^k$ — *момент порядка k или k-й момент* случайной величины ξ;
2. $\mathbb{E}\,|\xi|^k$ — *абсолютный k-й момент*;
3. $\mu_k = \mathbb{E}(\xi - \mathbb{E}\,\xi)^k$ — *центральный k-й момент*;
4. $\mathbb{E}\,|\xi - \mathbb{E}\,\xi|^k$ — *абсолютный центральный k-й момент* случайной величины ξ

Определение. Число $\mathbb{D}\,\xi = \mathbb{E}(\xi - \mathbb{E}\,\xi)^2$ (центральный момент второго порядка) называется *дисперсией* случайной величины ξ, $\sigma = \sqrt{\mathbb{D}\,\xi}$ — её *среднеквадратичным отклонением.*

Утверждение. *Если существует момент порядка $t > 0$ случайной величины ξ, то существует и ее момент порядка s, где $0 < s < t$.*

Доказательство. Заметим, что $|\xi|^s \leqslant |\xi|^t + 1$. В силу следствия из свойства 4 для математического ожидания можно получить из неравенства для случайных величин такое же неравенство для их математических ожиданий: $\mathbb{E}\,|\xi|^s \leqslant \mathbb{E}\,|\xi|^t + 1 < \infty$. ■

Свойства дисперсии. *Везде далее предполагается, что вторые моменты рассматриваемых случайных величин существуют. Тогда, в силу вышеописанной теоремы, существуют и матожидания.*

1. *Дисперсия может быть вычислена по формуле:* $\mathbb{D}\,\xi = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2$*;*
2. *При умножении случайной величины на постоянную c дисперсия увеличивается в c^2 раз:* $\mathbb{D}(c\xi) = c^2\,\mathbb{D}\,\xi$*;*
3. *Дисперсия всегда неотрицательна:* $\mathbb{D}\,\xi \geqslant 0$*;*
4. *Дисперсия обращается в нуль лишь для вырожденного распределения: если $\mathbb{D}\,\xi = 0$, то $\xi \overset{п.н.}{=} const$, и наоборот;*
5. *Если ξ и η независимы, то* $\mathbb{D}(\xi + \eta) = \mathbb{D}\,\xi + \mathbb{D}\,\eta$, $\mathbb{D}(\xi - \eta) = \mathbb{D}\,\xi + \mathbb{D}\,\eta$*;*

6. *Дисперсия не зависит от сдвига случайной величины на постоянную:* $\mathbb{D}(\xi + c) = \mathbb{D}\,\xi$;

Доказательство.

1. Обозначим для удобства $a = \mathbb{E}\,\xi$. Тогда

$$\mathbb{D}\,\xi = \mathbb{E}(\xi - a)^2 = \mathbb{E}\left(\xi^2 - 2a\xi + a^2\right) = \mathbb{E}\,\xi^2 - 2a\,\mathbb{E}\,\xi + a^2 = \mathbb{E}\,\xi^2 - a^2.$$

2. $\mathbb{D}(c\xi) = \mathbb{E}(c\xi)^2 - (\mathbb{E}(c\xi))^2 = c^2\,\mathbb{E}\,\xi^2 - (c\,\mathbb{E}\,\xi)^2 = c^2(\mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2) = c^2\,\mathbb{D}\,\xi$

3. Пусть $a = \mathbb{E}\,\xi$. Дисперсия есть математическое ожидание неотрицательной случайной величины $(\xi - a)^2$, откуда (и из свойства (4) матожидания) следует неотрицательность дисперсии.

4. $\mathbb{D}\,\xi = 0 \implies (\xi - a)^2 \overset{\text{п.н.}}{=} 0$, $\xi \overset{\text{п.н.}}{=} a = \text{const}$. И наоборот: если $\xi \overset{\text{п.н.}}{=} c$, то $\mathbb{D}\,\xi = \mathbb{E}(c - \mathbb{E}\,c)^2 = 0$.

5. Действительно, применяя свойство (5) матожидания, получим:

$$\begin{aligned}\mathbb{D}(\xi + \eta) &= \mathbb{E}(\xi + \eta)^2 - (\mathbb{E}(\xi + \eta))^2 = \\ &= \mathbb{E}\,\xi^2 + \mathbb{E}\,\eta^2 + 2\,\mathbb{E}(\xi\eta) - (\mathbb{E}\,\xi)^2 - (\mathbb{E}\,\eta)^2 - 2\,\mathbb{E}\,\xi\ \mathbb{E}\,\eta = \mathbb{D}\,\xi + \mathbb{D}\,\eta.\end{aligned}$$

   Обратное, аналогично замечанию к свойству (5) матожидания, неверно. Для разности, соответственно, имеем:

$$\mathbb{D}(\xi - \eta) = \mathbb{D}(\xi + (-\eta)) = \mathbb{D}\,\xi + \mathbb{D}(-\eta) = \mathbb{D}\,\xi + (-1)^2\,\mathbb{D}\,\eta = \mathbb{D}\,\xi + \mathbb{D}\,\eta.$$

   Следствие. *Для произвольных случайных величин ξ и η имеет место равенство:*

$$\mathbb{D}(\xi \pm \eta) = \mathbb{D}\,\xi + \mathbb{D}\,\eta \pm 2(\mathbb{E}(\xi\eta) - \mathbb{E}\,\xi\ \mathbb{E}\,\eta),$$

   где величина $\mathbb{E}(\xi\eta) - \mathbb{E}\,\xi\ \mathbb{E}\,\eta$ называется ковариацией случайных величин ξ и η ($cov(\xi, \eta)$).

6. $\mathbb{D}(\xi + c) = \mathbb{D}\,\xi + \mathbb{D}\,c = \mathbb{D}\,\xi$

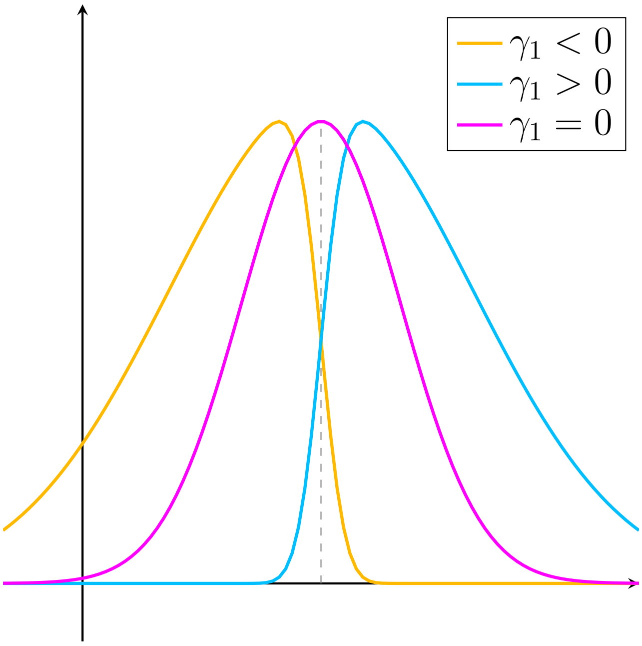
■

Прочие числовые характеристики

Определение. *Коэффициент асимметрии* случайной величины ξ:

$$\gamma_1 = \mathbb{E}\left(\frac{\xi - \mathbb{E}\,\xi}{\sqrt{\mathbb{D}\,\xi}}\right)^3 = \frac{\mu^3}{\sigma^3}$$

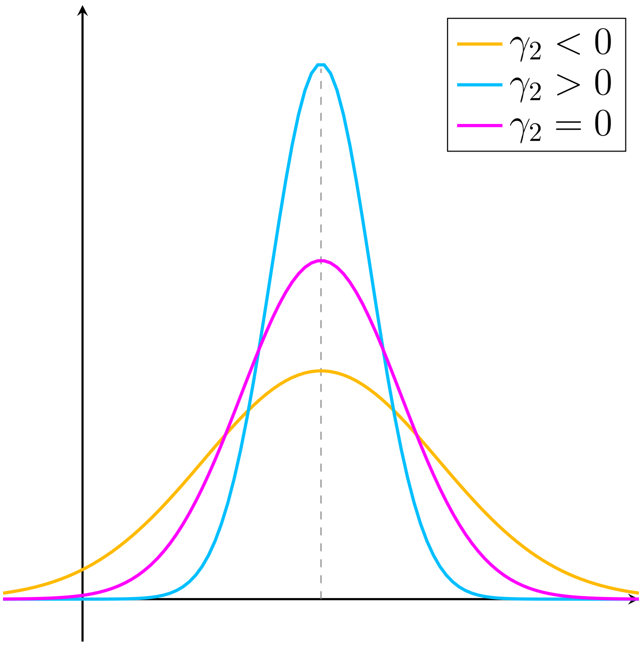
Характеризует «скошенность» графика плотности распределения:



Определение. *Коэффициент эксцесса* случайной величины ξ:

$$\gamma_2 = \mathbb{E}\left(\frac{\xi - \mathbb{E}\,\xi}{\sqrt{\mathbb{D}\,\xi}}\right)^4 - 3 = \frac{\mu^4}{\sigma^4} - 3$$

Характеризует «островершинность» графика плотности распределения:



Замечание. Слагаемое -3 добавлено, чтобы коэффициент эксцесса стандартного нормального распределения был равен нулю. Иногда его не учитывают и считают, что коэффициент эксцесса $\mathrm{N}_{0,1}$ равен 3.

1.6 Числовые характеристики случайных величин: квантили. Медиана и ее свойства. Интерквартильный размах

Определение. *Медианой* $\mathrm{Med}\,\xi$ распределения случайной величины ξ называется любое из *чисел* μ таких, что

$$\mathbb{P}(\xi \leqslant \mu) \geqslant \frac{1}{2}, \quad \mathbb{P}(\xi \geqslant \mu) \geqslant \frac{1}{2}.$$

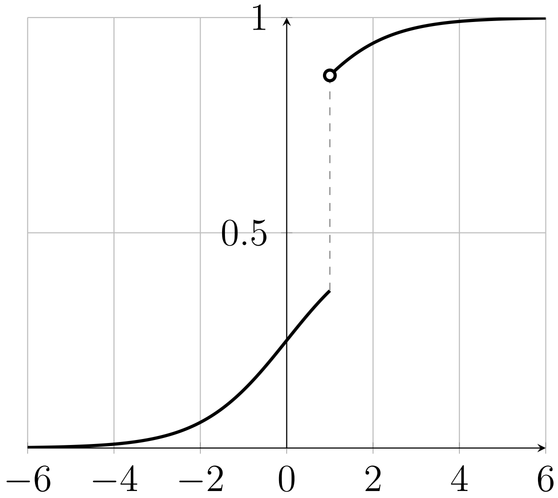
Замечание. Медиана распределения всегда существует, но может быть не единственна. Например, в случае дискретного распределения $\mathbb{P}(\xi = -1) = \mathbb{P}(\xi = 1) = \frac{1}{2}$ и 1, и -1 удовлетворяют определению медианы.

Определение. *Квантиль порядка* γ — это такое число κ_γ, для которого выполняется

$$\begin{cases} \mathbb{P}(\xi \leqslant \kappa_\gamma) = F(\kappa_\gamma) \geqslant \gamma, \\ \mathbb{P}(\xi \geqslant \kappa_\gamma) = 1 - F(\kappa_\gamma + 0) \geqslant 1 - \gamma \end{cases}$$

Замечание. Если функция распределения F непрерывна и строго монотонна, то *квантилем* порядка (уровня) γ, где $\gamma \in (0; 1)$, является решение x_γ уравнения $F(x_\gamma) = \gamma$. Тогда квантиль порядка γ отрезает от области под графиком плотности область с площадью γ слева от себя. Справа от κ_γ площадь области равна $1 - \gamma$.

Если же случайная величина не является абсолютно непрерывной, то уравнение $F(x_\gamma) = \gamma$ может не иметь решений. Например, для приведенного ниже графика не существует $x_{\frac{1}{2}} : F(x_{\frac{1}{2}}) = \frac{1}{2}$.



Определение. Квантили уровней, кратных 0.01, называют *процентилями*, квантили уровней, кратных 0.1, — *децилями*, уровней, кратных 0.25, — *квартилями*.

Замечание. Медиана является квантилем уровня $1/2$.

Свойства медианы.

1. *Медиана случайной величины ξ минимизирует средний модуль отклонения ξ:*
$$\mathbb{E}\,|\xi - \operatorname{Med}\xi| = \min_a \mathbb{E}\,|\xi - a|;$$

2. *Отклонение медианы случайной величины ξ от её матожидания $\mathbb{E}\,\xi$ не превышает по модулю среднеквадратичного отклонения $\sigma = \sqrt{\mathbb{D}\,\xi}$:*
$$|\,\mathbb{E}\,\xi - \operatorname{Med}\xi| \leqslant \sigma.$$

Доказательство.

1. Рассмотрим случайную величину $\eta = \xi - \operatorname{Med}\xi$. Очевидно, что $\operatorname{Med}\eta = 0$. Тогда нам надо показать, что $\forall c \in \mathbb{R}$ справедливо
$$\mathbb{E}\,|\eta - c| - \mathbb{E}\,|\eta| \geqslant 0.$$
Рассмотрим случай $c > 0$. Заметим, что
$$\begin{cases} |\eta - c| - |\eta| = c, & \eta < 0; \\ |\eta - c| - |\eta| \geqslant -c, & \eta \geqslant 0. \end{cases}$$
Тогда
$$\mathbb{E}\big(|\eta - c| - |\eta|\big) = \mathbb{E}\big((|\eta - c| - |\eta|)\cdot \mathrm{I}(\eta < 0) + (|\eta - c| - |\eta|)\cdot \mathrm{I}(\eta \geqslant 0)\big)$$
$$\mathbb{E}\big(|\eta - c| - |\eta|\big) \geqslant c\ \mathbb{P}(\eta < 0) - c\ \mathbb{P}(\eta \geqslant 0).$$
Так как $\operatorname{Med}\eta = 0$, то $\mathbb{P}(\eta \leqslant 0) = \mathbb{P}(\eta \geqslant 0) = \frac{1}{2} \implies \mathbb{E}\big(|\eta - c| - |\eta|\big) \geqslant 0.$.
Случай $c < 0$ сводится к предыдущему умножением случайной величины и c на -1. Отсюда следует, что медиана действительно минимизирует средний модуль отклонения.

2. Рассмотрим цепочку неравенств:
$$|\,\mathbb{E}\,\xi - \operatorname{Med}\xi| = |\,\mathbb{E}(\operatorname{Med}\xi - \xi)| \leqslant \{\text{св-во (6) матожидания}\} \leqslant \mathbb{E}\,|\operatorname{Med}\xi - \xi| \leqslant$$
$$\{\text{св-во (1) медианы}\} \leqslant \mathbb{E}\,|\,\mathbb{E}\,\xi - \xi| = \mathbb{E}\,\sqrt{|\,\mathbb{E}\,\xi - \xi|^2} \leqslant \{\text{нер-во Йенсена}\} \leqslant$$
$$\leqslant \sqrt{\mathbb{E}\,|\,\mathbb{E}\,\xi - \xi|^2} = \sqrt{\mathbb{D}\,\xi} = \sigma.$$

$\blacksquare$

Определение. *Интерквартильным размахом* называется разность между третьим и первым квартилями, то есть $x_{0{,}75} - x_{0{,}25}$.

В каком-то смысле эту величину можно считать аналогом дисперсии случайной величины, устойчивой к выбросам.

1.7 Испытания Бернулли. Биномиальное распределение. Теорема Пуассона. Распределение Пуассона

Определение. *Испытание Бернулли* — случайный эксперимент, у которого есть ровно два возможных исхода: «успех» и «неудача». Как правило, вероятность успеха обозначается буквой p, вероятность неудачи $q = 1 - p$.

Определение. *Схема Бернулли* — последовательность из n *независимых однородных* испытаний Бернулли с вероятностью успеха p и неудачи $q = 1 - p$.

Со схемой Бернулли можно связать последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, где $\xi_k = \begin{cases} 1, & \text{с вероятностью } p; \\ 0, & \text{с вероятностью } q, \end{cases} \quad k = \overline{1, n}$.

Таким образом, принятие случайной величиной ξ_k значения 1 интерпретируется как успех в k-м испытании. Так как испытания в схеме Бернулли независимы и однородны, данные случайные величины должны быть *независимы и одинаково распределены.*

Формула Бернулли. *Пусть ξ — случайная величина, равная числу успехов в n испытаниях. Тогда $\forall k = \overline{1, n}$ вероятность получить в n испытаниях ровно k успехов равна*

$$\mathbb{P}\left(\xi = k\right) = C_n^k p^k q^{n-k}$$

Доказательство. Рассмотрим один элементарный исход события $A = \{\xi = k\}$:

$$(\underbrace{у, у, \ldots, у}_{k}, \underbrace{н,\ н, \ldots,\ н}_{n-k})$$

когда первые k испытаний завершились успехом (у), остальные неудачей (н). Поскольку испытания независимы, вероятность такого элементарного исхода равна $p^k(1-p)^{n-k}$. Другие элементарные исходы из события A отличаются лишь расположением k успехов на n местах. Поэтому событие A состоит из C_n^k элементарых исходов, вероятность каждого из которых равна $p^k q^{n-k}$. $\blacksquare$

Определение. Пусть $\xi_1, \ldots, \xi_n$ — последовательность независимых случайных величин, имеющих одинаковое распределение Бернулли с параметром p (B_p), то есть принимает значение 1 («успех») с вероятностью p и 0 («неудача») с вероятностью $1-p=q$. Тогда говорят, что случайная величина $\xi = \xi_1 + \ldots + \xi_n$ имеет *биномиальное распределение* с параметрами n и p ($\mathrm{Bi}_{n,p}$).

Числовые характеристики B_p

1. Математическое ожидание:

$$\mathbb{E}\,\xi = 1\cdot p + 0 \cdot q = p$$

2. Дисперсия:

$$\mathbb{E}\,\xi^2 = 1^2\cdot p + 0^2\cdot q = p; \quad \mathbb{D}\,\xi = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2 = p - p^2 = p\cdot(1-p) = pq$$

Числовые характеристики $\mathrm{Bi}_{n,p}$

1. Математическое ожидание:

$$\mathbb{E}\,\xi = \mathbb{E}(\xi_1 + \ldots + \xi_n) = \mathbb{E}\,\xi_1 + \ldots + \mathbb{E}\,\xi_n = \underbrace{p + \ldots + p}_{n} = np$$

2. Дисперсия:

$$\mathbb{D}\,\xi = \mathbb{D}(\xi_1 + \ldots + \xi_n) = \mathbb{D}\,\xi_1 + \ldots + \mathbb{D}\,\xi_n = \underbrace{p\cdot q + \ldots + p\cdot q}_{n} = npq$$

Теорема Пуассона. *Пусть проводится n обобщённых испытаний Бернулли (т.е. вероятность успеха испытания зависит от n) с вероятностью успеха p_n, ξ — количество успехов в этих испытаниях и $np_n \xrightarrow[n\to+\infty]{} \lambda$. Тогда*

$$\forall k \in \mathbb{Z},\ 0 \leqslant k \leqslant n: \quad \mathbb{P}\,(\xi = k) \xrightarrow[n\to+\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

Доказательство. По условию теоремы, $np_n \xrightarrow[n\to+\infty]{} \lambda$. Тогда $p_n \xrightarrow[n\to+\infty]{} 0$. Рассмотрим формулу Бернулли:

$$\mathbb{P}(\xi = k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{-k}\cdot\left(1-\frac{\lambda}{n}\right)^{n} =$$
$$= \frac{\lambda^k}{k!}\cdot\frac{(n-k+1)\cdot\ldots\cdot(n-1)\cdot n}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n}\cdot\left(1-\frac{\lambda}{n}\right)^{-k}$$

Перейдём к пределу при $n \to +\infty$:

$$\frac{(n-k+1)\cdot\ldots\cdot(n-1)\cdot n}{n^k} \to 1,\ \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda},\ \left(1-\frac{\lambda}{n}\right)^{-k} \to 1$$

Таким образом, получим

$$\mathbb{P}(\xi = k) \xrightarrow[n\to+\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

$\blacksquare$

Определение. Набор вероятностей $\{\frac{\lambda^k}{k!}e^{-\lambda}\}$, где k принимает значения $0, 1, 2, \ldots$, называется *распределением Пуассона* с параметром $\lambda > 0$ (Pois_λ).

Замечание. Распределение Пуассона представляет собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью (за которую отвечает параметр λ) и независимо друг от друга.

Числовые характеристики Pois_λ

1. Математическое ожидание:

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

$$\mathbb{E}\,\xi = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda\underbrace{\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}}_{=e^{\lambda}} = \lambda$$

2. Дисперсия:

$$\mathbb{E}\,\xi(\xi-1) = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2$$

$$\mathbb{E}\,\xi^2 = \mathbb{E}\,\xi(\xi-1) + \mathbb{E}\,\xi = \lambda^2 + \lambda \quad \Longrightarrow \quad \mathbb{D}\,\xi = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2 = \lambda$$

1.8 Испытания Бернулли. Геометрическое распределение. Теорема Реньи. Показательное распределение

Рассмотрим бесконечную схему экспериментов Бернулли с вероятностью успеха p, неудачи — $q = 1-p$. Вероятность того, что первый успех произойдёт в испытании с номером $k \in \mathbb{N}$, очевидно, равна $\mathbb{P}(\tau = k) = pq^{k-1}$. Действительно, это событие равносильно тому, что $\xi_1 = \xi_2 = \ldots = \xi_{k-1} = 0,\ \xi_k = 1$. В силу независимости ξ_i

$$\mathbb{P}(\xi_1 = 0,\ \xi_2 = 0, \ldots, \xi_{k-1} = 0,\ \xi_k = 1) = \\ \mathbb{P}(\xi_1 = 0) \cdot \mathbb{P}(\xi_2 = 0) \cdot \ldots \cdot \mathbb{P}(\xi_{k-1} = 0) \cdot \mathbb{P}(\xi_k = 1) = q^{k-1}p.$$

Определение. Набор вероятностей $\{pq^{k-1}\}$, где k принимает любые значения из множества натуральных чисел, называется *геометрическим распределением* вероятностей (Geom_p).

Аналогично можно ввести геометрическое распределение как «число неудач до первого успеха». Тогда k будет принимать значения из множества $\{0, 1, 2, \ldots\}$, и $\mathbb{P}(\tau' = k) = pq^k$.

Числовые характеристики Geom_p

1. Математическое ожидание:

$$\mathbb{E}\,\xi = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} = p\sum_{k=1}^{\infty} \frac{dq^k}{dq} = \\ = p\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\frac{1}{(1-q)^2} = \frac{1}{p}.$$

2. Дисперсия:

$$\mathbb{E}\,\xi(\xi-1) = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = pq\sum_{k=0}^{\infty} \frac{d^2q^k}{dq^2} = pq\frac{d^2}{dq^2}\left(\sum_{k=0}^{\infty} q^k\right) = \\ = pq\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) = pq\frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$\mathbb{D}\,\xi = \mathbb{E}\,\xi(\xi-1) + \mathbb{E}\,\xi - (\mathbb{E}\,\xi)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q-1+p}{p^2} = \frac{q}{p^2}.$$

Замечание. Если определять геометрическое распределение как количество неудач до первого успеха, его математическое ожидание изменится:

$$\mathbb{E}\,\xi = \sum_{k=0}^{\infty} kpq^{k} = qp\sum_{k=1}^{\infty} kq^{k-1} = qp\sum_{k=1}^{\infty} \frac{dq^k}{dq} =$$

$$= qp\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) = qp\frac{d}{dq}\left(\frac{q}{1-q}\right) = qp\frac{1}{(1-q)^2} = \frac{q}{p}$$

Так как $q \in (0, 1)$, математическое ожидание станет меньше, и это логично — ведь количество неудач до первого успеха всегда на единицу меньше номера первого успеха. (Используя это наблюдение, можно посчитать матожидание ещё проще — $\mathbb{E}(\xi - 1) = \frac{1}{p} - 1 = \frac{1-p}{p} = \frac{q}{p}$). Дисперсия же не зависит от сдвига и останется прежней.

Определение. Случайная величина ξ имеет *показательное (экспоненциальное) распределение* с параметром $\lambda > 0$ (Exp_λ), если ξ имеет следующие плотность и функцию распределения:

$$f(x) = \begin{cases} 0, & \text{если } x < 0; \\ \lambda e^{-\lambda x}, & \text{если } x \geqslant 0. \end{cases} \quad F(x) = \begin{cases} 0, & \text{если } x < 0; \\ 1 - e^{-\lambda x}, & \text{если } x \geqslant 0. \end{cases}$$

Замечание. Показательное распределение моделирует время между двумя последовательными свершениями одного и того же события. К примеру, пусть есть магазин, в который время от времени заходят покупатели. При определённых допущениях время между появлениями двух последовательных покупателей будет случайной величиной с экспоненциальным распределением. Среднее время ожидания нового покупателя равно $\frac{1}{\lambda}$. Сам параметр λ тогда может быть интерпретирован как среднее число новых покупателей за единицу времени.

Числовые характеристики Exp_λ

Найдём для произвольного $k \in \mathbb{N}$ момент порядка k:

$$\mathbb{E}\,\xi^k = \int_{-\infty}^{\infty} x^k f_\xi(x)dx = \int_{0}^{\infty} x^k \lambda e^{-\lambda x}dx = \frac{1}{\lambda^k}\int_{0}^{\infty} (\lambda x)^k e^{-\lambda x} d(\lambda x) = \frac{k!}{\lambda^k}$$

В последнем равенстве была использована формула для гамма-функции:

$$\Gamma(k+1) = \int_{0}^{\infty} u^k e^{-u} du = k!$$

1. Математическое ожидание: $\mathbb{E}\,\xi = \frac{1}{\lambda}$.
2. Дисперсия: $\mathbb{E}\,\xi^2 = \frac{2}{\lambda^2}$, $\mathbb{D}\,\xi = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2 = \frac{1}{\lambda^2}$.

Важным требованием в схеме Бернулли является однородность, т.е. постоянство параметра p на протяжении всех испытаний. Но иногда есть желание посмотреть, например, что получится, если устремить вероятность успеха к нулю, а количество испытаний к бесконечности. В этом случае придётся прибегнуть к нетривиальной модели эксперимента.

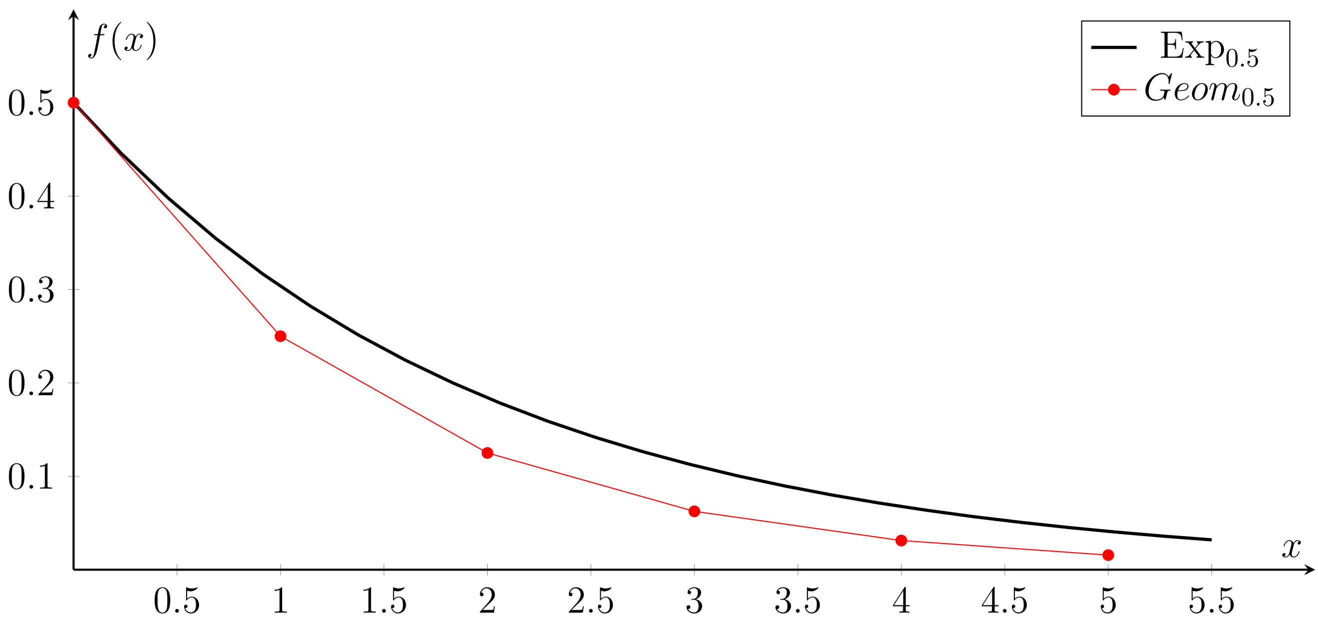
Утверждение. *Рассмотрим серию схем Бернулли — последовательность схем Бернулли с количеством испытаний n и вероятностями успеха p_n (т.е. в первой схеме — одно испытание с вероятностью успеха p_1, в второй — два испытания с вероятностью успеха p_2 и т.д.). В каждой схеме рассмотрим случайную величину $\tau_n \sim \mathrm{Geom}_{p_n}$. Пусть $np_n \xrightarrow[n\to\infty]{} \lambda > 0$.*

Тогда распределение случайной величины $\frac{\tau_n}{n}$ сходится к показательному с параметром λ при $n \to +\infty$.

Доказательство. Пусть F_n — функция распределения случайной величины $\frac{\tau_n}{n}$. Тогда для $x \geqslant 0$:

$$F_n(x) = \mathbb{P}\left(\frac{\tau_n}{n} \leqslant x\right) = \mathbb{P}\left(\tau_n \leqslant nx\right) = \mathbb{P}\left(\tau_n \leqslant \lfloor nx \rfloor\right) =$$
$$= \sum_{k=1}^{\lfloor nx \rfloor} p_n q_n^{k-1} = p_n \frac{1 - q_n^{\lfloor nx \rfloor + 1}}{1 - q_n} = 1 - (1 - p_n)^{\lfloor nx \rfloor + 1}$$

Далее, т.к. $(1-p_n)^n = \left(1 - \frac{np_n}{n}\right)^n \xrightarrow[n\to+\infty]{} e^{-\lambda}$, то $(1-p_n)^{nx} \xrightarrow[n\to+\infty]{} e^{-\lambda x}$. По определению, $\lfloor nx \rfloor \leqslant nx < \lfloor nx \rfloor + 1$, или, что эквивалентно, $nx < \lfloor nx \rfloor + 1 \leqslant nx + 1$. Таким образом верно $(1-p_n)^{\lfloor nx \rfloor + 1} \xrightarrow[n\to+\infty]{} e^{-\lambda x}$. Следовательно, $F_n(x) \xrightarrow[n\to+\infty]{} 1 - e^{-\lambda x}$, что есть функция показательного распределения. ■



Теорема Реньи. *Пусть даны случайная величина* $N \sim \mathrm{Geom}_p$, $\xi_1, \xi_2, \ldots$ — *независимые одинаково распределённые случайные величины,* $\xi_i \geqslant 0$ *и* $0 < a = \mathbb{E}\,\xi_i < \infty$, $S_n = \sum\limits_{i=1}^{N} \xi_i$. *Тогда*

$$\sup_x \left| \mathbb{P}\left(\frac{p}{a} S_N < x\right) - G(x) \right| \underset{p \to 0}{\longrightarrow} 0,$$

где $G(x) = (1 - e^{-x})\,\mathrm{I}(x \geqslant 0)$ — *функция стандартного показательного распределения* Exp_1.

Если $b^2 = \mathbb{E}\,\xi_i^2$, *то* $\sup\limits_x \left| \mathbb{P}\left(\frac{p}{a} S_N < x\right) - G(x) \right| \leqslant \frac{pb^2}{(1-p)a^2}$.

1.9 Испытания Бернулли. Теорема Муавра—Лапласа. Нормальное распределение

Локальная предельная теорема Муавра—Лапласа. *Пусть* S_n — *число успехов в* n *испытаниях Бернулли с вероятностью успеха* $0 < p < 1$. *Пусть* $n \to \infty$, *тогда* $np(1-p) \longrightarrow \infty$, *и*

$$\forall m \in \mathbb{Z} : 0 \leqslant m \leqslant n \quad \mathbb{P}\left(S_n = m\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(1 + \underline{O}\left(\frac{1}{\sigma}\right)\right),$$

где $x = \frac{m - np}{\sigma}$, *а* $\sigma = \sqrt{\mathbb{D}\,S_n} = \sqrt{np(1-p)}$.

Интегральная теорема Муавра—Лапласа. *Пусть* S_n — *число успехов в* n *испытаниях Бернулли с вероятностью успеха* $0 < p < 1$, C — *произвольная положительная константа, тогда равномерно по* a *и* b *из отрезка* $[-C, C]$ *(пусть* $b \geqslant a$*)*

$$\mathbb{P}\left(a \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant b\right) \xrightarrow[n \to +\infty]{} \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}} dx.$$

Определение. Случайная величина ξ имеет *нормальное (гауссовское) распределение* с параметрами a и σ^2 (N_{a,σ^2}), где $a \in \mathbb{R}, \sigma > 0$, если ξ имеет следующую плотность распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Числовые характеристики N_{a,σ^2}

Найдем матожидание и дисперсию для *стандартного* нормального распределения ($\alpha = 0$, $\sigma^2 = 1$):

$$\mathbb{E}\,\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x)dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} xe^{\frac{-x^2}{2}}dx = -\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} d\left(e^{\frac{-x^2}{2}}\right) = -e^{\frac{-x^2}{2}}\Big|_{-\infty}^{\infty} = 0$$

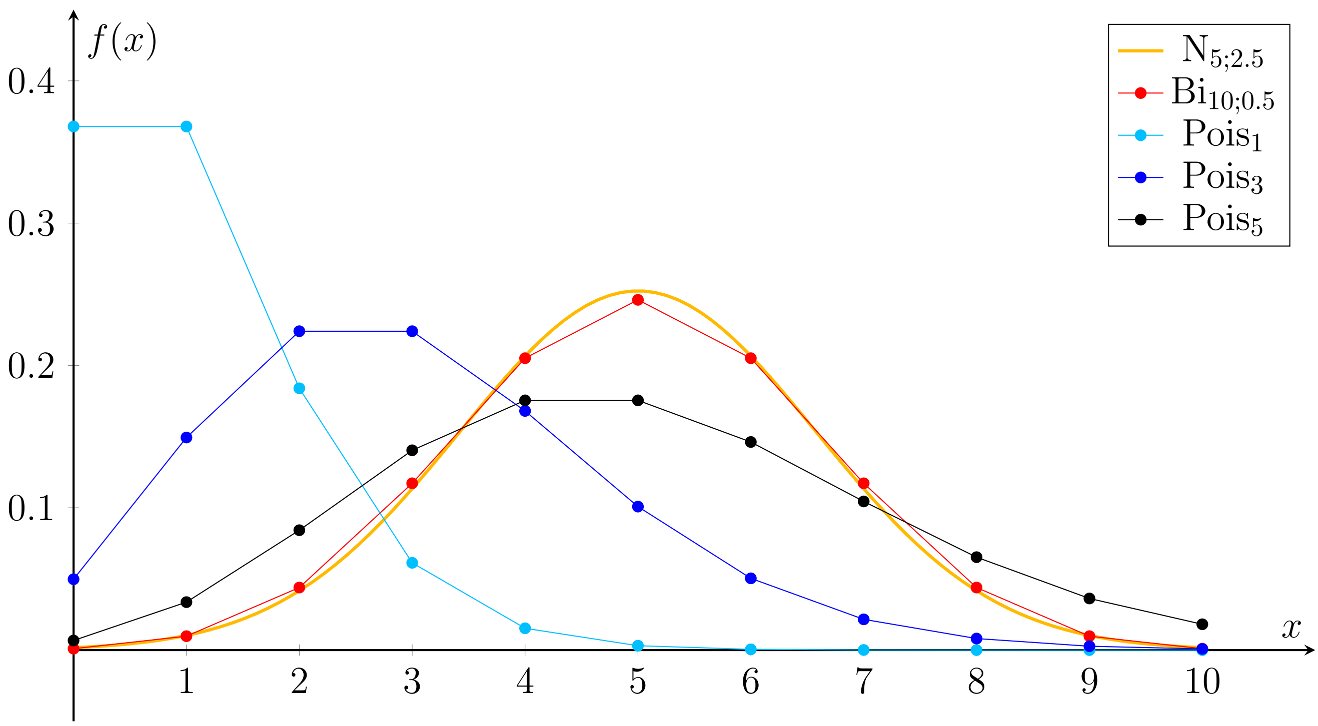
$$\mathbb{E}\,\xi^2 = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x^2e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty} x^2e^{-x^2/2}dx = -\frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty} xde^{-x^2/2} =$$

$$= -\frac{2x}{\sqrt{2\pi}}e^{-x^2/2}\Big|_0^{\infty} + 2\int\limits_0^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = 0 + \int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = 1.$$

Поэтому $\mathbb{D}\,\xi = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2 = 1 - 0 = 1$.

Теперь рассмотрим случайную величину η с нормальным распределением в общем виде (с параметрами α и σ^2). Тогда $\xi = \frac{\eta-\alpha}{\sigma}$ - случайная величина со *стандартным* нормальным распределением. Далее, т.к. $\mathbb{E}\,\xi = 0$, $\mathbb{D}\,\xi = 1$, то

$$\mathbb{E}\,\eta = \mathbb{E}(\sigma\xi + \alpha) = \sigma\,\mathbb{E}\,\xi + \alpha = \alpha,$$
$$\mathbb{D}\,\eta = \mathbb{D}(\sigma\xi + \alpha) = \sigma^2\,\mathbb{D}\,\xi = \sigma^2.$$



Замечание. Формулу из теоремы Муавра—Лапласа можно переписать в следующем виде:

$$\mathbb{P}\left(a \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant b\right) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2}dt$ — функция распределения стандартного нормального распределения.

1.10 Совокупности случайных величин. Совместная функция распределения. Независимость случайных величин. Критерии независимости. Ковариация, коэффициент корреляции

Совместное распределение, его свойства

Пусть случайные величины $\xi_1, \ldots, \xi_n$ заданы на одном вероятностном пространстве $(\Omega, \mathcal{F}\,\mathbb{P})$.

Определение. *Совместное распределение* случайных величин $(\xi_1, \ldots, \xi_n)$ — функция $\mathbb{P}\colon \mathfrak{B}(\mathbb{R}^n) \to \mathbb{R}$:

$$\mathbb{P}(\xi \in B) = \mathbb{P}(\omega\colon (\xi_1(\omega), \ldots, \xi_n(\omega)) \in B),\ B \subset \mathfrak{B}(\mathbb{R}^n).$$

Определение. *Функция совместного распределения* случайных величин $(\xi_1, \ldots, \xi_n)$ — функция $F\colon \mathbb{R}^n \to \mathbb{R}$:

$$F(x_1, \ldots, x_n) = \mathbb{P}(\xi_1 < x_1, \ldots, \xi_n < x_n).$$

Замечание. Для функции совместного распределения выполняются свойства, аналогичные одномерному случаю — сохраняется монотонность, непрерывность слева по каждой переменной. При этом частные функции распределения восстанавливаются по совместной следующим образом:

$$\lim_{\substack{x_k\to+\infty\\ k\neq i}} F(x_1, \ldots, x_i, \ldots, x_n) = F_i(x), \quad i = \overline{1,n}.$$

Следует, однако, обратить внимание на предельные свойства. Для простоты проиллюстрируем их на двумерном случае.

$$\lim_{x_1\to-\infty} F(x_1, x_2) = 0$$

$$\lim_{x_2 \to -\infty} F(x_1, x_2) = 0$$
$$\lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty}} F(x_1, x_2) = 1$$

В самом деле, по определению $F(x_1, x_2) = \mathbb{P}(\xi_1 < x_1, \xi_2 < 2) = \mathbb{P}(B_1 \cap B_2)$, где $B_1 = \{\xi_1 < x_1\}, B_2 = \{\xi_2 < x_2\}$. Если хоть одно из событий «стремится» к пустому множеству (что и происходит при $x_1 \to -\infty$ или $x_2 \to -\infty$), то и пересечение делает то же самое.

Если же мы хотим получить в пересечении Ω (и, следовательно, вероятность 1), то необходимо устремить x_1 и x_2 к $+\infty$ *одновременно.*

Далее рассматриваем совместные распределения двух случайных величин.

Виды многомерных распределений

Определение. Случайные величины ξ_1, ξ_2 имеют дискретное совместное распределение, если существует не более чем счётный набор пар неотрицательных чисел $\{a_i, b_j\}$ такой, что

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbb{P}\left(\xi_1 = a_i, \xi_2 = b_j\right) = 1.$$

Таблицу, на пересечении i-й строки и j-го столбца которой стоит вероятность $\mathbb{P}\left(\xi_1 = a_i, \xi_2 = b_j\right)$, называют *таблицей совместного распределения* случайных величин ξ_1 и ξ_2.

Определение. Случайные величины ξ_1, ξ_2 имеют абсолютно непрерывное совместное распределение, если существует неотрицательная функция $f_{\xi_1, \xi_2}(x, y)$ такая, что для любого борелевского множества $B \in \mathfrak{B}\left(\mathbb{R}^2\right)$ имеет место равенство

$$\mathbb{P}\big((\xi_1, \xi_2) \in B\big) = \iint\limits_B f_{\xi_1, \xi_2}(x, y)\, dxdy.$$

Если такая функция $f_{\xi_1, \xi_2}(x, y)$ существует, она называется *плотностью совместного распределения* случайных величин ξ_1, ξ_2.

Функция совместного распределения в этом случае имеет вид:

$$F(x, y) = \mathbb{P}(\xi_1 < x, \xi_2 < y) = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{y} f_{\xi_1, \xi_2}(u, v)\, dv \right) du.$$

Замечание. Плотность совместного распределения имеет те же свойства, что и плотность распределения одной случайной величины: неотрицательность и нормированность:

$$f(x,y) \geqslant 0 \; \forall x,y \in \mathbb{R}, \quad \iint\limits_{\mathbb{R}^2} f(x,y)\,dxdy = 1.$$

По функции совместного распределения его плотность находится как смешанная частная производная (в точках, где она существует):

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

Замечание. Из существования плотностей ξ_1 и ξ_2 не следует абсолютная непрерывность совместного распределения этих случайных величин. Например, вектор (ξ,ξ) принимает значения только на диагонали в $\mathbb{R}^2$ и уже поэтому не имеет плотности распределения (его распределение сингулярно). Обратное же свойство, как показывает следующая теорема, всегда верно.

Утверждение. *Если случайные величины ξ_1 и ξ_2 имеют абсолютно непрерывное совместное распределение с плотностью $f(x,y)$, то ξ_1 и ξ_2 в отдельности также имеют абсолютно непрерывное распределение с плотностями:*

$$f_{\xi_1}(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy, \quad f_{\xi_2}(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx.$$

Для $n > 2$ плотности случайных величин $\xi_1, \ldots, \xi_n$ находятся по плотности их совместного распределения $f(x_1, \ldots, x_n)$ интегрированием функции f по всем «лишним» координатам.

Доказательство.

$$F_{\xi_1}(x_1) = \lim_{x_2 \to +\infty} F_{\xi_1,\xi_2}(x_1,x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{\infty} f(x,y)dy \right) dx = \int\limits_{-\infty}^{x_1} f_{\xi_1}(x)dx$$

■

Независимость случайных величин

Определение. Случайные величины $\xi_1, \ldots, \xi_n$ называют *независимыми в совокупности*, если для любого набора борелевских множеств $B_1, \ldots, B_n \in \mathfrak{B}(\mathbb{R})$:

$$\mathbb{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathbb{P}(\xi_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(\xi_n \in B_n)$$

Критерий независимости. *Случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности $\iff$ имеет место равенство:*

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n).$$

В частности, в случае дискретного совместного распределения:

$$\mathbb{P}(\xi_1 = a_1, \ldots, \xi_n = a_n) = \mathbb{P}(\xi_1 = a_1) \cdot \ldots \cdot \mathbb{P}(\xi_n = a_n) \quad \forall a_1, \ldots, a_n \in \mathbb{R}.$$

В случае абсолютно непрерывного:

$$f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = f_{\xi_1}(x_1) \cdot \ldots \cdot f_{\xi_n}(x_n).$$

Формула свёртки

Пусть ξ_1, ξ_2 — случайные величины с плотностью совместного распределения $f_{\xi_1,\xi_2}(x_1, x_2)$, задана борелевская функция $g\colon \mathbb{R}^2 \to \mathbb{R}$. Требуется найти функцию распределения (и плотность, если она существует) случайной величины $\eta = g(\xi_1, \xi_2)$.

Лемма. *Пусть $x \in \mathbb{R}$, задана область $D_x \subseteq \mathbb{R}^2$, $D_x = \{(u, v)\colon g(u, v) < x\}$ Тогда случайная величина $\eta = g(\xi_1, \xi_2)$ имеет функцию распределения*

$$F_\eta(x) = \mathbb{P}(g(\xi_1, \xi_2) < x) = \mathbb{P}((\xi_1, \xi_2) \in D_x) = \iint\limits_{D_x} f_{\xi_1,\xi_2}(u, v)\,du\,dv.$$

Далее считаем, что случайные величины ξ_1 и ξ_2 независимы, т. е. $f_{\xi_1,\xi_2}(u, v) \equiv f_{\xi_1}(u) f_{\xi_2}(v)$. В этом случае распределение величины $g(\xi_1, \xi_2)$ полностью определяется частными распределениями величин ξ_1 и ξ_2.

Формула свёртки. *Если случайные величины ξ_1 и ξ_2 независимы и имеют абсолютно непрерывные распределения с плотностями $f_{\xi_1}(u)$ и $f_{\xi_2}(v)$, то плотность распределения суммы $\xi_1 + \xi_2$ существует и равна «свёртке» плотностей f_{ξ_1} и f_{ξ_2}:*

$$f_{\xi_1+\xi_2}(t) = \int\limits_{-\infty}^{\infty} f_{\xi_1}(u) f_{\xi_2}(t - u)\,du = \int\limits_{-\infty}^{\infty} f_{\xi_2}(u) f_{\xi_1}(t - u)\,du.$$

Доказательство. Воспользуемся утверждением вышеуказанной леммы для борелевской функции $g(u, v) = u + v$. Интегрирование по двумерной области $D_x = \{(u, v)\colon u + v < x\}$ можно заменить последовательным вычислением двух интегралов: наружного — по переменной u, меняющейся в пределах от

$-\infty$ до $+\infty$, и внутреннего — по переменной v, которая при каждом u должна быть меньше, чем $x-u$. Поэтому

$$F_{\xi_1+\xi_2}(x) = \iint\limits_{D_x} f_{\xi_1}(u)f_{\xi_2}(v)dvdu = \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{x-u} f_{\xi_1}(u)f_{\xi_2}(v)dv \right) du.$$

Сделаем в последнем интеграле замену $v = t-u$. При этом $v \in (-\infty, x-u) \iff t \in (-\infty, x), dv = dt$. В полученном интеграле меняем порядок интегрирования:

$$F_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{x} f_{\xi_1}(u)f_{\xi_2}(t-u)dtdu = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{\infty} f_{\xi_1}(u)f_{\xi_2}(t-u)du \right) dt.$$

Из функции распределения $F_{\xi_1+\xi_2}(x)$ выражается плотность $f_{\xi_1+\xi_2}(t)$. ■

Ковариация, коэффициент корреляции, их свойства

Определение. Рассмотрим случайные величины ξ и η. Дисперсия их суммы в общем случае равна $\mathbb{D}(\xi+n) = \mathbb{D}\,\xi + \mathbb{D}\,\eta + 2\big(\mathbb{E}(\xi\eta) - \mathbb{E}\,\xi\;\mathbb{E}\,\eta\big)$. Величина $\operatorname{cov}(\xi,\eta) \equiv \mathbb{E}(\xi\eta) - \mathbb{E}\,\xi\;\mathbb{E}\,\eta$ называется *ковариацией* случайных величин ξ и η. Если ξ и η независимы, то $\operatorname{cov}(\xi,\eta) = 0$. Обратное, вообще говоря, неверно.

Пример. Рассмотрим $\xi \sim \mathrm{U}_{[-\pi;\pi]}$, случайные величины $\eta_1 = \cos\xi$ и $\eta_2 = \sin\xi$.

1. Докажем некореллированность данных случайных величин.

$$\mathbb{E}\,\eta_1 = \int\limits_{-\pi}^{\pi} \cos x \cdot \frac{1}{2\pi}dx = 0, \quad \mathbb{E}\,\eta_2 = \int\limits_{-\pi}^{\pi} \sin x \cdot \frac{1}{2\pi}dx = 0$$

$$\mathbb{E}\,\eta_1\eta_2 = \int\limits_{-\pi}^{\pi} (\cos x \sin x)\frac{1}{2\pi}dx = \frac{1}{4\pi}\int\limits_{-\pi}^{\pi} \sin 2x dx = 0$$

   Следовательно, $\operatorname{cov}(\eta_1,\eta_2) = 0$.

2. Докажем зависимость η_1 и η_2. Рассмотрим события:

$$A = \left\{\omega\colon \eta_1(\omega) \in \left[0, \frac{1}{2}\right]\right\}, \quad B = \left\{\omega\colon \eta_2(\omega) \in \left[0, \frac{1}{2}\right]\right\},$$

Проверим по критерию независимости:

$$\mathbb{P}\left\{\eta_1 \in \left[0, \frac{1}{2}\right]\right\} = \mathbb{P}\left\{\xi \in \left[-\frac{\pi}{2}, -\frac{\pi}{3}\right] \cup \left[\frac{\pi}{3}, \frac{\pi}{2}\right]\right\} = \frac{1}{2\pi} \cdot 2 \cdot \frac{\pi}{6} = \frac{1}{6}$$
$$\mathbb{P}\left\{\eta_2 \in \left[0, \frac{1}{2}\right]\right\} = \mathbb{P}\left\{\xi \in \left[0, \frac{\pi}{6}\right] \cup \left[\frac{5\pi}{6}, \pi\right]\right\} = \frac{1}{2\pi} \cdot 2 \cdot \frac{\pi}{6} = \frac{1}{6}$$
$$\mathbb{P}\left\{\eta_1 \in \left[0, \frac{1}{2}\right], \eta_2 \in \left[0, \frac{1}{2}\right]\right\} = \mathbb{P}\{\varnothing\} = 0 \neq \frac{1}{6} \cdot \frac{1}{6}$$

Следовательно, η_1 и η_2 зависимы.

Свойства ковариации.

1. $\operatorname{cov}(\xi, \xi) = \mathbb{D}\,\xi$;
2. $\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\eta, \xi)$;
3. $\forall a, b \in \mathbb{R}\ \operatorname{cov}(a\xi + b, \eta) = a \operatorname{cov}(\xi, \eta)$;
4. $\operatorname{cov}(\eta + \zeta, \xi) = \operatorname{cov}(\eta, \xi) + \operatorname{cov}(\zeta, \xi)$;
5. $\operatorname{cov}^2(\xi, \eta) \leqslant \mathbb{D}\,\xi\, \mathbb{D}\,\eta$,
   $\operatorname{cov}^2(\xi, \eta) = \mathbb{D}\,\xi\, \mathbb{D}\,\eta \iff \xi \overset{\text{п.н.}}{=} a\eta + b,\ a, b \in \mathbb{R}$ *(аналог неравенства Коши-Буняковского);*

Величина ковариации характеризует меру линейной зависимости случайных величин. Однако от умножения на константу (не равную нулю) зависимость случайных величин не изменяется, в отличие от ковариации. Введём новый термин.

Определение. *Коэффициент корреляции* $\rho(\xi, \eta)$ случайных величин ξ и η, дисперсии которых существуют и отличны от нуля:

$$\rho(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathbb{D}\,\xi}\sqrt{\mathbb{D}\,\eta}}.$$

Свойства коэффициента корреляции.

1. *Коэффициент корреляции независимых случайных величин равен нулю;*
2. *Для любых двух случайных величин (для которых выполнены условия определения) их коэффициент корреляции по модулю не превосходит единицы;*

3. *Если* $\left|\rho(X,Y)\right| = 1$*, то с вероятностью один* X *и* Y *линейно выражаются друг через друга:*

$$\left|\rho(X,Y)\right| = 1 \Longrightarrow \exists a \neq 0,\ b \in \mathbb{R}\colon\ \mathbb{P}(X - aY = b) = 1.$$

При этом знак коэффициента a *совпадает со знаком коэффициента корреляции.*

Доказательство.

1. В числителе дроби, которой равен коэффициент корреляции, окажется ноль. В знаменателе нуля быть не должно, это обеспечивается определением.

2. Обозначим эти две случайные величины как ξ и η и центрируем: $\xi_c = \xi - \mathbb{E}\,\xi$ и $\eta_c = \eta - \mathbb{E}\,\eta$. Так как $\operatorname{cov}(\xi,\eta) = \operatorname{cov}(\xi_c,\eta_c)$, а дисперсия случайной величины не меняется от смещения случайной величины на константу, коэффициент корреляции не изменится.

   Далее, т.к. $\mathbb{E}\,\xi_c = \mathbb{E}\,\eta_c = 0$:

$$\mathbb{D}\,\xi_c = \mathbb{E}\,\xi_c^2 - \left(\mathbb{E}\,\xi_c\right)^2 = \mathbb{E}\,\xi_c^2,\ \mathbb{D}\,\eta_c = \mathbb{E}\,\eta_c^2;$$
$$\operatorname{cov}\left(\xi_c,\eta_c\right) = \mathbb{E}\left(\xi_c\,\eta_c\right) - \mathbb{E}\,\xi_c\ \mathbb{E}\,\eta_c = \mathbb{E}\left(\xi_c\,\eta_c\right).$$

   Далее идут те же рассуждения, что часто используются при доказательстве неравенства Коши-Буняковского:

$$\forall a \in \mathbb{R}\quad 0 \leqslant \mathbb{D}\left(\xi_c - a\eta_c\right) = \mathbb{E}\left(\xi_c - a\eta_c\right)^2 - \Big(\mathbb{E}\left(\xi_c - a\eta_c\right)\Big)^2 = \mathbb{E}\left(\xi_c - a\eta_c\right)^2.$$

   Полученное неравенство можно рассматривать как квадратное неравенство относительно a, а именно

$$\mathbb{E}\left(\xi_c - a\eta_c\right)^2 = \mathbb{E}\,\xi_c^2 - 2a\,\mathbb{E}\left(\xi_c\eta_c\right) + a^2\,\mathbb{E}\,\eta_c^2 \geqslant 0.$$

   Поскольку верно это для любого a, то дискриминанту нельзя быть больше нуля. То есть:

$$\Big(\mathbb{E}\left(\xi_c\eta_c\right)\Big)^2 - \mathbb{E}\,\xi_c^2\ \mathbb{E}\,\eta_c^2 \leqslant 0 \Longleftrightarrow \left|\mathbb{E}\left(\xi_c\eta_c\right)\right| \leqslant \sqrt{\mathbb{E}\,\xi_c^2\ \mathbb{E}\,\eta_c^2} \Longrightarrow$$
$$\Longrightarrow \left|\operatorname{cov}\left(\xi_c,\eta_c\right)\right| \leqslant \sqrt{\mathbb{D}\,\xi_c\ \mathbb{D}\,\eta_c}.$$

   По доказанному выше «стирание» индексов не изменит коэффициентов.

3. Доказательство этого свойства целиком опирается на доказательство предыдущего: если выполнилось равенство $|\operatorname{cov}(\xi,\eta)| = \sqrt{\mathbb{D}\,\xi\ \mathbb{D}\,\eta}$, то квадратное неравенство относительно a обратилось в равенство. Но это равенство означает, что равна нулю $\mathbb{D}(\xi - a\eta)$, а это сразу говорит о том, что с вероятностью один $\xi - a\eta$ равна константе. Обозначим эту константу за b и получим то, что нужно было доказать.

   Знак коэффициента корреляции совпадает с знаком ковариации, так дисперсии по предположению положительны. Выразив ξ через η, мы можем воспользоваться свойствами ковариации и получить

$$\rho(\xi,\eta) = \rho(a\eta+b,\eta) = \frac{\operatorname{cov}(a\eta+b,\eta)}{\sqrt{\mathbb{D}(a\eta+b)\ \mathbb{D}\,\eta}} = \frac{a\operatorname{cov}(\eta,\eta)}{\sqrt{a^2\ \mathbb{D}\,\eta\ \mathbb{D}\,\eta}} = \frac{a\,\mathbb{D}\,\eta}{|a|\,\mathbb{D}\,\eta} = \operatorname{sign}(a).$$

■

1.11 Неравенства Маркова, Чебышёва и Гаусса. Правило «трех сигм». Закон больших чисел в форме Чебышёва

Неравенство Маркова. *Если* $\mathbb{E}\,|\xi| < \infty$, *то для любого* $x > 0$

$$\mathbb{P}\big(|\xi| \geqslant x\big) \leqslant \frac{\mathbb{E}\,|\xi|}{x}.$$

Доказательство. $\mathrm{I}(A) \sim \mathrm{B}_p,\ p = \mathbb{P}(\mathrm{I}(A)=1) = \mathbb{P}(A) = \mathbb{E}\ \mathrm{I}(A)$.

Индикаторы прямого и противоположного событий связаны равенством $\mathrm{I}(A) + \mathrm{I}(\overline{A}) = 1$, поэтому

$$|\xi| = |\xi|\cdot\mathrm{I}\big(|\xi| < x\big) + |\xi|\cdot\mathrm{I}\big(|\xi| \geqslant x\big) \geqslant |\xi|\cdot\mathrm{I}\big(|\xi| \geqslant x\big) \geqslant x\cdot\mathrm{I}\big(|\xi| \geqslant x\big).$$

Тогда $\mathbb{E}\,|\xi| \geqslant \mathbb{E}\big(x\cdot\mathrm{I}(|\xi| \geqslant x)\big) = x\cdot\mathbb{P}\big(|\xi| \geqslant x\big)$. Осталось разделить обе части этого неравенства на положительное число x. ■

Неравенство Чебышёва. *Если* $\mathbb{D}\,\xi$ *существует, то для любого* $\varepsilon > 0$

$$\mathbb{P}(|\xi - \mathbb{E}\,\xi| \geqslant \varepsilon) \leqslant \frac{\mathbb{D}\,\xi}{\varepsilon^2}.$$

Доказательство. Для $\varepsilon > 0$ неравенство $|\xi - \mathbb{E}\,\xi| \geqslant \varepsilon \iff (\xi - \mathbb{E}\,\xi)^2 \geqslant \varepsilon^2$, поэтому $\mathbb{P}(|\xi - \mathbb{E}\,\xi| \geqslant \varepsilon) = \mathbb{P}\left((\xi - \mathbb{E}\,\xi)^2 \geqslant \varepsilon^2\right) \leqslant \dfrac{\mathbb{E}(\xi - \mathbb{E}\,\xi)^2}{\varepsilon^2} = \dfrac{\mathbb{D}\,\xi}{\varepsilon^2}$. ■

Определение. В неравенстве Чебышёва в качестве ε можно брать любое положительное число. Если взять в качестве ε величину 3σ, где σ — стандартное отклонение, то получим

$$\mathbb{P}\big(|\xi - \mathbb{E}\,\xi| > 3\sigma\big) \leqslant \frac{\mathbb{D}\,\xi}{9\sigma^2} = \frac{\mathbb{D}\,\xi}{9\,\mathbb{D}\,\xi} = \frac{1}{9} \iff \mathbb{P}\big(|\xi - \mathbb{E}\,\xi| \leqslant 3\sigma\big) \geqslant 1 - \frac{1}{9} = \frac{8}{9}.$$

Это соотношение называется *правилом трёх сигм.*

Неравенство Гаусса. *Пусть X — одномодальная случайная величина с модой m, a^2 — математическое ожидание $(X-m)^2$. Тогда*

$$\mathbb{P}(|X-m| > k) \leq \begin{cases} \left(\dfrac{2a}{3k}\right)^2, & \text{если } k \geqslant \dfrac{2a}{\sqrt{3}}; \\ 1 - \dfrac{k}{a\sqrt{3}}, & \text{если } 0 \leqslant k \leqslant \dfrac{2a}{\sqrt{3}}. \end{cases}$$

Определение. Говорят, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ с конечными первыми моментами *удовлетворяет закону больших чисел*, если

$$\frac{\Big(\xi_1 + \ldots + \xi_n\Big) - \Big(\mathbb{E}\,\xi_1 + \ldots + \mathbb{E}\,\xi_n\Big)}{n} \xrightarrow[n\to+\infty]{\mathrm{P}} 0$$

Закон больших чисел в форме Чебышёва. *Для любой последовательности $\xi_1, \xi_2, \ldots$ попарно независимых и одинаково распределённых случайных величин с конечным вторым моментом $\mathbb{E}\,\xi_1^2 < \infty$ имеет место сходимость*

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{p} \mathbb{E}\,\xi_1.$$

Доказательство. Обозначим через $S_n = \xi_1 + \ldots + \xi_n$ сумму первых n случайных величин. Из линейности матожидания получим

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{\mathbb{E}\,\xi_1 + \ldots + \mathbb{E}\,\xi_n}{n} = \frac{n\,\mathbb{E}\,\xi_1}{n} = \mathbb{E}\,\xi_1.$$

Пусть $\varepsilon > 0$. Воспользуемся неравенством Чебышёва:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathbb{D}\,S_n}{n^2\varepsilon^2} = \frac{\mathbb{D}\,\xi_1 + \ldots + \mathbb{D}\,\xi_n}{n^2\varepsilon^2} =$$
$$= \frac{n\,\mathbb{D}\,\xi_1}{n^2\varepsilon^2} = \frac{\mathbb{D}\,\xi_1}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0,$$

так как $\mathbb{D}\,\xi_1 < \infty$. Дисперсия суммы превратилась в сумму дисперсий в силу попарной независимости слагаемых, из-за которой все ковариации $\mathrm{cov}(\xi_i, \xi_j)$ по свойству ковариации обратились в нуль при $i \neq j$. ■

Замечание. Условие попарной независимости является избыточным — достаточно равенства нулю коварций, т.е. некоррелированности случайных величин. Интересно, что даже это условие можно ослабить и потребовать только неотрицательности ковариаций. Доказательство этого факта оставим читателю в качестве упражнения.

1.12 Виды сходимости последовательностей случайных величин

Пусть случайные величины ξ, $\xi_1, \ldots, \xi_n, \ldots$ определены на одном вероятностном пространстве $(\Omega, \mathcal{F}\,\mathbb{P})$.

Определение. Последовательность случайных величин $\{\xi_n\}_{n=1}^{+\infty}$ *почти наверное сходится* к случайной величине ξ ($\xi_n \xrightarrow{\text{п.н.}} \xi$), если

$$\mathbb{P}\left(\left\{\omega\colon \lim_{n\to\infty} \xi_n(w) = \xi(w)\right\}\right) = 1.$$

Определение. Последовательность случайных величин $\{\xi_n\}_{n=1}^{+\infty}$ *сходится по вероятности* к случайной величине ξ ($\xi_n \xrightarrow{\mathrm{P}} \xi$), если

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\left\{\omega\colon |\xi_n(\omega) - \xi(\omega)| > \varepsilon\right\}\right) \xrightarrow[n\to+\infty]{} 0.$$

Определение. Последовательность случайных величин $\{\xi_n\}_{n=1}^{+\infty}$ *сходится в среднем* к случайной величине ξ ($\xi_n \xrightarrow{(\mathrm{r})} \xi$ или $\xi_n \xrightarrow{L_p} \xi$), если

$$\mathbb{E}\,|\xi_n - \xi|^r \xrightarrow[n\to+\infty]{} 0, \quad r \geqslant 1.$$

В последующих опеределениях случайные величины могут принадлежать разным вероятностным пространствам.

Определение. Последовательность случайных величин $\{\xi_n\}_{n=1}^{+\infty}$ *сходится по распределению* к случайной величине ξ ($\xi_n \xrightarrow{\mathrm{d}} \xi$), если

$$F_{\xi_n}(x) \xrightarrow[n\to+\infty]{} F_\xi(x) \quad \forall x, \text{ в которых } F_\xi \text{ непрерывна.}$$

Определение. Последовательность случайных величин $\{\xi_n\}_{n=1}^{+\infty}$ *слабо сходится* к случайной величине ξ ($\xi_n \xRightarrow{\mathrm{w}} \xi$), если

$$\mathbb{E}\, f\,(\xi_n) \to \mathbb{E}\, f(\xi) \quad \forall \text{ непрерывной ограниченной } f(x).$$

Утверждение. *Вышеуказанные виды сходимости последовательностей случайных величин связаны следующими отношениями:*

$$\begin{tikzcd}
\text{п.н.} \arrow[dr, Rightarrow] & & & \\
& \mathrm{p} \arrow[r, Rightarrow] & \mathrm{d} \arrow[r, Leftrightarrow] & \mathrm{w} \\
(\mathrm{r}) \arrow[ur, Rightarrow] & & &
\end{tikzcd}$$

Доказательство.
(r) $\Longrightarrow$ p Используем неравенство Маркова:

$$\mathbb{P}\big(|\xi_n - \xi| \geqslant \varepsilon\big) = \mathbb{P}\big(|\xi_n - \xi|^r \geqslant \varepsilon^r\big) \leqslant \frac{\mathbb{E}\,|\xi_n - \xi|^r}{\varepsilon^r} \xrightarrow[n\to+\infty]{} 0.$$

(r) $\not\Longleftarrow$ p Рассмотрим последовательность случайных величин:

$$\xi_n = \begin{cases} 0, & \frac{1}{n} \leqslant \omega \leqslant 1; \\ \sqrt[r]{n}, & 0 \leqslant \omega \leqslant \frac{1}{n}. \end{cases} \Longrightarrow p_1 = 1 - \frac{1}{n},\ p_2 = \frac{1}{n}$$
$$\mathbb{P}\big(|\xi_n| > \varepsilon\big) \xrightarrow[n\to+\infty]{} 0, \text{ однако } \mathbb{E}\,|\xi_n|^r = 1.$$

п.н. $\Longrightarrow$ p От противного: допустим, что выполняется определение сходимости почти наверное, но нет сходимости по вероятности. Распишем определение сходимости по вероятности (обратите внимание, что здесь есть ε' и ε — один из определения предела, другой — из определения сходимости):

$$\forall \varepsilon > 0,\ \varepsilon' > 0 \quad \exists N \in \mathbb{N}\colon \forall n \geqslant N \quad \mathbb{P}\Big(\Big\{\omega\colon \big|\xi_n(\omega) - \xi(\omega)\big| > \varepsilon\Big\}\Big) < \varepsilon'.$$

По предположению сходимость по вероятности отсутствует:

$$\exists \varepsilon_0 > 0, \varepsilon_0' > 0\colon\ \forall N \in \mathbb{N}\ \exists n \geqslant N \quad \mathbb{P}\Big(\Big\{\omega\colon \big|\xi_n(\omega) - \xi(\omega)\big| > \varepsilon_0\Big\}\Big) \geqslant \varepsilon_0' > 0.$$

Распишем теперь определение сходимости почти наверное:

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N}\colon \forall n \geqslant N \quad \mathbb{P}\Big(\Big\{\omega\colon \big|\xi_n(\omega) - \xi(\omega)\big| < \varepsilon\Big\}\Big) = 1.$$

Подставив ε_0 и поменяв знак неравенства, получим

$$\mathbb{P}\Bigg(\Big\{\omega\colon \big|\xi_n(\omega)-\xi(\omega)\big|\geqslant\varepsilon_0\Big\}\Bigg)=0.$$

Но это противоречит второму неравенству. Значит, наше предположение неверно, и из сходимости почти наверное следует сходимость по вероятности, что и требовалось доказать.

п.н. $\not\Longleftarrow$ p Рассмотрим вероятностное пространство $([0,1],\mathfrak{B}_{[0,1]},\lambda)$ (λ — мера Лебега). Положим $\xi\equiv 0$, $\xi_{2^k}=\mathrm{I}\big(\big[0,\frac{1}{2^k}\big]\big)$, $\xi_{2^k+p}=\mathrm{I}\big(\big[\frac{p}{2^k},\frac{p+1}{2^k}\big]\big), 1\leqslant p<2^k$. Тогда $\xi_n\xrightarrow{\mathrm{p}}0$, т.к. $\mathbb{P}(\xi_{2^k+p}>0)\leqslant\lambda\big(\big[\frac{p}{2^k},\frac{p+1}{2^k}\big]\big)\leqslant\frac{1}{2^k}\to 0$, но $\xi_n\overset{\text{п.н.}}{\not\to}0$, т.к. для любого ω существует бесконечно много n, таких что $\xi_n(\omega)=1$.

(r) $\not\Longleftarrow$ п.н. Рассмотрим то же вероятностное пространство $([0,1],\mathfrak{B}_{[0,1]},\lambda)$. Определим для $k\geqslant 1$ $\xi_k=2^k\cdot\mathrm{I}\big(\big[0,\frac{1}{2^k}\big]\big)$. Тогда $\forall k\quad \mathbb{E}\,\xi_k=1$, но $\xi=\mathrm{I}(\omega=0)$.

p $\Longrightarrow$ w Пусть f - ограниченная и непрерывная функция, $|f|\leqslant C$. Зафиксируем $\varepsilon>0$. $\exists N,\exists\delta$:

1. $\mathbb{P}(|\xi|>N)\leqslant\frac{\varepsilon}{6C}$, (это возможно, т.к. $\mathbb{P}(\xi>x)\xrightarrow[x\to+\infty]{}0$)
2. $\mathbb{P}(|\xi_n-\xi|>\delta)\leqslant\frac{\varepsilon}{6C}\quad\forall n\geqslant N$ (из сходимости по вероятности)
3. $\forall x,y\colon |x|<N, |x-y|<\delta\implies |f(x)-f(y)|\leqslant\frac{\varepsilon}{3}$, т.к. f по теореме Кантора равномерно непрерывна на отрезке $[-N,N]$.

Рассмотрим следующие события:

$$\begin{gathered}A_1=\{|\xi_n-\xi|\leqslant\delta\}\cap\{|\xi|<N\},\\ A_2=\{|\xi_n-\xi|\leqslant\delta\}\cap\{|\xi|\geqslant N\},\\ A_3=\{|\xi_n-\xi|>\delta\}.\end{gathered}$$

Эти события образуют разбиение $\Omega=A_1\cup A_2\cup A_3$.
Оценим $|\,\mathbb{E}\,f(\xi_n)-\mathbb{E}\,f(\xi)|$:

$$\begin{aligned}\big|\mathbb{E}\,f\,(\xi_n)-\mathbb{E}\,f(\xi)\big|&=\big|\mathbb{E}\big(f\,(\xi_n)-f(\xi)\big)\big|\leqslant\mathbb{E}\big|f\,(\xi_n)-f(\xi)\big|=\\ &=\mathbb{E}\Big[\big|f\,(\xi_n)-f(\xi)\big|\cdot(\mathrm{I}(A_1)+\mathrm{I}(A_2)+\mathrm{I}(A_3))\Big]\leq\\ &\leqslant\frac{\varepsilon}{3}\,\mathbb{P}\,(A_1)+2C\big(\mathbb{P}\,(A_2)+\mathbb{P}\,(A_3)\big)\leqslant\frac{\varepsilon}{3}+2C\left(\frac{\varepsilon}{6C}+\frac{\varepsilon}{6C}\right)=\varepsilon,\end{aligned}$$

откуда следует, что $\big|\mathbb{E}\,f(\xi_n)-\mathbb{E}\,f(\xi)\big|\to 0\implies\xi_n\xrightarrow{\mathrm{w}}\xi$.

p $\not\Longleftarrow$ d Пусть $\xi_n=\begin{cases}1,\ p_1=\frac{1}{2}\\ 0,\ p_0=\frac{1}{2}\end{cases}$, $\xi=\begin{cases}0,\ p_1=\frac{1}{2}\\ 1,\ p_0=\frac{1}{2}\end{cases}$.

Тогда $|\xi_n - \xi| = \begin{cases} 1, \ p_1 = \frac{1}{2} \\ 0, \ p_0 = \frac{1}{2} \end{cases}$, и не выполняется определение сходимости по вероятности, например, при $\varepsilon_0 = \frac{1}{3}$, т.к.

$$\mathbb{P}\left(|\xi_n - \xi| > \frac{1}{3}\right) = 1 \nrightarrow 0 \text{ при } n \to \infty.$$

p $\not\Leftarrow$ w Пусть $\Omega = \{\omega_1, \omega_2\}, \mathbb{P}\big(\{\omega_1\}\big) = \mathbb{P}\big(\{\omega_2\}\big) = \frac{1}{2}$.
Определим для любого n $\xi_n(\omega_1) = 1, \xi_n(\omega_2) = -1$. Положим $\xi = -\xi_n$. Тогда:

$$\mathbb{E}\, f(\xi_n) = \frac{f(1) + f(-1)}{2} = \mathbb{E}\, f(\xi),$$

но $\forall n \quad |\xi_n - \xi| = 2 \implies \xi_n \overset{\text{p}}{\nrightarrow} \xi$.

■

Замечание. Слабая сходимость всё же не есть сходимость случайных величин, и ею нельзя оперировать как сходимостями п.н. и по вероятности, для которых предельная случайная величина единственна (с точностью до значений на множестве нулевой вероятности).

Замечание. Во многих источниках слабая сходимость и сходимость по распределению вводятся как один и тот же вид сходимости. Те немногие доказательства эквивалентности, которые удалось найти авторам, слишком техничны и предоставляются читателю.

1.13 Характеристические функции и их свойства

Определение. *Характеристическая функция случайной величины* ξ — функция $\varphi_\xi \colon \mathbb{R} \to \mathbb{C}$:

$$\varphi_\xi(t) = \mathbb{E}\, e^{it\xi} = \mathbb{E}\cos(t\xi) + i\,\mathbb{E}\sin(t\xi) = \int\limits_{\mathbb{R}} e^{itx} dF_\xi(x),$$

где интеграл справа называется *интегралом Фурье-Стильтъеса.*

Для абсолютно непрерывного распределения характеристическая функция имеет вид

$$\varphi_\xi(t) = \int\limits_{\mathbb{R}} e^{itx} f(x)\, dx.$$

Для дискретного, соответственно,

$$\varphi_\xi(t) = \sum_i e^{itx_i}\ \mathbb{P}\,(\xi = x_i).$$

Пример. Характеристическая функция стандартной нормальной случайной величины $\xi \sim \mathrm{N}_{0;1}$:

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{itx}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-t^2/2}e^{-x^2/2+itx+t^2/2}dx =$$

$$= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-t^2/2}e^{-(x-it)^2/2}dx = e^{-t^2/2}\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-(x-it)^2/2}d(x-it) = e^{-t^2/2}.$$

Свойства характеристической функции.

1. *Характеристическая функция существует для любой случайной величины ξ.*

2. $\forall\xi,\ \forall a, b \in \mathbb{R}\colon \varphi_{a\xi+b}(t) = e^{itb}\varphi_\xi(at)$.

3. (a) $\left|\varphi_\xi(t)\right| = \left|\mathbb{E}\,e^{it\xi}\right| \leqslant 1$;

   (b) $\varphi_\xi(0) = 1$;

   (c) $\overline{\varphi_\xi(t)} = \varphi_\xi(-t) = \varphi_{-\xi}(t) \quad \forall t \in \mathbb{R}$.

   Следствие *Если характеристическая функция вещественнозначна, то она является чётной.*

4. *Если случайные величины ξ и η независимы, то $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t)$.*

5. *Характеристическая функция равномерно непрерывна.*

6. *Если существует абсолютный момент k-го порядка $\mathbb{E}\,|\xi|^k < \infty,\ k \geqslant 1$, то существует непрерывная k-я производная характеристической функции:*

$$\left.\frac{\partial^k}{\partial t^k}\varphi_\xi(t)\right|_{t=0} = i^k\ \mathbb{E}\,\xi^k.$$

   Если существует непрерывная производная характеристической функции порядка $k = 2n, n \in \mathbb{N}$, то существует абсолютный момент порядка $k = 2n:\ \mathbb{E}\,|\xi|^k = \mathbb{E}\,\xi^k$ (а следовательно, и все предыдущие) и его можно вычислить по той же формуле.

7. *Характеристическая функция случайно величины ξ однозначно определяет её функцию распределения $F_\xi(x)$. Ряд распределения или плотность восстанавливаются по характеристической функции с помощью преобразования Фурье.*

   Дискретное распределение:

$$\mathbb{P}(\xi = k) = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} e^{-itk} \varphi_\xi(t)\, dt, k \in \mathbb{Z}.$$

   Абсолютно непрерывное распределение:

$$f_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} \varphi_\xi(t)\, dt, \quad x \in \mathbb{R}.$$

8. $\xi_n \xrightarrow[n\to\infty]{d} \xi \iff \varphi_{\xi_n}(t) \xrightarrow[n\to\infty]{} \varphi_\xi(t)$ *(теорема Леви о непрерывном соответствии).*

Доказательство.

1. Существование характеристической функции равносильно равномерной сходимости соответствующего интеграла. Докажем её по признаку Вейерштрасса:

$$|\varphi_\xi(t)| = \left| \int\limits_{\mathbb{R}} e^{itx} dF(x) \right| \leqslant \int\limits_{\mathbb{R}} \left| e^{itx} \right| dF(x) = \int\limits_{\mathbb{R}} dF(x) = 1.$$

2. $\varphi_{a\xi+b}(t) = \mathbb{E}\, e^{it(a\xi+b)} = e^{itb}\, \mathbb{E}\, e^{ita\xi} = e^{itb} \varphi_\xi(at).$

3. Неравенство доказано в пункте 1, равенство (b) очевидно.

$$\varphi_\xi(-t) = \mathbb{E}\, cos(-t\xi) + i\, \mathbb{E}\, sin(-t\xi) = \mathbb{E}\, cos(t\xi) - i\, \mathbb{E}\, sin(t\xi) = \overline{\varphi_\xi(t)}.$$

   Оставшиеся равенства следуют из второго свойства.

4. $\varphi_{\xi+\eta}(t) = \mathbb{E}\, e^{it(\xi+\eta)} = \{\text{независимость}\} = \mathbb{E}\, e^{it\xi}\, \mathbb{E}\, e^{it\eta} = \varphi_\xi(t)\varphi_\eta(t).$

5. Выберем сколь угодно малое $\varepsilon > 0$ и оценим разность значений характеристической функции в точках t и $t + h$:

$$\left|\varphi(t+h)-\varphi(t)\right| = \left|\int\limits_{\mathbb{R}} \left(e^{i(t+h)x} - e^{itx}\right) dF(x)\right| = \left|\int\limits_{\mathbb{R}} e^{itx}\left(e^{ihx}-1\right) dF(x)\right| \leqslant$$

$$\leqslant \int\limits_{\mathbb{R}} \left|e^{ihx}-1\right| dF(x) = \int\limits_{|x|\leqslant R} \left|e^{ihx}-1\right| dF(x) + \int\limits_{|x|>R} \left|e^{ihx}-1\right| dF(x)$$

Теперь выберем R настолько большим, чтобы $\mathbb{P}\big(|x| > R\big) < \frac{\varepsilon}{4}$. Поскольку $\left|e^{ihx}-1\right| \leqslant 2$, второй интеграл при этом не превосходит по величине $\frac{\varepsilon}{2}$. После этого выберем h столь малым, чтобы $\left|e^{ihx}-1\right| < \frac{\varepsilon}{2}$ при всех $|x| \leqslant R$. Тогда и первый интеграл не превосходит $\frac{\varepsilon}{2}$ и, таким образом, по заданному $\varepsilon > 0$ подобрано столь малое $h > 0$, что $|\varphi(t+h)-\varphi(t)| < \varepsilon \ \forall t \in \mathbb{R}$.

6. Если существует $\mathbb{E}\,\xi^k < \infty,\ k \geqslant 1$, то для всех $m = \overline{1,k}$ существуют $\mathbb{E}\,\xi^m < \infty$. Следовательно,

$$\left|\int\limits_{\mathbb{R}} (ix)^m e^{itx} dF(x)\right| \leqslant \int\limits_{\mathbb{R}} |x|^k dF(x) = \mathbb{E}\,|\xi|^m < \infty \quad \forall m = \overline{1,k}.$$

Т.е. интегралы $\int\limits_{\mathbb{R}} (ix)^m e^{itx} dF(x)$ сходятся равномерно по t, а значит, дифференцирование по t можно менять местами с операцией интегрирования, откуда

$$\varphi_\xi^{(m)}(t) = i^m \int\limits_{\mathbb{R}} x^m e^{itx} dF(x),\ \varphi_\xi^{(m)}(0) = i^m \int\limits_{\mathbb{R}} x^m dF(x) = i^m\,\mathbb{E}\,\xi^m.$$

Пусть у характеристической функции существует непрерывная производная чётного порядка k. Характеристическая функция и её производные непрерывны, функция e^{itx} бесконечно (а значит, и нужные нам k раз) дифференцируема по t, и можно показать, что при этих условиях можно поменять знаки интегрирования и дифференцирования местами. Тогда мы получим, что

$$\varphi^{(k)}(0) = i^k\,\mathbb{E}\,\xi^k e^{itx}\Big|_{t=0} = i^k\,\mathbb{E}\,\xi^k = \int\limits_{\mathbb{R}} x^k dF(x).$$

$x^k \geqslant 0$ в силу чётности k, а значит, указанный интеграл сходится абсолютно, что и означает существование искомого математического ожидания.

$\blacksquare$

1.14 Закон больших чисел в форме Хинчина

Закон больших чисел в форме Хинчина.

Для любой последовательности $\xi_1, \xi_2, \ldots$ *независимых и одинаково распределённых случайных величин с конечным первым моментом* $E\,|\xi_1| < \infty$[1] *имеет место сходимость:*

$$\frac{S_n}{n} = \frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{p} \mathbb{E}\,\xi_1.$$

Для доказательства теоремы нам потребуется следующая лемма.

Лемма. *Если* $\xi_n \overset{w}{\Longrightarrow} c = const$*, то* $\xi_n \xrightarrow{p} c$*.*

Доказательство. Пусть $\xi_n \overset{\mathrm{w}}{\Longrightarrow} c$, т.е.

$$F_{\xi_n}(x) \to F_c(x) = \begin{cases} 0, & x \leqslant c; \\ 1, & x > c. \end{cases}$$

при любом x, являющемся точкой непрерывности предельной функции $F_c(x)$, т. е. $\forall x \neq c$.

Возьмём произвольное $\varepsilon > 0$ и докажем, что $\mathbb{P}\left(|\xi_n - c| < \varepsilon\right) \to 1$:

$$\mathbb{P}\left(-\varepsilon < \xi_n - c < \varepsilon\right) = \mathbb{P}\left(c - \varepsilon < \xi_n < c + \varepsilon\right) \geqslant \mathbb{P}\left(c - \varepsilon/2 \leqslant \xi_n < c + \varepsilon\right) = \\ = F_{\xi_n}(c + \varepsilon) - F_{\xi_n}(c - \varepsilon/2) \to F_c(c + \varepsilon) - F_c(c - \varepsilon/2) = 1 - 0 = 1$$

поскольку в точках $c + \varepsilon$ и $c - \varepsilon/2$ функция F_c непрерывна, и, следовательно, имеет место сходимость последовательностей $F_{\xi_n}(c + \varepsilon)$ к $F_c(c + \varepsilon) = 1$ и $F_{\xi_n}(c - \varepsilon/2)$ к $F_c(c - \varepsilon/2) = 0$.

Осталось заметить, что $\mathbb{P}\left(|\xi_n - c| < \varepsilon\right)$ не бывает больше 1, так что по свойству предела зажатой последовательности $\mathbb{P}\left(|\xi_n - c| < \varepsilon\right) \to 1$. ■

Перейдём к доказательству теоремы.

Доказательство. По вышеприведённому свойству сходимость по вероятности к постоянной эквивалентна слабой сходимости. Так как a — постоянная, достаточно доказать слабую сходимость $\frac{S_n}{n}$ к a. По теореме о непрерывном соответствии, эта сходимость имеет место тогда и только тогда, когда для любого $t \in \mathbb{R}$ сходятся характеристические функции

$$\varphi_{S_n/n}(t) \to \varphi_a(t) = \mathbb{E}\,e^{ita} = e^{ita}$$

[1] Т.к. в определении математического ожидания требуется абсолютная сходимость, существование $\mathbb{E}\,\eta$ и $\mathbb{E}\,|\eta|$ равносильно.

Найдём характеристическую функцию случайной величины $\frac{S_n}{n}$. Пользуясь свойствами характеристической функции, получаем

$$\varphi_{S_n/n}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n.$$

Вспомним, что первый момент ξ_1 существует, поэтому мы можем разложить $\varphi_{\xi_1}(t)$ в ряд Тейлора в окрестности нуля:

$$\varphi_{\xi_1}(t) = 1 + it\,\mathbb{E}\,\xi_1 + o(t) = 1 + ita + o(t).$$

В точке $\frac{t}{n}$ соответственно:

$$\varphi_{\xi_1}\left(\frac{t}{n}\right) = 1 + \frac{ita}{n} + o\left(\frac{t}{n}\right);$$
$$\varphi_{S_n/n}(t) = \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{ita}{n} + o\left(\frac{t}{n}\right)\right)^n.$$

При $n \to +\infty$ воспользуемся «замечательным пределом» $\left(1 + \frac{x}{n}\right)^n \to e^x$ и получим:

$$\varphi_{S_n/n}(t) = \left(1 + \frac{ita}{n} + o\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{ita}{n}\right)^n + o\left(\frac{t}{n}\right)(\ldots) \xrightarrow[n\to+\infty]{} e^{ita}.$$

■

1.15 Центральная предельная теорема

Центральная предельная теорема. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных невырожденных*[1] *случайных величин с $\mathbb{E}\,\xi_1^2 < \infty$ и $S_n = \xi_1 + \ldots + \xi_n$. Тогда*

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}\,S_n}{\sqrt{\mathbb{D}\,S_n}} \leqslant x\right) \xrightarrow[n\to+\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du \quad \forall x \in \mathbb{R}$$

Доказательство. Пусть $\mathbb{E}\,\xi_1 = m$, $\mathbb{D}\,\xi_1 = \sigma^2$. Введём $X = \xi_1 - m$ и $\varphi_X(t) = \mathbb{E}\,e^{itX}$. Введём также

$$\varphi_n(t) = \mathbb{E}\,e^{it\frac{S_n - \mathbb{E}\,S_n}{\sqrt{\mathbb{D}\,S_n}}} = \left[\varphi_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

[1] Т.е. их дисперсии отличны от нуля. Важное требование, т.к. нам нужно будет делить на дисперсию.

В силу разложения характеристической функции (при существовании соответствующих моментов)

$$\varphi_X(t) = 1 + it\,\mathbb{E}\,X + \ldots + \frac{(it)^n}{n!}\,\mathbb{E}\,X^n + R_n(t)$$

Учитывая то, что $\mathbb{E}\,X = \mathbb{E}\,[\xi_1 - m] = 0$, при $n = 2$ получим

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + \overline{o}\left(t^2\right), \quad t \to 0$$

Следовательно, для любого $t \in \mathbb{R}$ при $n \to +\infty$

$$\varphi_n(t) = \left[1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + \overline{o}\left(\frac{1}{n}\right)\right]^n \to e^{-\frac{t^2}{2}}$$

Функция $e^{-\frac{t^2}{2}}$ является характеристической функцией $\mathrm{N}_{0,1}$. В силу теоремы о непрерывном соответствии между функциями распределения и характеристическими функциями центральная предельная теорема доказана. ■

1.16 Условное математическое ожидание

Рассмотрим дискретные случайные величины ξ и η на вероятностном пространстве $(\Omega, \mathcal{F}\,\mathbb{P})$. ξ принимает значения $y_1, y_2, \ldots$, η — $x_1, x_2, \ldots$ Вспомним определение условной вероятности: $\mathbb{P}\,(\xi = y_i | \eta = x_j) = \dfrac{\mathbb{P}\,(\xi = y_i,\, \eta = x_j)}{\mathbb{P}\,(\eta = x_j)}$. Ранее мы проверяли, что условная вероятность является вероятностной мерой на σ-алгебре $\mathcal{F}$. Рассмотрим теперь следующую величину:

$$\mathbb{E}\,[\xi | \eta = x_j] = \sum_i y_i\ \mathbb{P}\,(\xi = y_i | \eta = x_j) \equiv f(x_j). \tag{1.2}$$

$f(x)$ — *регрессия* ξ *на* η. В математическом анализе изучаются функциональные зависимости — каждому значению независимой переменной x соответствует одно определённое значение величины $y = f(x)$. В теории вероятностей и статистике изучаются зависимости между случайными величинами — например, ξ и η. В этом случае при принятии величиной η значения x_j случайная величина ξ может принимать разные значения — $y_1, y_2, \ldots$. Логично сопоставить x_j среднее этих значений. Таким образом строится *стохастическая зависимость* $f(x)$. Можно записать (пока — просто формально)

$$\mathbb{E}\big[\xi | \eta\big] = f(\eta).$$

Условное математическое ожидание — не число, а случайная величина.

Заметим теперь, что введённая нами величина обладает следующим свойством:

$$\forall B \in \mathfrak{B} \quad \mathbb{E}\big[\mathbb{E}\,(\xi|\eta)\,\mathrm{I}(\eta \in \mathrm{B})\big] = \mathbb{E}\big[\xi \cdot \mathrm{I}(\eta \in \mathrm{B})\big].$$

В самом деле,

$$\mathbb{E}\big[\mathbb{E}\,(\xi|\eta)\,\mathrm{I}(\eta \in \mathrm{B})\big] =$$

$$\sum_{j\colon x_j \in B} f(x_j)\;\mathbb{P}(\eta = x_j) = \sum_{j\colon x_j \in B}\left(\sum_k y_k\;\mathbb{P}(\xi = y_k\,|\,\eta = x_j)\right)\mathbb{P}(\eta = x_j) =$$

$$\sum_{j\colon x_j \in B}\sum_k y_k\;\mathbb{P}(\xi = y_k,\,\eta = x_j) = \mathbb{E}\big[\xi \cdot \mathrm{I}(\eta \in \mathrm{B})\big].$$

Рассмотрим теперь абсолютно непрерывный случай. Пусть $\xi, \eta \sim p_{\xi,\eta}(x, y)$, где $p_{\xi,\eta}(x, y)$ — плотность совместного распределения.

$$\forall B \in \mathfrak{B}_{\mathbb{R}^2} \quad \mathbb{P}\big((\xi,\eta) \in B\big) = \iint\limits_B p_{\xi,\eta}(x, y)\,dxdy,$$

$$p_\eta = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(x, y)\,dxdy.$$

В это случае некоторая проблема заключается в том, что при абсолютно непрерывном распределении вероятность попасть в конкретную точку равна нулю, и не получится так просто записать условную вероятность:

$$\mathbb{P}\,(\xi < x\,|\,\eta = y) = \frac{\mathbb{P}(\xi < x,\,\eta = y)}{\mathbb{P}(\eta = y)} = \frac{0}{0} = ?$$

Возьмём тогда произвольный $\varepsilon > 0$ и рассмотрим

$$\mathbb{P}(\xi < x\,|\,y \leqslant \eta < y + \varepsilon) = \frac{\mathbb{P}(\xi < x,\,y \leqslant \eta < y + \varepsilon)}{\mathbb{P}(y \leqslant \eta < y + \varepsilon)} =$$

$$\frac{\int\limits_{-\infty}^{x}\int\limits_{y}^{y+\varepsilon} p_{\xi,\eta}(x, y)\,dydx}{\int\limits_{y}^{y+\varepsilon} p_\eta(y)\,dy} = \{\text{теорема о среднем, } \hat{y}, \tilde{y} \in (y, y + \varepsilon)\} =$$

$$\frac{\varepsilon \cdot \int\limits_{-\infty}^{x} p_{\xi,\eta}(x,\hat{y})\,dx}{\varepsilon \cdot p_\eta(\tilde{y})} \xrightarrow[\varepsilon \to 0]{} = \int\limits_{-\infty}^{x} \frac{p_{\xi,\eta}(x,y)}{p_\eta(y)}\,dx \quad \forall x \in \mathbb{R}.$$

Положим по определению

$$p_{\xi|\eta=y}(x) \overset{\text{def}}{=} \frac{p_{\xi,\eta}(x,y)}{p_\eta(y)}.$$

Тогда

$$\boxed{\mathbb{E}\big[\xi|\eta=y\big] = \int\limits_{-\infty}^{+\infty} x\, p_{\xi|\eta=y}(x)\,dx = \int\limits_{-\infty}^{+\infty} x\,\frac{p_{\xi,\eta}(x,y)}{p_\eta(y)}\,dx \equiv f(y),} \qquad (1.3)$$

$$\mathbb{E}\big[\xi|\eta\big] = f(\eta).$$

Условное матожидание $\mathbb{E}\big(\xi|\eta\big)$ *— это (борелевская) функция от* η*.*

Заметим, что эта величина также удовлетворяет свойству

$$\forall B \in \mathfrak{B} \quad \mathbb{E}\big[\mathbb{E}\,(\xi|\eta)\,\mathrm{I}(\eta \in \mathrm{B})\big] = \mathbb{E}\big[\xi \cdot \mathrm{I}(\eta \in \mathrm{B})\big],$$

Так как

$$\mathbb{E}\left[\mathbb{E}\,(\xi|\eta)\,\mathrm{I}(\eta \in \mathrm{B})\right] = \mathbb{E}\left[f(\eta)\mathrm{I}(\eta \in \mathrm{B})\right] = \int\limits_B f(y)p_\eta(y)\,dy =$$

$$\int\limits_B \left(\int\limits_{-\infty}^{+\infty} x\,\frac{p_{\xi,\eta}(x,y)}{p_\eta(y)}\,dx\right) p_\eta(y)\,dy = \int\limits_B \int\limits_{-\infty}^{+\infty} x\,p_{\xi,\eta}(x,y)\,dxdy = \mathbb{E}\big[\xi \cdot \mathrm{I}(\eta \in \mathrm{B})\big].$$

Итак, мы обсудили два частных случая и убедились в том, что обе рассмотренные величины обладают некоторым общим свойством. Перейдём теперь к более строгому, формальному опеределению условного математического ожидания и поймём, зачем мы установили вышеописанный факт.

Определения, примеры

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}\,\mathbb{P})$. Пусть $\mathcal{A} \subseteq \mathcal{F}$ — некоторая σ-подалгебра σ-алгебры $\mathcal{F}$.

Определение. Случайная величина η называется $\mathcal{A}$-измеримой, если для $\forall B \in \mathfrak{B} \quad \eta^{-1}(B) \in \mathcal{A}$.

Определение. *Условным математическим ожиданием* случайной величины ξ *относительно σ-алгебры* $\mathcal{A}$ называется случайная величина $\mathbb{E}\,(\xi|\mathcal{A})$, такая что:

1. $\mathbb{E}\,(\xi|\mathcal{A})$ — $\mathcal{A}$-измерима;
2. Для любого события $C \in \mathcal{A}$ $\quad \mathbb{E}\big[\mathbb{E}\,(\xi|\mathcal{A})\,\mathrm{I(C)}\big] \;=\; \mathbb{E}\big[\xi\,\mathrm{I(C)}\big]$.

Пример. Пусть $\Omega = \{1,2,3,4\}$, $\mathcal{F} = 2^{\Omega}$, $\mathbb{P}(\omega) = 1/4$ для $\omega = 1,\ldots,4$. Положим $\mathcal{A} = \big\{\varnothing, \{1,2\}, \{3,4\}, \Omega\big\}$. Тогда $\mathcal{A}$ — σ-алгебра, и $\mathcal{A} \subset \mathcal{F}$. Пусть случайная величина ξ имеет вид $\xi(\omega) = \omega^2$, $\omega = 1,\ldots,4$. Тогда

$$\mathbb{E}\big(\xi|\mathcal{A}\big) = \begin{cases} 1^2 \cdot \mathbb{P}\big(\omega = 1|\omega \in \{1,2\}\big) + 2^2 \cdot \mathbb{P}\big(\omega = 2|\omega \in \{1,2\}\big), & \omega = 1,2 \\ 3^2 \cdot \mathbb{P}\big(\omega = 3|\omega \in \{3,4\}\big) + 4^2 \cdot \mathbb{P}\big(\omega = 4|\omega \in \{3,4\}\big), & \omega = 3,4. \end{cases}$$

$$\mathbb{E}\big(\xi|\mathcal{A}\big) = \begin{cases} \frac{5}{2}, & \omega = 1,2 \\ \frac{25}{2}, & \omega = 3,4. \end{cases}$$

Определение. σ-алгебра $\mathcal{A}$ называется *порождённой* случайной величиной η, если $\mathcal{A} = \big\{\eta^{-1}(B)\colon B \in \mathfrak{B}\big\}$. Обозначается $\mathcal{A} = \sigma(\eta)$.

Определение. Условное математическое ожидание случайной величины ξ *относительно случайной величины* η — это условное математическое ожидание ξ относительно σ-алгебры $\sigma(\eta)$:

$$\mathbb{E}\,(\xi|\eta) \equiv \mathbb{E}\big(\xi|\sigma(\eta)\big).$$

Таким образом, величины 1.2 и 1.3 удовлетворяют определению и являются выражениями для УМО в дискретном и абсолютно непрерывном случае соответственно.

Пример. Возьмём вероятностное пространство из предыдущего примера. Рассмотрим случайную величину

$$\eta(\omega) = \begin{cases} 0, & \omega = 1,2 \\ 1, & \omega = 3,4. \end{cases}$$

Легко проверить, что $\sigma(\eta) = \mathcal{A}$ из предыдущего примера. Таким образом, совпадёт и условное математическое ожидание:

$$\mathbb{E}\big(\xi|\eta\big) = \begin{cases} \frac{5}{2}, & \omega = 1,2 \\ \frac{25}{2}, & \omega = 3,4. \end{cases}$$

Свойства условного математического ожидания.

1. *Условное математическое ожидание линейно:*

$$\mathbb{E}\big(a\xi + b\zeta|\mathcal{A}\big) \overset{\text{п.н.}}{=} a\,\mathbb{E}\big(\xi|\mathcal{A}\big) + b\,\mathbb{E}\big(\zeta|\mathcal{A}\big).$$

2. *Если ζ — $\mathcal{A}$-измеримая случайная величина, то*

$$\mathbb{E}\left(\xi\cdot\zeta|\mathcal{A}\right) = \zeta\cdot\mathbb{E}\left(\xi|\mathcal{A}\right).$$

   В частности, если мы рассмотрим условное математическое ожидание относительно случайной величины η, и положим $\zeta = f(\eta)$, где $f(x)$ — борелевская функция, то

$$\mathbb{E}\left(\xi\cdot\zeta|\eta\right) = \zeta\cdot\mathbb{E}\left(\xi|\eta\right).$$

3. *Если $\mathbb{E}\big|g(\xi,\eta)\big| < \infty$, то*

$$\mathbb{E}\,g(\xi,\eta) = \mathbb{E}\left[\mathbb{E}\big(g(\xi,y)|\eta\big)\Big|_{y=\eta}\right]$$

4. *Если ξ и η независимы, то $\mathbb{E}(\xi|\eta) = \mathbb{E}\,\xi$.*

5. *Взятие математического ожидания «убирает» условие:*

$$\mathbb{E}\big[\mathbb{E}\left(\xi|\mathcal{A}\right)\big] = \mathbb{E}\,\xi$$

С помощью условного математического ожидания можно ввести условную вероятность относительно σ-алгебры:

$$\mathbb{P}(B|\mathcal{A}) \overset{\text{def}}{=} \mathbb{E}\big(\mathrm{I(B)}|\mathcal{A}\big).$$

Математическая статистика

2.1 Статистическая структура. Выборка. Статистика. Порядковые статистики. Вариационный ряд. Эмпирическая функция распределения

Определение. *Статистическая структура* — совокупность $(\mathbb{R}^n, \mathfrak{B}^n, \mathcal{P}_\theta)$, где $\mathbb{R}^n$ — выборочное пространство, $\mathfrak{B}^n$ — борелевская σ-алгебра на $\mathbb{R}^n$, $\mathcal{P}_\theta$ — семейство распределений, определённых на $\mathfrak{B}^n$, параметризованное одно- или многомерным числовым параметром: $\mathcal{P}_\theta = (\mathbb{P}\,\theta \colon \theta \in \Theta \subseteq R^m)$.

Определение. *Выборка* $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ объёма n — набор из n независимых и одинаково распределённых случайных величин[1], имеющих такое же распределение, как и наблюдаемая случайная величина ξ.

До того, как эксперимент проведён, выборка — набор случайных величин, после — набор чисел из множества возможных значений случайной величины. Числовой набор $\boldsymbol{x} = (x_1, \ldots, x_n)$ будем называть *реализацией выборки*.

Замечание. Статистическая структура очень похожа на многомерное *индуцированное вероятностное пространство*. Отличие заключается в том, что если ранее мы использовали распределение лишь одной случайной величины ξ, то здесь используется целое семейство. Почему? Дело в том, что мы ещё не знаем, каковы параметры исследуемого распределения. Задача математической статистики как раз и заключается в том, чтобы их найти, основываясь на наблюдениях — реализациях выборок.

Отметим также то, что в различных источниках $\mathcal{P}_\theta$ может называться семейством *вероятностных мер*. Это тоже верно, но может внести путаницу — распределение случайной величины действительно является мерой, определённой на борелевской σ-алгебре. Однако вероятностная мера, вообще говоря, не связана ни с какой случайной величиной — это просто отображение из некоторой σ-алгебры событий в вещественные числа, удовлетворяющее аксиомам неотрицательности, ограниченности и счётной аддитивности. А распределение случайной величины — это композиция прообраза и вероятностной меры: $P_\xi = \mathbb{P} \circ \xi^{-1}$. Т.е. распределение содержит информацию о конкретной случайной величине, с некоторыми фиксированными параметрами. Именно поэтому мы используем семейство распределений. Нам не нужно много вероятностных мер — нам нужно много распределений с разными наборами параметров.

[1] Вообще говоря, в приложениях возникают также выборки, состоящие из зависимых или разнораспределённых элементов, но изучение их свойств не входит в этот курс.

Следует обратить внимание на то, что поскольку распределение зависит от параметра θ, то от него будут зависеть и математическое ожидание, и дисперсия, и прочие числовые характеристики налюдаемой случайной величины. Например, для «честной» монетки математическое ожидание при 10 подбрасываниях должно быть равно 5, а дисперсия — $npq = 10 \cdot \frac{1}{2} \cdot \frac{1}{2} = 2.5$. Но в случае асимметричной монетки с вероятностью выпадения «решки», равной 0.7, математическое ожидание будет равно 7, а дисперсия — 2.1. Поэтому лучше явно указывать зависимость от параметра и писать $\mathbb{P}_\theta, \mathbb{E}_\theta, \mathbb{D}_\theta$.

Определение. *Статистика* или *оценка* — измеримая функция от выборки, не зависящая от любых других параметров. Чаще всего статистики используются для поиска неизвестного параметра распределения θ и имеют вид $T\colon \mathbb{R}^n \mapsto \Theta$.

Определение. *Вариационный ряд* — набор случайных величин $X_{(1)}, \ldots, X_{(n)}$, который получается при упорядочении выборки $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ по возрастанию на каждом элементарном исходе.

$X_{(1)}(\omega) = \min\big(X_1(\omega), \ldots, X_n(\omega)\big)$ — *минимальная порядковая статистика*, $X_{(n)}(\omega) = \max\big(X_1(\omega), \ldots, X_n(\omega)\big)$ — соответственно, *максимальная*. Элемент $X_{(k)}$ — *k-я порядковая статистика*.

Замечание. Согласно нашему определению, вариационный ряд *не является выборкой*, хотя бы потому, что разные порядковые статистики, как мы вскоре убедимся, имеют разное распределение. Однако он является статистикой $T\colon \mathbb{R}^n \mapsto \mathbb{R}^n$, так как зависит только от выборки[1]. Порядковые статистики тоже удовлетворяют определению статистики, но являются уже отображениями из $\mathbb{R}^n$ в $\mathbb{R}$.

Определение. *Эмпирическая функция распределения*, построенная по выборке $X_1, \ldots, X_n$ объёма n, — случайная функция F_n^*:

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(X_i < y) \quad \forall y \in \mathbb{R}$$

Эмпирическая функция распределения строится по вариационному ряду следующим образом:

$$F_n^*(y) = \begin{cases} 0, & y \leqslant X_{(1)} \\ k/n, & X_{(k)} < y \leqslant X_{(k+1)} \\ 1, & y > X_{(n)} \end{cases}$$

[1] Конечно, формально ещё необходимо проверить на измеримость функцию, осуществляющую переупорядочение

Пример. Найдём эмпирические функции распределения для крайних порядковых статистик.

$$F_{(1)}(x) = \mathbb{P}_\theta\big(X_{(1)} < x\big) = 1 - \mathbb{P}_\theta\big(X_{(1)} \geqslant x\big) = 1 - \mathbb{P}_\theta\big(x_1 \geqslant x, \ldots, x_n \geqslant x\big) =$$
$$= 1 - \prod_{i=1}^{n} \mathbb{P}_\theta(x_i \geqslant x) = 1 - \big(\mathbb{P}_\theta(x_1 \geqslant x)\big)^n = 1 - \big(1 - F_\theta(x)\big)^n.$$
$$F_{(n)}(x) = \mathbb{P}_\theta\big(X_{(n)} < x\big) = \mathbb{P}_\theta\big(x_1 < x, \ldots, x_n < x\big) =$$
$$= \prod_{i=1}^{n} \mathbb{P}_\theta(x_i < x) = \big(\mathbb{P}_\theta(x_1 < x)\big)^n = \big(F_\theta(x)\big)^n.$$

Свойства эмпирической функции распределения.

1. *Пусть $\big(X_1, \ldots, X_n\big)$ — выборка из семейства распределений $\mathcal{P}_\theta$ с функцией распределения F_θ и пусть F_n^* — эмпирическая функция распределения, построенная по этой выборке. Тогда $F_n^*(y) \xrightarrow[n\to\infty]{p} F_\theta(y)$ для любого $y \in \mathbb{R}$ и $\forall \theta \in \Theta$.*

2. *Для любого $y \in \mathbb{R}$ и любого $\theta \in \Theta$:*

   1) $\mathbb{E}_\theta\, F_n^*(y) = F_\theta(y)$, т.е. $F_n^*(y)$ — несмещённая оценка для $F_\theta(y)$.

   2) $\mathbb{D}_\theta\, F_n^*(y) = \dfrac{F_\theta(y)\big(1 - F_\theta(y)\big)}{n} \leqslant \dfrac{1}{4n}$.

   3) Пусть $\sigma^2(y) = \big(1 - F_\theta(y)\big)F_\theta(y)$. Тогда

   $$\sqrt{n}\big(F_n^*(y) - F_\theta(y)\big) \xrightarrow[n\to\infty]{d} \mathrm{N}_{0,\sigma^2(y)},$$

   т.е. $F_n^*(y)$ — асимптотически нормальная оценка для $F_\theta(y)$.

   4) $nF_n^*(y) \sim \mathrm{Bi}_{n,F_\theta(y)}$.

   5) $F_n^*(y) \xrightarrow[n\to\infty]{п.н.} F_\theta(y)$.

Доказательство.

1. $F_n^*(y) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)$, при этом случайные величины $\mathrm{I}(X_1 < y)$, $\mathrm{I}(X_2 < y), \ldots$ независимы и одинаково распределены, их математическое ожидание конечно:

   $$\mathbb{E}_\theta\, \mathrm{I}(X_1 < y) = 1 \cdot \mathbb{P}_\theta(X_1 < y) + 0 \cdot \mathbb{P}_\theta(X_1 \geqslant y) = \mathbb{P}_\theta(X_1 < y) = F_\theta(y) < \infty$$

Следовательно, можно применить ЗБЧ в форме Хинчина:

$$F_n^*(y) = \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)}{n} \xrightarrow[n\to\infty]{\mathrm{p}} \mathbb{E}_\theta \, \mathrm{I}(X_1 < y) = F_\theta(y) \quad \forall \theta \in \Theta.$$

2. Заметим:

$$\mathrm{I}(X_1 < y) \sim \mathrm{Bi}_{1, F_\theta(y)} \implies \mathbb{E}_\theta \, \mathrm{I}(X_1 < y) = F_\theta(y),$$
$$\mathbb{D}_\theta \, \mathrm{I}(X_1 < y) = F_\theta(y)(1 - F_\theta(y)) \quad \forall \theta \in \Theta.$$

1) Случайные величины $\mathrm{I}(X_i < y)$ одинаково распределены, поэтому:

$$\mathbb{E}_\theta \, F_n^*(y) = \mathbb{E}_\theta \, \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)}{n} = \frac{\sum\limits_{i=1}^{n} \mathbb{E}_\theta \, \mathrm{I}(X_i < y)}{n} = \frac{n \, \mathbb{E}_\theta \, \mathrm{I}(X_1 < y)}{n} = F_\theta(y)$$

2) Случайные величины $\mathrm{I}(X_i < y)$ независимы и одинаково распределены, поэтому:

$$\mathbb{D}_\theta \, F_n^*(y) = \mathbb{D}_\theta \, \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)}{n} = \frac{\sum\limits_{i=1}^{n} \mathbb{D}_\theta \, \mathrm{I}(X_i < y)}{n^2} =$$
$$= \frac{n \, \mathbb{D}_\theta \, \mathrm{I}(X_1 < y)}{n^2} = \frac{F_\theta(y)\big(1 - F_\theta(y)\big)}{n}$$

Значения $F_\theta(y)$ принадлежат отрезку $[0, 1]$, а значит, произведение $F_\theta(y)\big(1 - F_\theta(y)\big) \leqslant \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{1}{4}$ (нетрудно убедиться, что $1/2$ — точка максимума). А значит, $\mathbb{D}_\theta \, F_n^* \leqslant \frac{1}{4n}$.

3. Применим ЦПТ:

$$\sqrt{n}\big(F_n^*(y) - F_\theta(y)\big) = \sqrt{n}\left(\frac{\sum \mathrm{I}(X_i < y)}{n} - F_\theta(y)\right) =$$
$$= \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y) - nF_\theta(y)}{\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y) - n \, \mathbb{E}_\theta \, \mathrm{I}(X_1 < y)}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathrm{d}}$$
$$\xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}_{0, \mathbb{D}_\theta \, \mathrm{I}(X_1 < y)} = \mathrm{N}_{0, (1 - F_\theta(y))F_\theta(y)} \, .$$

4. Следует из устойчивости по суммированию биномиального распределения. Поскольку $\mathrm{I}\,(X_i < y)$ независимы и имеют биномиальное распределение $\mathrm{Bi}_{1,F_\theta(y)}$, то их сумма

$$nF_n^*(y) = \mathrm{I}\,(X_1 < y) + \ldots + \mathrm{I}\,(X_n < y)$$

имеет биномиальное распределение $\mathrm{Bi}_{n,F_\theta(y)}$.

5. Выберем произвольный $y \in \mathbb{R}$. $\xi_i = \mathrm{I}(X_i < y)$ независимы, одинаково распределены и $\exists\, \mathbb{E}_\theta\, \xi_i = F_\theta(y)$. Тогда можно применить усиленный закон больших чисел в форме Колмогорова: $\mathbb{P}_\theta\left(\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i = F_\theta(y)\right) = 1\ \forall\theta \in \Theta$, что является непосредственным определением сходимости эмпирической функции распределения почти наверное к теоретической.

■

2.2 Точечная оценка. Несмещённость, состоятельность, оптимальность. Теорема о единственности оптимальной оценки

Определение. *Статистика* или *оценка* $T(\mathbf{X})$ — измеримая функция от выборки.

Определение. *Несмещённая оценка* параметра θ — статистика $T(\mathbf{X})$, т.ч. $\forall\theta \in \Theta\colon\ \mathbb{E}_\theta\, T(X) = \theta$.

Обозначим $T_n(\mathbf{X}) = T(\mathbf{X})$, чтобы подчеркнуть зависимость от объёма выборки.

Определение. *Асимптотически несмещённая оценка* параметра θ — статистика $T_n(\mathbf{X})$, т.ч. $\forall\theta \in \Theta\colon\ \ \mathbb{E}_\theta\, T_n(\mathbf{X}) \xrightarrow[n\to\infty]{} \theta$.

Определение. *Состоятельная оценка* параметра θ — статистика $T_n(\mathbf{X})$, т.ч. $\forall\theta \in \Theta\colon\ T_n(\mathbf{X}) \xrightarrow[n\to\infty]{p} \theta$.

Оценки также могут вводиться и для функций $\tau(\theta)$ параметра θ, для них все вышеуказанные определения вводятся аналогично.

Несмещённость означает отсутствие ошибки «в среднем», т. е. при систематическом использовании данной оценки. Несмещённость является желательным, но не обязательным свойством оценок. Достаточно, чтобы смещение оценки (разница между её средним значением и истинным параметром) уменьшалось с ростом объёма выборки. Поэтому асимптотическая несмещённость является весьма желательным свойством оценок.

Свойство состоятельности означает, что последовательность оценок приближается к неизвестному параметру при увеличении количества наблюдений. Вспомним определение сходимости по вероятности: $\mathbb{P}_\theta\Big(\big|T_n(\mathbf{X}) - \theta\big| > \varepsilon\Big) \xrightarrow[n\to\infty]{} 0 \ \forall \varepsilon > 0$. Мы можем зафиксировать некоторый ε_0 — допустимую погрешность, и найти такое n, что указанная вероятность будет мала — например, 0.01. Тогда значение оценки $T_n(\mathbf{X})$ с вероятностью 0.99 отклоняется от истинного значения не более чем на ε_0.
В отсутствие этого свойства статистика совершенно «несостоятельна» как оценка.

Замечание. Отметим некоторые свойства несмещённых и состоятельных оценок.

1. Несмещённые оценки не единственны.

   К примеру, в качестве несмещённой оценки для математического ожидания $\mathbb{E}_\theta X$ могут выступать $\mathbb{E}_\theta X_1$ или $\mathbb{E}_\theta \overline{X}$.

2. Несмещённые оценки могут не существовать.

   Пример. Дано распределение Pois_θ, над которым произведено одно наблюдение. Найти несмещённую оценку для функции $\tau(\theta) = \dfrac{1}{\theta}$.

$$\mathbb{E}_\theta T(\mathbf{X}) = \tau(\theta)$$
$$\sum_{x=0}^{\infty} T(x)\, e^{-\theta}\frac{\theta^x}{x!} = \frac{1}{\theta}$$
$$\sum_{x=0}^{\infty} T(x)\, \frac{\theta^{x+1}}{x!} = e^\theta, \text{ но } e^\theta = \sum_{k=0}^{\infty}\frac{\theta^k}{k!} \implies$$
$$T(x) \equiv \frac{1}{\theta} \quad \forall x \in \{0, 1, 2, 3, \ldots\}$$

   Мы получили, что функция $T(x)$ зависит от θ, но тогда она не может быть статистикой — ведь статистика должна зависеть только от выборки.

3. Несмещённые оценки могут существовать, но быть бессмысленными.

Пример. Дано геометрическое распределение $\mathrm{Geom}_{1-\theta}$ с вероятностью успеха $1-\theta$, над которым произведено одно наблюдение. Найти несмещённую оценку для параметра θ.

$$\mathbb{E}_\theta\, T(\mathbf{X}) = \theta$$

$$\sum_{x=0}^{\infty} T(x)(1-\theta)\theta^x = \theta$$

$$\sum_{x=0}^{\infty} T(x)\theta^x = \frac{\theta}{1-\theta} = \sum_{k=1}^{\infty}\theta^k \implies$$

$$T(x) = \begin{cases} 0, & \text{если } x = 0 \\ 1, & \text{если } x \geqslant 1 \end{cases}$$

Значения этой статистики не принадлежат параметрическому множеству $\Theta = (0,1)$, следовательно, эта оценка бессмысленна.

4. Состоятельные оценки не единственны.

   Как будет показано в следующем параграфе, выборочная дисперсия S^2 и несмещённая выборочная дисперсия S_0^2 являются состоятельными оценками теоретической дисперсии.

5. Состоятельные оценки могут быть смещёнными.

   Например, выборочная дисперсия является состоятельной, но смещённой оценкой теоретической дисперсии.

Как мы увидели, несмещённые и состоятельные оценки не единственны. Возникает вопрос — как определить, какая из нескольких имеющихся оценок лучше?

Рассмотрим несмещённые оценки $T(\mathbf{X})$ параметра θ, для которых существует дисперсия: $\mathbb{E}_\theta\big(T(\mathbf{X})-\theta\big)^2 = \mathbb{D}_\theta\, T(\mathbf{X})$. Обозначим класс всех таких оценок $\mathcal{T}$. Тогда можно оценивать точность оценок дисперсией. Если $T, T^* \in \mathcal{T}$ и $\mathbb{D}_\theta\, T^* \leqslant \mathbb{D}_\theta\, T \quad \forall\theta\in\Theta$, то по T^* *равномерно (по θ) не хуже T*.

Введём понятие оптимальной оценки.

Определение. *Оптимальная оценка* параметра θ — статистика $T^*(\mathbf{X})$, т.ч.:

1. $T^* \in \mathcal{T}$, т.е. T^* — несмещённая.

2. T^* имеет равномерно минимальную дисперсию, т.е. для любой другой **несмещённой** оценки $T_1 \in \mathcal{T}$ параметра θ: $\mathbb{D}_\theta\, T^* \leqslant \mathbb{D}_\theta\, T_1 \quad \forall\theta\in\Theta$.

Утверждение. *Если существует оптимальная оценка параметра θ, то она единственна.*

Доказательство. Предположим обратное: пусть существуют две оптимальные оценки $T_1(\mathbf{X})$ и $T_2(\mathbf{X})$ параметра θ. Тогда в силу их несмещённости: $\mathbb{E}_\theta T_1(\mathbf{X}) = \mathbb{E}_\theta T_2(\mathbf{X}) = T(\mathbf{X})$, а в силу того, что они имеют равномерно минимальную дисперсию: $\mathbb{D}_\theta T_1(\mathbf{X}) = \mathbb{D}_\theta T_2(\mathbf{X}) \quad \forall\theta \in \Theta$.

Введём новую статистику:

$$T_3(\mathbf{X}) = \frac{T_1(\mathbf{X}) + T_2(\mathbf{X})}{2}$$

Так как $\mathbb{E}_\theta T_3(\mathbf{X}) = \dfrac{\mathbb{E}_\theta T_1(\mathbf{X}) + \mathbb{E}_\theta T_2(\mathbf{X})}{2} = \theta$, то $T_3(\mathbf{X})$ — несмещённая оценка параметра θ.

Имеем также:

$$\mathbb{D}_\theta T_3(\mathbf{X}) = \frac{\mathbb{D}_\theta\big(T_1(\mathbf{X}) + T_2(\mathbf{X})\big)}{2^2} = \frac{\mathbb{D}_\theta T_1(\mathbf{X}) + \mathbb{D}_\theta T_2(\mathbf{X}) + 2\operatorname{cov}\big(T_1(\mathbf{X})T_2(\mathbf{X})\big)}{4}$$

По свойствам ковариции $|\operatorname{cov}(\xi, \eta)| \leqslant \sqrt{\mathbb{D}_\theta\xi\ \mathbb{D}_\theta\eta}$, а значит

$$\mathbb{D}_\theta T_3(\mathbf{X}) \leqslant \frac{\mathbb{D}_\theta T_1(\mathbf{X}) + \mathbb{D}_\theta T_2(\mathbf{X}) + 2\sqrt{\mathbb{D}_\theta T_1(\mathbf{X})}\sqrt{\mathbb{D}_\theta T_2(\mathbf{X})}}{4} = \mathbb{D}_\theta T_1(\mathbf{X})$$

В силу того, что $T_1(\mathbf{X})$ и $T_2(\mathbf{X})$ — оптимальные, дисперсия $T_3(\mathbf{X})$ не может быть меньше дисперсии $T_1(\mathbf{X})$, следовательно, неравенство должно обратиться в равенство, но тогда

$$T_1(\mathbf{X}) = aT_2(\mathbf{X}) + b \implies \mathbb{E}_\theta T_1(\mathbf{X}) = a\,\mathbb{E}_\theta T_2(\mathbf{X}) + b \iff$$
$$\iff \theta = a\theta + b\ \forall\theta \in \Theta \implies a = 1, b = 0$$

■

2.3 Выборочные моменты. Их свойства

В параграфе 2.1 мы предположили, что все случайные величины выборки $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ имеют одно и то же распределение, т.е. $X_i \sim \xi\ \ \forall i = \overline{1, n}$ для некоторой случайной величины ξ. Попробуем найти приближения некоторых числовых характеристик этой случайной величины.

Определение. *Выборочное математическое ожидание:*

$$\tilde{\mathbb{E}}\xi = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

Выборочное матожидание функции $g(\xi)$:

$$\tilde{\mathbb{E}} g\left(\xi\right) = \frac{1}{n} \sum_{i=1}^{n} g\left(X_i\right) = \overline{g(X)}$$

Определение. *Выборочная дисперсия:*

$$\tilde{\mathbb{D}}\xi = \sum_{i=1}^{n} \frac{1}{n} \big(X_i - \tilde{\mathbb{E}}\xi\big)^2 = \frac{1}{n} \sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2 = S^2$$

Определение. *Несмещённая выборочная дисперсия:*

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \frac{n}{n-1} S^2$$

Определение. *Выборочный момент k-го порядка:*

$$\tilde{\mathbb{E}} \left[\xi^k\right] = \sum_{i=1}^{n} \frac{1}{n} X_i^k = \frac{1}{n} \sum_{i=1}^{n} X_i^k = \overline{X^k}$$

Все вышеперечисленные характеристики являются случайными величинами как функции от выборки $\big(X_1, \ldots, X_n\big)$ и оценками для истинных моментов искомого распределения.

Введём ещё одно определение.

Определение. Статистика $T(\mathbf{X})$ называется *асимптотически нормальной*, если существуют такие $a_n(\theta), \sigma_n(\theta)$, что $\dfrac{T_n(\mathbf{X}) - a_n(\theta)}{\sigma_n(\theta)} \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}_{0,1}$. Иными словами, оценка называется асимптотически нормальной, если с ростом объёма выборки её функция распределения (оценка, будучи функцией от выборки, сама является случайной величиной) стремится к функции нормального распределения.

Утверждение. *Выборочное среднее $\overline{X}$ является несмещённой, состоятельной и асимптотически нормальной оценкой для теоретического среднего (математического ожидания), то есть:*

1. *Если $\mathbb{E}_\theta \left|X_1\right| < \infty$, то $\mathbb{E}_\theta \overline{X} = \mathbb{E}_\theta X_1 = a$;*

2. *Если* $\mathbb{E}_\theta |X_1| < \infty$, *то* $\overline{X} \xrightarrow[n\to\infty]{p} \mathbb{E}_\theta X_1 = a$;

3. *Если* $\mathbb{D}_\theta X_1 < \infty$, $\mathbb{D}_\theta X_1 \neq 0$,
то $\dfrac{\overline{X} - \mathbb{E}_\theta \overline{X}}{\sqrt{\mathbb{D}_\theta \overline{X}}} = \sqrt{n}\,\dfrac{\overline{X} - \mathbb{E}_\theta X_1}{\sqrt{\mathbb{D}_\theta X_1}} \xrightarrow[n\to\infty]{d} \mathrm{N}_{0,1}$.

Доказательство.

1. Из линейности математического ожидания:

$$\mathbb{E}_\theta \overline{X} = \frac{1}{n}\big(\mathbb{E}_\theta X_1 + \ldots + \mathbb{E}_\theta X_n\big) = \frac{1}{n}\cdot n\,\mathbb{E}_\theta X_1 = \mathbb{E}_\theta X_1 = a.$$

2. Из ЗБЧ в форме Хинчина:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{\mathrm{p}} \mathbb{E}_\theta X_1 = a.$$

3. Раскроем дисперсию суммы, пользуясь тем, что $X_1, \ldots X_n$ независимы и одинаково распределены, а затем домножим числитель и знаменатель на n. Тогда можно будет применить ЦПТ:

$$\frac{\big(\overline{X} - \mathbb{E}_\theta \overline{X}\big)}{\sqrt{\mathbb{D}_\theta \overline{X}}} = \frac{\big(\overline{X} - \mathbb{E}_\theta X_1\big)}{\sqrt{\mathbb{D}_\theta \left[\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right]}} = \frac{\big(\overline{X} - \mathbb{E}_\theta X_1\big)}{\sqrt{\frac{1}{n^2}\mathbb{D}_\theta \left[\sum\limits_{i=1}^{n} X_i\right]}} =$$

$$\frac{\big(\overline{X} - \mathbb{E}_\theta X_1\big)}{\sqrt{\frac{1}{n^2}\,n\,\mathbb{D}_\theta X_1}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mathbb{E}_\theta X_1}{\sqrt{\frac{1}{n}\,\mathbb{D}_\theta X_1}} = \sqrt{n}\,\frac{\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mathbb{E}_\theta X_1}{\sqrt{\mathbb{D}_\theta X_1}} =$$

$$\sqrt{n}\,\frac{\sum\limits_{i=1}^{n} X_i - n\,\mathbb{E}_\theta X_1}{n\sqrt{\mathbb{D}_\theta X_1}} = \frac{\sum\limits_{i=1}^{n} X_i - n\,\mathbb{E}_\theta X_1}{\sqrt{n\,\mathbb{D}_\theta X_1}} \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}_{0,1}$$

■

Замечание. Аналогичными свойствами обладает выборочный k-й момент, являющийся несмещённой, состоятельной и асимптотически нормальной оценкой для теоретического k-го момента.

Замечание. Применив УЗБЧ Колмогорова, можно показать, что выборочные k-е моменты сходятся к теоретическим почти наверное. Такие оценки называются *сильно состоятельными.* На практике обычно достаточно и состоятельности в обычном смысле (т.е. сходимости к теоретическому моменту по вероятности с ростом объёма выборки).

Утверждение. *Пусть* $\mathbb{D}_\theta\, X_1 < \infty$.

1. *Выборочные дисперсии* S^2 *и* S_0^2 *являются состоятельными оценками для истинной дисперсии:*
$$S^2 \xrightarrow[n\to\infty]{p} \mathbb{D}_\theta\, X_1 = \sigma^2, \quad S_0^2 \xrightarrow[n\to\infty]{p} \mathbb{D}_\theta\, X_1 = \sigma^2 \quad \forall\theta\in\Theta.$$

2. S^2 *— смещённая оценка дисперсии, а* S_0^2 *— несмещённая:*
$$\mathbb{E}_\theta\, S^2 = \frac{n-1}{n}\,\mathbb{D}_\theta\, X_1 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad \mathbb{E}_\theta\, S_0^2 = \mathbb{D}_\theta\, X_1 = \sigma^2 \quad \forall\theta\in\Theta.$$

3. *Если* $0 \neq \mathbb{D}_\theta\Big[\big(X_1 - \mathbb{E}_\theta\, X_1\big)^2\Big] < \infty$, *то* S^2 *и* S_0^2 *являются асимптотически нормальными оценками истинной дисперсии:*
$$\sqrt{n}\,\big(S^2 - \mathbb{D}_\theta\, X_1\big) \xrightarrow[n\to\infty]{d} \mathrm{N}_{0,\mathbb{D}_\theta\left(X_1-\mathbb{E}_\theta\, X_1\right)^2} \quad \forall\theta\in\Theta.$$

Доказательство.

1. $S^2 = \dfrac{1}{n}\sum\limits_{i=1}^n \big(X_i - \overline{X}\big)^2 = \dfrac{1}{n}\sum\limits_{i=1}^n\Big(X_i^2 - 2X_i\overline{X} + \big(\overline{X}\big)^2\Big) =$
$\dfrac{1}{n}\Big(n\overline{X^2} - 2\overline{X}\sum\limits_{i=1}^n X_i + n\big(\overline{X}\big)^2\Big) = \overline{X^2} - 2\big(\overline{X}\big)^2 + \big(\overline{X}\big)^2 = \overline{X^2} - \big(\overline{X}\big)^2.$
Используя состоятельность первого и второго выборочных моментов и свойства сходимости по вероятности, получаем:
$$S^2 = \overline{X^2} - (\overline{X})^2 \xrightarrow[n\to\infty]{\mathrm{p}} \mathbb{E}_\theta\, X_1^2 - \big(\mathbb{E}_\theta\, X_1\big)^2 = \sigma^2$$
$$\frac{n}{n-1} \xrightarrow[n\to\infty]{} 1 \implies S_0^2 = \frac{n}{n-1}S^2 \xrightarrow[n\to\infty]{\mathrm{p}} \sigma^2$$

2. Используя несмещённость первого и второго выборочных моментов:
$$\mathbb{E}_\theta\, S^2 = \mathbb{E}_\theta\left(\overline{X^2} - (\overline{X})^2\right) = \mathbb{E}_\theta\,\overline{X^2} - \mathbb{E}_\theta\big(\overline{X}\big)^2 = \mathbb{E}_\theta\, X_1^2 - \mathbb{E}_\theta\big(\overline{X}\big)^2 =$$
$$= \mathbb{E}_\theta\, X_1^2 - \left(\big(\mathbb{E}_\theta\,\overline{X}\big)^2 + \mathbb{D}_\theta\,\overline{X}\right) = \mathbb{E}_\theta\, X_1^2 - \big(\mathbb{E}_\theta\, X_1\big)^2 - \mathbb{D}_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) =$$

$$= \mathbb{D}_\theta X_1 - \mathbb{D}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \sigma^2 - \frac{1}{n^2} n \, \mathbb{D}_\theta X_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

Откуда следует:

$$\mathbb{E}_\theta S_0^2 = \frac{n}{n-1} \, \mathbb{E}_\theta S^2 = \sigma^2.$$

3. Введём случайные величины $Y_i = X_i - a$; $\mathbb{E}_\theta Y_i = 0$, $\mathbb{D}_\theta Y_1 = \mathbb{D}_\theta X_1 = \sigma^2$.

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - a - (\overline{X} - a))^2 = \overline{Y^2} - \left(\overline{Y}\right)^2.$$

$$\sqrt{n}\left(S^2 - \sigma^2\right) = \sqrt{n}\left(\overline{Y^2} - \left(\overline{Y}\right)^2 - \sigma^2\right) = \sqrt{n}\left(\overline{Y^2} - \mathbb{E}_\theta Y_1^2\right) - \sqrt{n}\left(\overline{Y}\right)^2 =$$

$$= \frac{\sum\limits_{i=1}^{n} Y_i^2 - n \, \mathbb{E}_\theta Y_1^2}{\sqrt{n}} - \overline{Y} \cdot \sqrt{n} \, \overline{Y} \xrightarrow[n \to \infty]{\mathrm{d}} \mathrm{N}^2_{0, \mathbb{D}_\theta (X_1 - \mathbb{E}_\theta X_1)}),$$

поскольку $\dfrac{\sum\limits_{i=1}^{n} Y_i^2 - n \, \mathbb{E}_\theta Y_1^2}{\sqrt{n}} \xrightarrow[n \to \infty]{\mathrm{d}} \mathrm{N}_{0, \mathbb{D}_\theta Y_1^2}$ по ЦПТ, а $\overline{Y} \cdot \sqrt{n} \, \overline{Y} \xrightarrow[n \to \infty]{\mathrm{d}} 0$ как произведение последовательностей $\overline{Y} \xrightarrow[n \to \infty]{p} 0$ и $\sqrt{n} \, \overline{Y} \xrightarrow[n \to \infty]{\mathrm{d}} \mathrm{N}_{0, \mathbb{D}_\theta X_1}$.

■

2.4 Функция правдоподобия. Достаточные статистики, полные статистики. Теорема факторизации

В зависимости от типа семейства распределения $\mathcal{P}_\theta$ обозначим через $f_\theta(y)$ одну из следующих функций:

$$f_\theta(y) = \begin{cases} \text{плотность } f_\theta(y), & \text{если } \mathcal{P}_\theta \text{ абсолютно непрерывно,} \\ P_\theta\left(X_1 = y\right), & \text{если } \mathcal{P}_\theta \text{ дискретно.} \end{cases}$$

Определение. *Функция правдоподобия* выборки $\mathbf{X}$:

$$L(\mathbf{X}, \theta) = f_\theta\left(X_1\right) \cdot f_\theta\left(X_2\right) \cdot \ldots \cdot f_\theta\left(X_n\right) = \prod_{i=1}^{n} f_\theta\left(X_i\right).$$

В дискретном случае функция правдоподобия принимает вид:

$$L(\mathbf{X}, \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \mathbb{P}_\theta(X_1 = x_1) \cdot \ldots \cdot \mathbb{P}_\theta(X_n = x_n) =$$
$$= \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n).$$

Замечание. Функция правдоподобия зависит от выборки $\mathbf{X}$ и от параметра θ. Если мы зафиксируем некоторое θ_0, то получим вероятностную меру над выборочным пространством $\mathbb{R}^n$. Тогда значение функции правдоподобия $L(\mathbf{x}, \theta_0)$ на некоторой реализации выборки x — это вероятность этой реализации в предположении, что параметр θ_0 — истинное значение.
Если же мы напротив, зафиксируем некоторую реализацию выборки $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ (что соответствует реальным условиям — как правило, у нас есть некоторые наблюдения, а параметр неизвестен), то значение функции правдоподобия на некотором параметре θ — это «правдоподобие», «вероятность» этого параметра. Но, строго говоря, при фиксированной реализации выборки x^* функция правдоподобия $L(\mathbf{x}^*, \theta)$ не является вероятностной мерой над Θ. Рассмотрим пример.

Пример. При двух последовательных независимых подбрасываниях монетки выпало два «орла». Запишем функцию правдоподобия выборки $\mathbf{X} = \big(X_1, X_2\big),\ X_i \in \{0, 1\}$. Обозначим вероятность выпадения «орла» $\big(\mathbb{P}_\theta(X_i = 1)\big)$ за $\theta,\ \theta \in [0, 1]$. Тогда

$$L(\mathbf{X}, \theta) = \mathbb{P}_\theta(X_1 = x_1) \cdot \mathbb{P}_\theta(X_2 = x_2).$$

Для любого фиксированнного значения параметра θ_0 функция правдоподобия $L(\mathbf{X}, \theta_0)$ является функцией совместного распределения (X_1, X_2):

	$X_1 = 1$	$X_1 = 0$
$X_2 = 1$	θ_0^2	$(1-\theta_0)\theta_0$
$X_2 = 0$	$\theta_0(1-\theta_0)$	$(1-\theta_0)^2$

$$\sum_{x_1=0}^{1}\sum_{x_2=0}^{1} L(\mathbf{x}, \theta_0) = \sum_{x_1=0}^{1}\sum_{x_2=0}^{1} \mathbb{P}\,\theta_0(X_1 = x_1) \cdot \mathbb{P}\,\theta_0(X_2 = x_2) =$$
$$= \theta_0^2 + (1-\theta_0)\theta_0 + \theta_0(1-\theta_0) + (1-\theta_0)^2 = \theta_0^2 + 2\theta_0 - 2\theta_0^2 + (1 - 2\theta_0 + \theta_0^2) = 1.$$

Но если мы зафиксируем реализацию выборки $x^* = (x_1^* = 1, x_2^* = 1)$ и проинтегрируем по $\Theta = [0, 1]$, то получим

$$\int\limits_0^1 L(\mathbf{x}^*, \theta)\, d\theta = \int\limits_0^1 \mathbb{P}_\theta(X_1 = 1) \cdot \mathbb{P}_\theta(X_2 = 1)\, d\theta = \int\limits_0^1 \theta^2\, d\theta = \frac{\theta^3}{3}\bigg|_0^1 = \frac{1}{3}.$$

Выходит, что $L(\mathbf{x}^*, \theta)$ не может быть вероятностной мерой, так как интеграл по Θ не равен 1.

Определение. *Достаточная статистика* — статистика $T(\mathbf{X})$ такая, что $\forall t \in \mathbb{R}^m$, [1] $\forall B \in \mathfrak{B}^n$ условное распределение $\mathbb{P}_\theta\Big(\big(X_1, \ldots, X_n\big) \in B | \, T = t\Big)$ не зависит от параметра θ. Иными словами, статистика $T(\mathbf{X})$ называется достаточной, если $\mathbb{E}_\theta\big(\mathbf{X} \,|\, T(\mathbf{X})\big)$ не зависит от θ.

Иными словами, достаточная статистика содержит в себе всю информацию о параметре, которую можно извлечь из выборки.

Достаточная статистика всегда существует — например, сама выборка $\mathbf{X}$ является достаточной статистикой. Однако этот тривиальный пример не очень полезен. Как правило, нас интересуют достаточные статистики малой размерности. Для того, чтобы проверить, является ли статистика достаточной, можно использовать критерий факторизации.

Критерий факторизации. $T(\mathbf{X})$ *— достаточная статистика* $\iff$ *её функция правдоподобия представима в виде*

$$L(X_1, \ldots, X_n, \theta) \overset{\text{п.н.}}{=} g\big(T(\mathbf{X}), \theta\big) \cdot h(\mathbf{X})$$

Доказательство. Рассмотрим только дискретный случай.

$\Longrightarrow$ Пусть $T(\mathbf{X})$ — достаточная статистика. Возьмём произвольную реализацию выборки $\boldsymbol{x}$ и обозначим $t = T(\boldsymbol{x})$. Тогда

$$\begin{aligned} L(\boldsymbol{x}, \theta) = \mathbb{P}_\theta(\mathbf{X} = \boldsymbol{x}) = \mathbb{P}_\theta(\mathbf{X} = \boldsymbol{x}, T(\mathbf{X}) = t) = \\ = \underbrace{\mathbb{P}_\theta(T(\mathbf{X}) = t)}_{g(T(\mathbf{X}), \theta)} \underbrace{\mathbb{P}_\theta(\mathbf{X} = \boldsymbol{x} | T(\mathbf{X}) = t)}_{h(\mathbf{X})} \end{aligned}$$

$\Longleftarrow$ Пусть теперь функция правдоподобия имеет вид $L(\boldsymbol{x}, \theta) = g\big(T(\boldsymbol{x}), \theta\big) h(\boldsymbol{x})$. Тогда, если x таково, что $T(\boldsymbol{x}) = t$, то:

[1] Напомним, что вообще говоря, статистика — это отображение $T(X) \colon \mathbb{R}^n \mapsto \mathbb{R}^m$

$$\mathbb{P}_\theta(\mathbf{X}=\boldsymbol{x}|\,T(\mathbf{X})=t)=\frac{\mathbb{P}_\theta(\mathbf{X}=\boldsymbol{x},T(\mathbf{X})=t)}{\mathbb{P}_\theta(T(\mathbf{X})=t)}=\frac{\mathbb{P}_\theta(\mathbf{X}=\boldsymbol{x})}{\sum\limits_{\boldsymbol{x}':T(\boldsymbol{x}')=t}\mathbb{P}_\theta(\mathbf{X}=\boldsymbol{x}')}=$$

$$=\frac{g(t,\theta)h(\boldsymbol{x})}{\sum\limits_{\boldsymbol{x}':T(\boldsymbol{x}')=t}g(t,\theta)h(\boldsymbol{x}')}=\frac{h(\boldsymbol{x})}{\sum\limits_{\boldsymbol{x}':T(\boldsymbol{x}')=t}h(\boldsymbol{x}')}$$

Т.е. вероятность не зависит от θ, а значит, $T(\mathbf{X})$ достаточная.

■

Определение. Статистика $T(\mathbf{X})$ называется *полной*, если для любой борелевской функции $\varphi(T)$, для которой $\mathbb{E}_\theta\,\varphi(T)=0\ \forall\theta\in\Theta$, справедливо: $\varphi(T)\overset{\text{п.н.}}{=}0$.

Пример. Рассмотрим равномерное распределение $\mathrm{U}_{[0,\theta]}$ и докажем, что статистика $T(\mathbf{X})=X_{(n)}$ является полной. Напоминаем, что плотность распредения $X_{(n)}$ для $U[0,\theta]$ - это $f(t)=\dfrac{nt^{n-1}}{\theta^n}\,\mathrm{I}(0\leqslant t\leqslant\theta)$.

$$\mathbb{E}_\theta\,\varphi(X_{(n)})=\int\limits_0^\theta\varphi(t)\frac{nt^{n-1}}{\theta^n}dt=0\implies\int\limits_0^\theta\varphi(t)t^{n-1}dt=0\implies$$
$$\{\text{продифференцируем по }\theta\}\implies\varphi(\theta)\theta^{n-1}=0\implies\varphi(T)\equiv0\ \ \forall T>0$$

Таким образом, $\varphi(T)\overset{\text{п.н.}}{=}0$ при $T\geqslant0$, и указанная статистика является полной.

2.5 Неравенство Рао—Крамера. Эффективные оценки

Пусть $X_1,\ldots,X_n$ — некоторая выборка с функцией правдоподобия $L(\mathbf{X},\theta)$ относительно некоторой меры μ. Введём функцию $\varphi(\theta)=\int\limits_{\mathbf{R}^n}T(x)L(\mathbf{x},\theta)\mu(dx)<\infty$, в дальнейшем считая, что она дифференцируема необходимое число раз.

Определение. Функция правдоподобия $L(\mathbf{X},\theta)$ *удовлетворяет условиям регулярности для m-й производной*, если существует

$$\frac{d^m\varphi(\theta)}{d\theta^m}=\int\limits_{\mathbb{R}^n}T(x)\frac{\partial^m}{\partial\theta^m}L(\mathbf{x},\theta)\mu(dx),$$

причём множество $\{x\colon L(\mathbf{x},\theta)>0\}$ не зависит от параметра θ.

Пример. Условиям регулярности удовлетворяют многие модели: биномиальная, пуассоновская, нормальная, гамма-распределение (а следовательно, и экспоненциальная, и χ^2) и т.д. Требование независимости множества значений исследуемой случайной величины от параметра θ существенно: например, равномерное распределение $\mathrm{U}_{[0,\theta]}$ относится к нерегулярным моделям. Это происходит из-за того, что при дифференцировании интеграла, пределы интегрирования которого зависят от параметра, появляются дополнительные слагаемые. Т.е. менять местами интегрирование и дифференцирование нельзя, и к нерегулярным моделям изложенная ниже теория неприменима.

Определение. Функция $U(\mathbf{X},\theta) = \dfrac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}$ называется *функцией вклада.*

Утверждение. Если функция правдоподобия удовлетворяет условиям регулярности для первой производной, то $\mathbb{E}_\theta\, U(\mathbf{X},\theta) = 0$. В самом деле,

$$\mathbb{E}_\theta\, U(\mathbf{X},\theta) = \int\limits_{\mathbb{R}^n} U(x,\theta)L(\mathbf{x},\theta)\mu(dx) = \int\limits_{\mathbb{R}^n} \frac{\partial \ln L(\mathbf{x},\theta)}{\partial\theta} L(\mathbf{x},\theta)\mu(dx) =$$

$$\int\limits_{\mathbb{R}^n} \frac{1}{L(\mathbf{x},\theta)}\frac{\partial L(\mathbf{x},\theta)}{\partial\theta} L(\mathbf{x},\theta)\mu(dx) = \int\limits_{\mathbb{R}^n} \frac{\partial L(\mathbf{x},\theta)}{\partial\theta}\mu(dx) = \{\text{регулярность}\}$$

$$= \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} L(\mathbf{x},\theta)\mu(dx) = \frac{\partial}{\partial\theta}\,1 = 0.$$

Из этого, в частности, вытекает, что для регулярных моделей $\mathbb{D}_\theta\, U(\mathbf{X},\theta) = \mathbb{E}_\theta\, U^2(\mathbf{X},\theta)$.

Посчитаем дисперсию функции вклада:

$$\mathbb{D}_\theta\, U(\mathbf{X},\theta) = \mathbb{D}_\theta\,\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta} = \mathbb{D}_\theta\,\frac{\partial \ln \prod\limits_{i=1}^{n} f_\theta(X_i)}{\partial\theta} =$$

$$\mathbb{D}_\theta\,\frac{\partial \sum\limits_{i=1}^{n} \ln f_\theta(X_i)}{\partial\theta} = \mathbb{D}_\theta \sum_{i=1}^{n}\frac{\partial \ln f_\theta(X_i)}{\partial\theta} = \{\text{независимость выборки}\} =$$

$$\sum_{i=1}^{n}\mathbb{D}_\theta\,\frac{\partial \ln f_\theta(X_i)}{\partial\theta} = \{\text{однородность выборки}\} =$$

$$n\;\mathbb{D}_\theta\,\frac{\partial \ln\, f_\theta(X_1)}{\partial\theta} = n\;\mathbb{D}_\theta\, U(X_1,\theta) = n\, i_1(\theta).$$

Здесь за $i_1(\theta)$ обозначена дисперсия функции вклада от выборки из одного элемента.

Определение. Пусть $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ — выборка объёма n. Величину $i_n(\theta) = \mathbb{D}_\theta\, U(\mathbf{X}, \theta)$ называют *фишеровской информацией, содержащейся в выборке размера n.*

Неравенство Рао—Крамера. *Пусть $\big(X_1, \ldots, X_n\big)$ — выборка, $L(\mathbf{X}, \theta)$ удовлетворяет условиям регулярности для первой производной и $\tau(\theta)$ — дифференцируемая функция θ. Тогда:*

1. *Для любой $T(\mathbf{X})$ — несмещённой оценки $\tau(\theta)$, справедливо неравенство:*

$$\mathbb{D}_\theta\, T(\mathbf{X}) \geqslant \frac{\big(\tau'(\theta)\big)^2}{n\, i_1(\theta)} = \frac{\big(\tau'(\theta)\big)^2}{\mathbb{D}_\theta\, U(\mathbf{X}, \theta)} \quad \forall \theta \in \Theta$$

2. *Равенство достигается $\iff$ $\exists a_n(\theta)\colon\ T(\mathbf{X}) - \tau(\theta) = a_n(\theta) \cdot U(\mathbf{X}, \theta)$*

Доказательство. Из условий регулярности $L(\mathbf{X}, \theta)$ для следует:

$$\int L(\mathbf{x}, \theta)\mu(dx) = 1 \quad \Longrightarrow \quad \int \frac{\partial L(\mathbf{x}, \theta)}{\partial \theta}\mu(dx) = 0$$

$$\int T(x) L(\mathbf{x}, \theta)\mu(dx) = \mathbb{E}_\theta\, T(\mathbf{X}) = \tau(\theta) \quad \Longrightarrow \quad \int T(x)\frac{\partial L(\mathbf{x}, \theta)}{\partial \theta}\mu(dx) = \tau'(\theta)$$

Заметим, что

$$\frac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} \cdot L(\mathbf{x}, \theta)$$

Откуда следует:

$$\int U(x, \theta) L(\mathbf{x}, \theta)\mu(dx) = 0 \iff \mathbb{E}_\theta\big[U(\mathbf{X}, \theta)\big] = 0$$

$$\int T(x) U(x, \theta) L(\mathbf{x}, \theta)\mu(dx) = \tau'(\theta) \iff \mathbb{E}_\theta\big[T(\mathbf{X}) U(\mathbf{X}, \theta)\big] = \tau'(\theta)$$

Вычитая из второго равенства первое, умноженного на $\tau(\theta) = \mathbb{E}_\theta\,[T(\mathbf{X})]$, получаем:

$$\mathbb{E}_\theta\,[T(\mathbf{X}) U(\mathbf{X}, \theta)] - \mathbb{E}_\theta\,[T(\mathbf{X})]\ \mathbb{E}_\theta\,[U(\mathbf{X}, \theta)] = \tau'(\theta) - 0 \cdot \tau(\theta) = \tau'(\theta)$$

В левой части полученного равенства стоит ковариация случайных величин $T(\mathbf{X})$ и $U(\mathbf{X}, \theta)$:

$$\mathrm{cov}_\theta\big(T(\mathbf{X}), U(\mathbf{X}, \theta)\big) = \tau'(\theta)$$

Из неравенства Коши-Буняковского:

$$\left(\tau'(\theta)\right)^2 = \operatorname{cov}^2_\theta\big(T(\mathbf{X}),\, U(\mathbf{X},\theta)\big) \leqslant \mathbb{D}_\theta\, T(\mathbf{X})\ \mathbb{D}_\theta\, U(\mathbf{X},\theta) = \mathbb{D}_\theta\, T(\mathbf{X})\, n\, i_1(\theta),$$

что равносильно п.1 теоремы:

$$\mathbb{D}_\theta\, T(\mathbf{X}) \geqslant \frac{\left(\tau'(\theta)\right)^2}{n\, i_1(\theta)}$$

Равенство достигается, если статистика и функция вклада линейно связаны (опять же, следствие неравенства Коши-Буняковского):

$$T(\mathbf{X}) = \varphi(\theta)U(\mathbf{X},\theta) + \psi(\theta) \implies \tau(\theta) = \psi(\theta),\ a_n(\theta) = \varphi(\theta). \qquad (2.1)$$

■

Оценки, обращающие неравенство Рао—Крамера в равенство, выделяются в отдельный класс.

Определение. *Эффективная оценка* $T(\mathbf{X})$ — это несмещённая оценка параметра θ (или функции $\tau(\theta)$), дисперсия которой совпадает с нижней гранью в неравенстве Рао—Крамера.

Замечание. Если существует эффективная оценка для функции $\tau(\theta)$, то ни для какой другой функции от θ, кроме линейного преобразования $\tau(\theta)$, эффективной оценки существовать не будет. Это следует из (2.1).

Так как дисперсия любой оценки в регулярной модели не может быть меньше нижней грани, определяемой неравенством Рао—Крамера, то каждая эффективная оценка является оптимальной. Обратное, в силу предыдущего замечания, неверно.

2.6 Теорема Рао—Блекуэлла—Колмогорова. Оптимальность оценок, являющихся функцией полной достаточной статистики

Теорема Рао—Блекуэлла—Колмогорова. *Если существует оптимальная оценка для* $\tau(\theta)$*, то она является функцией от достаточной статистики.*

Доказательство. В доказательстве используются следующие свойства условного математического ожидания:

$$\mathbb{E}\, f(x,z) = \mathbb{E}\Big(\mathbb{E}\big(f(x,z)|z\big)\Big),$$

$$\mathbb{E}\big(g(z)|z\big) = g(z).$$

Мы докажем даже более сильное утверждение: для любой несмещённой оценки мы можем построить новую оценку, являющуюся функцией от достаточной статистики, при этом дисперсия построенной оценки будет не больше исходной. Отсюда вытекает и утверждение теоремы — ведь оптимальная оценка является несмещённой, соответственно, мы можем построить новую оценку, которая будет равномерно по θ не хуже оптимальной. Но оптимальная оценка единственна, а значит, она сама является функцией от достаточной статистики, что и требуется доказать.

1. Построим искомую оценку. Пусть $T(\mathbf{X})$ — достаточная статистика, $T_1(\mathbf{X})$ — несмещённая оценка $\tau(\theta)$, т.е. $\mathbb{E}_\theta\, T_1(\mathbf{X}) = \tau(\theta)$. Рассмотрим функцию $H(T) = \mathbb{E}_\theta\,(T_1|T)$. Тогда из первого свойства следует:

$$\mathbb{E}_\theta\, H(T) = \mathbb{E}_\theta\big(\mathbb{E}_\theta\,(T_1|T)\big) = \mathbb{E}_\theta\, T_1 = \tau(\theta) \implies$$
$$H(T) \text{ — несмещённая оценка } \tau(\theta).$$

2. Покажем, что её дисперсия не превосходит дисперсию исходной:

$$\mathbb{D}_\theta\, T_1 = \mathbb{E}_\theta\big(T_1 - \tau(\theta)\big)^2 =$$
$$\mathbb{E}_\theta\Big(T_1 - H(T) + H(T) - \tau(\theta)\Big)^2 =$$
$$\underbrace{\mathbb{E}_\theta\big(T_1 - H(T)\big)^2}_{\geqslant 0} + \underbrace{2\ \mathbb{E}_\theta\Big[\big(T_1 - H(T)\big)\big(H(T) - \tau(\theta)\big)\Big]}_{?} + \mathbb{D}_\theta\, H(T)$$

Оценим второе слагаемое, пользуясь тем, что $H(T)$ — функция от T и может быть вынесена из под условного математического ожидания как константа:

$$\mathbb{E}_\theta\Big[\big((T_1 - H(T)\big)\big(H(T) - \tau(\theta)\big)\Big] =$$
$$= \mathbb{E}_\theta\bigg(\mathbb{E}_\theta\Big[\big(T_1 - H(T)\big)\big(H(T) - \tau(\theta)\big)|\,T\Big]\bigg) =$$
$$= \mathbb{E}_\theta\bigg(\big(H(T) - \tau(\theta)\big)\ \mathbb{E}_\theta\Big[\big(T_1 - H(T)\big)|\,T\Big]\bigg) =$$
$$= \mathbb{E}_\theta\bigg(\big(H(T) - \tau(\theta)\big)\Big[\mathbb{E}_\theta\big(T_1|T\big) - \mathbb{E}_\theta\big(H(T)|\,T\big)\Big]\bigg) =$$
$$= \mathbb{E}_\theta\bigg(\big(H(T) - \tau(\theta)\big)\Big[H(T) - H(T)\Big]\bigg) = 0.$$

Отсюда и вытекает, что

$$\mathbb{D}_\theta\, T_1 = \underbrace{\mathbb{E}_\theta\big(T_1 - H(T)\big)^2}_{\geqslant 0} + \mathbb{D}_\theta\, H(T) \;\geqslant\; \mathbb{D}_\theta\, H(T).$$

Таким образом, если существует оптимальная оценка T_1, то $H(T)$ тоже оптимальна, но мы знаем, что оптимальная оценка единственна. Осталось заметить, что $H(T) = \mathbb{E}_\theta\big(T_1|T\big) \equiv f(T)$ — функция от достаточной статистики. ∎

Теорема Колмогорова. *Пусть $T(\mathbf{X})$ — полная достаточная статистика. Пусть $\varphi(x)$ — борелевская функция, и $\mathbb{E}_\theta\, \varphi\big(T(\mathbf{X})\big) = \tau(\theta)$. Тогда $\varphi\big(T(\mathbf{X})\big)$ — оптимальная оценка для $\tau(\theta)$.*

Доказательство. Из теоремы Рао—Блекуэлла—Колмогорова следует, что если существует несмещённая оценка для $\tau(\theta)$, то существует также и несмещённая оценка $\tau(\theta)$, являющаяся функцией от достаточной статистики.

Пусть $T_1 = \varphi_1(T)$ — несмещённая оценка $\tau(\theta)$ и T — полная достаточная статистика. Допустим, что существует ещё одна несмещённая оценка $T_2 = \varphi_2(T)$. Тогда $\forall \theta \in \Theta$

$$\mathbb{E}_\theta\Big(\varphi_1\big(T(\mathbf{X})\big) - \varphi_2\big(T(\mathbf{X})\big)\Big) = 0.$$

Но $T(\mathbf{X})$ полная, $\implies$ $\varphi_1(y) \overset{\text{п.н.}}{=} \varphi_2(y)$ на области значений $T(\mathbf{X})$. Таким образом, несмещённая оценка, являющаяся функцией от достаточной статистики, единственна. Значит, она и будет оптимальной оценкой $\tau(\theta)$.

∎

2.7 Метод моментов. Свойства оценок, полученных методом моментов

Пусть $\big(X_1, \ldots, X_n\big)$ — выборка объёма n из параметрического семейства распределений $\mathcal{P}_\theta$. Выберем функцию $g(y)\colon \mathbb{R} \mapsto \mathbb{R}$ так, чтобы существовал момент $\mathbb{E}_\theta\, g\,(X_1) = h(\theta)$ и функция $h(\theta)$ была обратима на Θ. Разрешим полученное уравнение относительно θ, а затем вместо истинного момента возьмём выборочный:

$$\theta = h^{-1}\big(\mathbb{E}_\theta\, g\,(X_1)\big), \quad \theta^* = h^{-1}\big(\overline{g(\mathbf{X})}\big) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g\,(X_i)\right)$$

Полученная оценка θ^* — *оценка метода моментов* для параметра θ. Чаще всего берут $g(y) = y^k$. В этом случае, при условии обратимости функции h

на Ω:

$$\mathbb{E}_\theta X_1^k = h(\theta), \quad \theta = h^{-1}\left(\mathbb{E}_\theta X_1^k\right), \quad \theta^* = h^{-1}\big(\overline{X^k}\big) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right)$$

Пример. Рассмотрим равномерное распределение $\mathrm{U}_{[0,\theta]}$. Найдём оценку метода моментов для параметра θ по первому моменту:

$$\mathbb{E}_\theta X_1 = \frac{\theta}{2}, \quad \theta = 2\,\mathbb{E}_\theta X_1, \quad \theta_1^* = 2\overline{X}$$

Найдём оценку метода моментов k по k-му моменту:

$$\mathbb{E}_\theta X_1^k = \int\limits_0^\theta y^k \frac{1}{\theta}\,dy = \frac{\theta^k}{k+1}, \quad \theta = \sqrt[k]{(k+1)\,\mathbb{E}_\theta X_1^k}, \quad \theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}$$

Утверждение. *Пусть $\theta^* = h^{-1}\big(\overline{g(\mathbf{X})}\big)$ — оценка параметра θ, полученная методом моментов, причём функция h^{-1} непрерывна. Тогда оценка θ^* состоятельна.*

Доказательство. По ЗБЧ Хинчина имеем:

$$\overline{g(\mathbf{X})} = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}_\theta\, g\left(X_1\right) = h(\theta)$$

Ввиду непрерывности функции h^{-1}:

$$\theta^* = h^{-1}\big(\overline{g(\mathbf{X})}\big) \xrightarrow[n\to\infty]{\mathbb{P}} h^{-1}\big(\mathbb{E}_\theta\, g\left(X_1\right)\big) = h^{-1}\big(h(\theta)\big) = \theta.$$

$\blacksquare$

Определение. *Асимптотически нормальная оценка* параметра θ с коэффициентом $\sigma^2(\theta)$ — оценка θ^*, т.ч. при $n \to \infty$ имеет место слабая сходимость к стандартному нормальному распределению:

$$\frac{\sqrt{n}(\theta^* - \theta)}{\sigma(\theta)} \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}_{0,1}\,.$$

Лемма. *Пусть функция $g(y)$ такова, что $0 \neq \mathbb{D}_\theta\, g\left(X_1\right) < \infty$. Тогда статистика $\overline{g(\mathbf{X})}$ является асимптотически нормальной оценкой для $\mathbb{E}_\theta\, g\left(X_1\right)$ с коэффициентом $\sigma^2(\theta) = \mathbb{D}_\theta\, g\left(X_1\right)$:*

$$\sqrt{n}\,\frac{\overline{g(\mathbf{X})} - \mathbb{E}_\theta\, g\left(X_1\right)}{\sqrt{\mathbb{D}_\theta\, g\left(X_1\right)}} \xrightarrow[n\to\infty]{d} \mathrm{N}_{0,1}$$

Доказательство. Следует непосредственно из ЦПТ. ■

Замечание. Следующая теорема утверждает асимптотическую нормальность оценок вида

$$\theta^* = H\big(\overline{g(\mathbf{X})}\big) = H\left(\frac{g\left(X_1\right) + \ldots + g\left(X_n\right)}{n}\right)$$

которые обычно получаются при использовании метода моментов, при этом всегда $\theta = H\big(\mathbb{E}_\theta\, g\left(X_1\right)\big)$.

Утверждение. *Пусть функция $g(y)$ такова, что $0 \neq \mathbb{D}_\theta\, g\left(X_1\right) < \infty$, функция $H(y)$ дифференцируема в точке $a = \mathbb{E}_\theta\, g\left(X_1\right)$ и её производная в этой точке $H'(a) = H'(y)\big|_{y=a}$ отлична от нуля. Тогда оценка $\theta^* = H\big(\overline{g(\mathbf{X})}\big)$ является асимптотически нормальной оценкой для параметра $\theta = H\big(\mathbb{E}_\theta\, g\left(X_1\right)\big) = H(a)$ с коэффициентом асимптотической нормальности $\sigma^2(\theta) = \big(H'(a)\big)^2 \cdot \mathbb{D}_\theta\, g\left(X_1\right)$.*

Доказательство. Согласно ЗБЧ последовательность $\overline{g(\mathbf{X})}$ стремится к $a = \mathbb{E}_\theta\, g\left(X_1\right)$ по вероятности с ростом n: Функция

$$G(y) = \begin{cases} \dfrac{H(y) - H(a)}{y - a}, & y \neq a \\ H'(a), & y = a \end{cases}$$

по условию непрерывна в точке a: поскольку сходимость по вероятности сохраняется под действием непрерывной функции, получим, что $G\big(\overline{g(\mathbf{X})}\big) \xrightarrow[n\to\infty]{\mathbb{P}} G(a) = H'(a)$.

Заметим также, что по вышеприведённой лемме величина $\sqrt{n}\big(\overline{g(\mathbf{X})} - a\big)$ слабо сходится к нормальному распределению $\mathrm{N}_{0,\mathbb{D}_\theta\, g(X_1)}$: пусть ξ — случайная величина из этого распределения. Тогда

$$\sqrt{n}\Big(H\big(\overline{g(\mathbf{X})}\big) - H(a)\Big) = \sqrt{n}\big(\overline{g(\mathbf{X})} - a\big) \cdot G\big(\overline{g(\mathbf{X})}\big) \xrightarrow[n\to\infty]{\mathrm{d}} \xi \cdot H'(a)$$

Мы использовали следующее свойство слабой сходимости: если $\xi_n \xrightarrow[n\to\infty]{\mathrm{d}} \xi$ и $\eta_n \xrightarrow[n\to\infty]{\mathbb{P}} c = const$, то $\xi_n\eta_n \xrightarrow[n\to\infty]{\mathrm{d}} c\xi$. Но распределение случайной величины $\xi \cdot H'(a)$ есть $\mathrm{N}_{0,(H'(a))^2\cdot\mathbb{D}_\theta\, g(X_1)}$, откуда следует

$$\sigma^2(\theta) = \big(H'(a)\big)^2 \cdot \mathbb{D}_\theta\, g\,(X_1)\,.$$

■

2.8 Метод максимального правдоподобия. Свойства оценок максимального правдоподобия

Определение. *Оценка максимального правдоподобия* $\theta^*(\mathbf{X})$ *параметра* θ — точка параметрического множества Θ, в которой функция правдоподобия $L(\mathbf{X},\theta)$ при заданной реализации выборки $\boldsymbol{x}$ достигает максимума, т.е.:

$$L(\boldsymbol{x},\theta^*) = \max_{\theta\in\Theta} L(\boldsymbol{x},\theta)$$

Замечание. Поскольку функция $\ln y$ монотонна, то точки максимума функций $L(\mathbf{X},\theta)$ и $\ln L(\mathbf{X},\theta)$ совпадают.

Если для каждого X максимум функции правдоподобия достигается во внутренней точке Θ, и $L(\mathbf{X},\theta)$ дифференцируема по θ, то оценка максимального правдоподобия $\theta^* = \theta^*(\mathbf{X})$ удовлетворяет уравнению:

$$\frac{\partial \ln L(\boldsymbol{x},\theta)}{\partial\theta} = 0$$

Если θ — векторный параметр: $\theta = (\theta_1,\ldots,\theta_n)$, то это уравнение заменяется системой уравнений:

$$\frac{\partial \ln L(\boldsymbol{x},\theta)}{\partial\theta_i} = 0,\ i=\overline{1,n}$$

Пример. Пусть дана выборка $\mathbf{X} = \big(X_1,\ldots,X_n\big)$, $X_i \sim \mathrm{Bi}_{1,\theta}$. Найдём оценку максимального правдоподобия для θ. Для этого запишем функцию правдоподобия и исследуем её (точнее, её логарифм) на экстремум:

$$L(\mathbf{X},\theta) = \prod_{i=1}^n \mathbb{P}_\theta\big(X_i = x_i\big) = \theta^{\sum\limits_{i=1}^n X_i}(1-\theta)^{n-\sum\limits_{i=1}^n X_i}$$

$$\ln L(\mathbf{X},\theta) = \sum_{i=1}^n X_i \ln\theta + \left(n - \sum_{i=1}^n X_i\right)\ln(1-\theta)$$

$$\frac{\partial}{\partial\theta}\ln L(\mathbf{X},\theta) = \frac{\sum\limits_{i=1}^n X_i}{\theta} - \frac{n-\sum\limits_{i=1}^n X_i}{1-\theta} =$$

$$= \frac{(1-\theta)\sum\limits_{i=1}^{n} X_i - \theta\left(n - \sum\limits_{i=1}^{n} X_i\right)}{\theta(1-\theta)} =$$

$$= \frac{\sum\limits_{i=1}^{n} X_i - n\theta}{\theta(1-\theta)}$$

Приравнивая числитель к нулю, получаем, что $\theta^* = \overline{X}$ — стационарная точка. При этом если $\theta > \overline{X}$, то производная отрицательна, а если $\theta < \overline{X}$ — положительна, т.е. это локальный максимум. Значит, это в самом деле оценка максимального правдоподобия для параметра θ.

Утверждение. *Если существует эффективная оценка $\widetilde{T}(\mathbf{X})$ скалярного параметра θ, то она совпадает с оценкой максимального правдоподобия.*

Доказательство. Если оценка $\widetilde{T}(\mathbf{X})$ скалярного параметра θ эффективна, то $\widetilde{T}(\mathbf{X})$ линейно выражается через функцию вклада (следствие неравенства Рао—Крамера):

$$\widetilde{T}(\mathbf{X}) - \theta = a_n(\theta)U(\mathbf{X},\theta) = a_n(\theta)\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}$$

Но если мы подставим вместо θ оценку максимального правдоподобия θ^*, то правая часть обнулится. А значит, $\widetilde{T}(\mathbf{X}) = \theta^*$. ■

Утверждение. *Если $T(\mathbf{X})$ — достаточная статистика, а оценка максимального правдоподобия θ^* существует и единственна, то она является функцией от $T(\mathbf{X})$.*

Доказательство. Из критерия факторизации следует, что если $T = T(\mathbf{X})$ достаточная статистика, то имеет место представление:

$$L(\mathbf{X},\theta) = g(T(\mathbf{X}),\theta)h(\mathbf{X})$$

Таким образом, максимизация $L(\mathbf{X},\theta)$ сводится к максимизации $g\big(T(\mathbf{X}),\theta\big)$ по θ. Следовательно, θ^* есть функция от $T(\mathbf{X})$. ■

Инвариантность метода максимального правдоподобия. *Пусть $f\colon \Theta \mapsto Y$ — некоторая биективная функция. Тогда, если θ^* есть оценка максимального правдоподобия для θ, то $f(\theta^*)$ — оценка максимального правдоподобия для $f(\theta)$.*

Доказательство.

$$L(\boldsymbol{x}, \theta^*) = \sup_{\theta \in \Theta} L(\boldsymbol{x}, \theta) = \sup_{y \in Y} L(\boldsymbol{x}, f^{-1}(y)) = L(\boldsymbol{x}, f^{-1}(y^*))$$

Тогда $\theta^* = f^{-1}(y^*)$ и $y^* = f(\theta^*)$. ■

Определение. *Асимптотически эффективная оценка* параметра $\tau(\theta)$ — оценка τ^*:

$$\mathbb{D}_\theta\, \tau^* \cdot \frac{i_n(\theta)}{\big(\tau'(\theta)\big)^2} \xrightarrow[n \to \infty]{} 1, \text{ где } i_n(\theta) = \mathbb{E}_\theta \left(\frac{\partial \ln L(X, \theta)}{\partial \theta} \right)^2 = \mathbb{E}_\theta\big(U^2\big(X, \theta\big)\big)$$

Утверждение. *Пусть выполнены следующие условия:*

1. *Функция правдоподобия $L(\mathbf{X}, \theta)$ удовлетворяет условиям регулярности для первых двух производных;*

2. *Существует единственная оценка максимального правдоподобия θ^* для всех θ, которая достигается во внутренней точке Θ.*

Тогда оценка θ^*:

1. *асимптотически несмещена;*

2. *состоятельна;*

3. *асимптотически эффективна;*

4. *асимптотически нормальна.*

2.9 Интервальное оценивание: Центральная статистика и использование точечной оценки

Определение. *Доверительный интервал* для параметра θ с коэффициентом доверия (или надёжности) $\gamma \in (0, 1)$ — интервал $\big(T_1(\mathbf{X}), T_2(\mathbf{X})\big)$, где T_i — статистики, т.ч.:

1. $T_1(\mathbf{X}) \overset{\text{п.н.}}{\leqslant} T_2(\mathbf{X})$;

2. $\mathbb{P}_\theta\Big(T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})\Big) \geqslant \gamma$.

Пример. Пусть $\big(X_1, \ldots, X_n\big)$ — выборка из $\mathrm{N}_{\theta,1}$. Тогда

$$\theta^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}_{\theta,\frac{1}{n}} \xrightarrow[n\to\infty]{\mathrm{d}} (\overline{X}-\theta)\sqrt{n} \sim \mathrm{N}_{0;1}$$

Для величины, имеющей стандартное нормальное распределение, можно построить доверительный интервал следующим образом: находим такое $t_{\gamma/2}$, что

$$\mathbb{P}_\theta\Big(\big|(\overline{X}-\theta)\sqrt{n}\big| \leqslant t_{\gamma/2}\Big) = \gamma.$$

Решаем уравнение относительно θ и получаем

$$\mathbb{P}_\theta\left(\overline{X} - \frac{t_{\gamma/2}}{\sqrt{n}} \leqslant \theta \leqslant \overline{X} + \frac{t_{\gamma/2}}{\sqrt{n}}\right) = \gamma.$$

Таким образом, $\left(\overline{X} - \dfrac{t_{\gamma/2}}{\sqrt{n}},\ \overline{X} + \dfrac{t_{\gamma/2}}{\sqrt{n}}\right)$ — γ-доверительный интервал для параметра θ.

Длина построенного нами в примере интервала — $2\,\dfrac{t_{\gamma/2}}{\sqrt{n}}$. Она зависит лишь от размера выборки. Однако в общем случае длина доверительного интервала является случайной величиной — $T_2(X) - T_1(X)$. Вместе с тем очевидно, что нам хотелось бы сделать интервал как можно более коротким. Рассмотрим достотаточно общий способ построения кратчайших доверительных интервалов. Для этого нам понадобится ввести следующее определение.

Определение. *Центральная статистика* — функция $G(X,\theta)$, т.ч.:

1. $G(X,\theta)$ непрерывна и строго монотонна по θ при любом фиксированной реализации выборки $\boldsymbol{x}$.
2. $F_G(t) \equiv \mathbb{P}_\theta\big(G(X,\theta) < t\big)$ непрерывна и не зависит от θ.

Замечание. Формально определённая выше величина не является статистикой, т.к. зависит от неизвестного параметра θ.

Построение кратчайшего доверительного интервала с помощью центральной статистики:

1. Найдём $g_1, g_2 \in \mathbb{R}$, т.ч.

$$\mathbb{P}_\theta\Big(g_1 \leqslant G(X,\theta) \leqslant g_2\Big) = \gamma \ \ \forall\theta\in\Theta \quad\iff\quad F_G(g_2+0) - F_G(g_1) = \gamma.$$

2. Пусть для определённости $G(X,\theta)$ возрастает по θ. Тогда существует обратная по θ функция $G^{-1}(X,g)$ и из условий

$$\begin{cases} G(X,\theta) \leqslant & g_2 \\ G(X,\theta) \geqslant & g_1 \end{cases}$$

можно найти статистики

$$\begin{cases} T_2(\mathbf{X})\colon G\big(X, T_2(\mathbf{X})\big) = g_2 \\ T_1(\mathbf{X})\colon G\big(X, T_1(\mathbf{X})\big) = g_1 \end{cases} \iff T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})$$

откуда $\mathbb{P}_\theta\Big(T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})\Big) \geqslant \gamma \ \forall\theta \in \Theta$.

3. Минимизируем длину получившегося интервала. В общем случае это сводится к поиску условного экстремума

$$\Big|G^{-1}(\mathbf{X}, g_2) - G^{-1}(\mathbf{X}, g_1)\Big| = \Big|T_2(\mathbf{X}) - T_1(\mathbf{X})\Big| \to \min_{F_G(g_2+0)-F_G(g_1)\geqslant\gamma}$$

Поиск условного экстремума зачастую является сложной задачей. Порой вместо минимизации длины интервала минимизируют среднюю длину $\mathbb{E}_\theta\big[T_2(\mathbf{X}) - T_1(\mathbf{X})\big]$, или даже отношение $\frac{g_2}{g_1}$ (считая, что $g_2 > g_1$).

Можно использовать и другой подход, который позволяет избежать решения сложных оптимизационных задач. Согласно определению, параметр попадает в доверительный интервал с вероятностью γ. Мы можем распределить оставшиеся $1-\gamma$ поровну, чтобы параметр был левее или правее нашего интервала с равной вероятностью.

Определение. *Центральный доверительный интервал* для параметра θ с коэффициентом доверия $\gamma \in (0,1)$ — интервал $\big(T_1(\mathbf{X}), T_2(\mathbf{X})\big)$, т.ч.:

$$\mathbb{P}_\theta\Big(T_1(\mathbf{X}) > \theta\Big) = \frac{1-\gamma}{2}$$
$$\mathbb{P}_\theta\Big(T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})\Big) = \gamma$$
$$\mathbb{P}_\theta\Big(T_2(\mathbf{X}) < \theta\Big) = \frac{1-\gamma}{2}$$

В этом случае при наличии центральной статистики мы однозначно находим g_1, g_2 как квантили распределения $F_G(x)\colon g_1 = F_G^{-1}\left(\frac{1-\gamma}{2}\right),\ g_2 = F_G^{-1}\left(\frac{1+\gamma}{2}\right)$, и остаётся лишь найти $T_1(\mathbf{X}) = G^{-1}(\mathbf{X}, g_1),\ T_2(\mathbf{X}) = G^{-1}(\mathbf{X}, g_2)$ (или наоборот, если $G(\mathbf{X},\theta)$ убывает по θ).

Построение доверительного интервала с помощью точечной оценки

Порой бывает невозможно найти центральную статистику, так как функция распределения имеет слишком сложный вид — например, в случае бернуллиевского или пуассоновского распределения. В таком случае можно использовать метод точечной оценки.

1. Пусть $T(\mathbf{X})$ — точечная оценка θ. Обозначим $F_T(t, \theta) = \mathbb{P}_\theta\Big(T(\mathbf{X}) < t\Big)$. Предположим, что $F_T(t, \theta)$ — непрерывная и строго монотонная функция θ при любом фиксированном t. Найдём такие $t_1(\theta), t_2(\theta)$, что $\mathbb{P}_\theta\Big(t_1(\theta) \leqslant T(\mathbf{X}) \leqslant t_2(\theta)\Big) \geqslant \gamma$. Для однозначности будем выбирать их так, чтобы

$$\begin{cases} \mathbb{P}_\theta\Big(T(\mathbf{X}) > t_2(\theta)\Big) \leqslant \dfrac{1-\gamma}{2} \\ \mathbb{P}_\theta\Big(T(\mathbf{X}) < t_1(\theta)\Big) \leqslant \dfrac{1-\gamma}{2} \end{cases} \iff \begin{cases} 1 - F_T\big(t_2(\theta), \theta\big) \leqslant \dfrac{1-\gamma}{2} \\ F_T\big(t_1(\theta), \theta\big) \leqslant \dfrac{1-\gamma}{2} \end{cases}$$

Т.е. речь идёт о построении центрального доверительного интервала. Подразумевается, что $t_1(\theta)$ — наибольшее, а $t_2(\theta)$ — наименьшее значения $T(\mathbf{X})$, удовлетворяющие этим требованиям. Если наблюдаемая случайная величина имеет непрерывное распределение, то неравенства заменяются на равенства.

2. Рассмотрим вспомогательную лемму, верную в абсолютно непрерывном случае:

Лемма. *Если $F_T\big(t, \theta\big)$ возрастает по θ, то $t_1(\theta)$ и $t_2(\theta)$ убывают. Если же $F_T\big(t, \theta\big)$ убывает по θ, то $t_1(\theta)$ и $t_2(\theta)$ возрастают.*

Доказательство. Докажем для t_1, для t_2 аналогично. Пусть $F_T(t, \theta)$ возрастает. От противного: предположим, что $\exists \theta_1 < \theta_2 \colon t_1\left(\theta_1\right) \leqslant t_1\left(\theta_2\right)$. Используя то, что $F_T\big(t, \theta\big)$ возрастает по θ и, как всякая функция распределения, не убывает по t, получим:

$$\frac{1-\gamma}{2} = F_T\big(t_1(\theta_1), \theta_1\big) < F_T\big(t_1(\theta_1), \theta_2\big) \leqslant F_T\big(t_1(\theta_2), \theta_2\big) = \frac{1-\gamma}{2}$$

Полученное противоречие завершает доказательство. ■

3. Заключительный шаг:

$$t_1(\theta) < T(\mathbf{X}) \iff \theta > \varphi_1\big(T(\mathbf{X})\big) \implies \mathbb{P}_\theta\Big(\theta > \varphi_1\big(T(\mathbf{X})\big)\Big) = \frac{1-\gamma}{2}$$

$$t_2(\theta) > T(\mathbf{X}) \iff \theta < \varphi_2\big(T(\mathbf{X})\big) \implies \mathbb{P}_\theta\Big(\theta < \varphi_2\big(T(\mathbf{X})\big)\Big) = \frac{1-\gamma}{2}$$

$$\implies \mathbb{P}_\theta\Big(\underbrace{\varphi_2\big(T(\mathbf{X})\big)}_{T_1(\mathbf{X})} \leqslant \theta \leqslant \underbrace{\varphi_1\big(T(\mathbf{X})\big)}_{T_2(\mathbf{X})}\Big) = \gamma$$

где φ_1, φ_2 — обратные к $t_1(\theta), t_2(\theta)$ функции. Доказанная выше лемма является, по существу, обоснованием применимости метода в абсолютно непрерывном случае — ведь если t_1, t_2 строго монотонны, то они обратимы.

В дискретном случае придётся проверять обратимость вручную.

2.10 Проверка гипотез. Лемма Неймана—Пирсона

Определение. *Статистическая гипотеза* H — любое предположение о распределении наблюдаемой случайной величины.

Гипотеза называется *простой*, если в ней явно задаётся одно (не параметризованное) распределение. Например, $H\colon X_i \sim \mathrm{N}_{0,1}$. В противном случае гипотеза называется *сложной*.

Как правило, рассматривается сразу две взаимоисключающие гипотезы. Одна из них называется *основной* и обозначается H_0, а другая — *альтернативной* и обозначается H_1.

Гипотезы могут быть самыми разнообразными. Приведём несколько примеров.

1. Гипотезы о виде распределения:
   Пусть производится n независимых наблюдений над некоторой случайной величиной ξ с неизвестной функцией распределения $F_\xi(x)$. Основная гипотеза — $H_0\colon F_\xi(x) = F(x)$ или $H_0\colon F_\xi(x) \in \mathcal{F}$, где $\mathcal{F}$ — некоторое подмножество в множестве всех распределений (как правило, $\mathcal{F}$ задаётся параметрически).

2. Гипотезы о проверке однородности выборки:
   Произведено k серий независимых наблюдений и получено k выборок $\mathbf{X}_1, \ldots, \mathbf{X}_k$. Основная гипотеза состоит в том, что эти выборки извлечены из одного распределения, т.е. $H_0\colon F_1(x) \equiv \ldots \equiv F_k(x)$, где F_i — функция распределения элементов i-й выборки.

3. Гипотеза независимости:
   Наблюдается двумерная случайная величина $\xi = (\xi_1, \xi_2)$ с неизвестной функцией распределения $F(x, y)$. Получена выборка $(X_1, Y_1), \ldots, (X_n, Y_n)$. Основная гипотеза заключается в том, что ξ_1 и ξ_2 независимы, то есть $H_0\colon F(x, y) = F_{\xi_1}(x) F_{\xi_2}(y)$, где F_{ξ_1}, F_{ξ_2} — некоторые одномерные функции распределения. В общем случае можно рассматривать k-мерную случайную величину и проверять гипотезу независимости её компонент.

4. Гипотеза случайности:
   Результат эксперимента описывается n-мерной случайной величиной $X = \big(X_1, \ldots, X_n\big)$ с неизвестной функцией распределения $F_X(x_1, \ldots, x_n)$. Можно ли рассматривать X как выборку из распределения некоторой случайной величины ξ (т.е. являются ли компоненты X_i независимыми и одинаково распределёнными)? В данном случае проверяется гипотеза случайности $H_0\colon F_X(x_1, \ldots, x_n) = F_\xi(x_1) \cdot \ldots \cdot F_\xi(x_n)$.

Замечание. Может показаться, что в общем случае задача имеет такой вид: «Дана выборка $\mathbf{X}$. Верна ли гипотеза H_0?» Но корректнее говорить «Согласуются ли наблюдаемые данные с высказанной гипотезой?».

В дальнейшем мы увидим, что процесс проверки гипотезы скорее напоминает доказательство неверности H_0. Мы оцениваем, насколько вероятна имеющаяся реализация выборки в предположении, что H_0 истинна. Если наблюдаемые значения достаточно маловероятны, то мы считаем, что гипотеза H_0 неправдоподобна и отвергаем её. В противном случае мы лишь можем сказать, что данные не противоречат высказанной гипотезе.

Приведём модельный пример: пусть есть выборка из одного элемента $\mathbf{X} = X_1$ и проверяется простая гипотеза $H_0\colon X_1 \sim U[-2, 2]$. Если наблюдаемая реализация выборки $x_1 = 3$, то наша гипотеза очевидно неверна, и мы можем спокойно её отвергнуть. Но если $x_1 = 1$, то мы не можем утверждать, что H_0 справедлива: мы могли получить $x_1 = 1$ и из стандартного нормального распределения, и из пуассоновского, или из отрицательного биномиального.

В приведённом выше замечании мы уже сформулировали одну гипотезу, и, по сути, даже проверили её, интуитивно построив алгоритм принятия решения: если $x_1 \notin [-2, 2]$, то H_0 отвергается, иначе оснований считать её неверной нет. Для того, чтобы проверять более сложные гипотезы, нам понадобится строить похожие алгоритмы, или критерии — правила, согласно которым *по выборке* делается заключение о верности гипотезы.

Определение. Критерий — это статистика $\varphi(\mathbf{X})$ (т.е. измеримая функция от выборки) со значениями из $[0, 1]$. Трактуется как «вероятность» отвергнуть H_0.

Если $\varphi(\mathbf{X}) \overset{\text{п.н.}}{\in} \{0,1\}$, то критерий называется *нерандомизированным*, иначе — *рандомизированным.*

Замечание. С нерандомизированным критерием всё понятно: на каждой реализации выборки он даёт однозначный ответ, принять ($\varphi(\boldsymbol{x}) = 0$) или отвергнуть ($\varphi(\boldsymbol{x}) = 1$) гипотезу. Но что делать, если $\varphi(\boldsymbol{x}) \in (0,1)$? Тогда мы отвергаем гипотезу H_0 с вероятностью $\varphi(\boldsymbol{x})$. По сути, мы принимаем решение, подбрасывая (асимметричную) монетку. Понятно, что нерандомизированные критерии лучше, но, к сожалению, некоторые задачи не получается решить, отвечая лишь «да» и «нет».

Каким образом строится критерий?

Выборка $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ объёма n — точка в пространстве $\mathbb{R}^n$. Выделим такое множество $S \subset \mathbb{R}^n$, что $\mathbb{P}_\theta\Big(\mathbf{X} \in S \,|\, \{H_0 \text{ верна}\}\Big) \leqslant \alpha$, где $\alpha \in (0,1)$ — некоторое наперёд заданное число. Это множество называется *критической областью* для гипотезы H_0. В этом случае критерий можно сформулировать следующим образом:

- $\mathbf{X} \in S \implies$ отвергаем H_0;
- $\mathbf{X} \notin S \implies$ нет оснований отвергать H_0, считаем её верной.

Или $\varphi(\mathbf{X}) = \mathrm{I}(\mathbf{X} \in S)$ *(нерандомизированный!).*

По смыслу критическая область — это множество таких значений выборки, которые маловероятны при условии истинности H_0. Поэтому при попадании выборки в критическую область основная гипотеза отвергается, как противоречащая статистическим данным.

Конечно, если мы построили критерий и выбрали какую-то гипотезу, это не значит, что она стопроцентно верна. Может оказаться, что мы отвергнули верную, или приняли ложную гипотезу. Ошибки при проверке гипотез делятся на два типа.

Определение. Говорят, что произошла *ошибка 1-го рода (false positive)*, если критерий отверг верную гипотезу H_0. Вероятность ошибки 1-го рода:

$$\alpha(S) = \mathbb{P}_\theta\left(\mathbf{X} \in S \middle|\, H_0\right) = \mathbb{P}\,0\left(\mathbf{X} \in S\right)$$

Аналогично вероятность *ошибки 2-го рода (false negative)*:

$$\beta(S) = \mathbb{P}_\theta\left(\mathbf{X} \notin S \middle|\, H_1\right) = \mathbb{P}\,1\left(\mathbf{X} \notin S\right)$$

Замечание. Вероятность ошибки 1-го рода также называется *уровнем значимости критерия.* Выше мы говорили о том, что критическая область —

это множество тех реализаций выборки, которые маловероятны при истинности основной гипотезы. При помощи $\alpha = \alpha(S)$ мы как раз и определяем, насколько маловероятно это событие. Естественным образом получается, что вероятность отвергнуть верную основную гипотезу равна α.

Истинная гипотеза	Результат принятия решения	
	H_0 принята	H_0 отклонена
H_0	$1-\alpha$	α
H_1	β	$1-\beta$

Если $1 - \beta(S) < \alpha(S)$, то попасть в S при условии истинности гипотезы H_1 труднее, чем при условии истинности гипотезы H_0, и тогда S — критическая область скорее для H_1. Хотелось бы, чтобы неравенство имело вид $1 - \beta(S) \geqslant \alpha(S)$.

Определение. Критерий называется *несмещённым*, если выполняется условие

$$1 - \beta(S) \geqslant \alpha(S).$$

В общем случае не получается сделать сумму вероятностей ошибок обоих родов сколь угодно малой, так как задачи минимизации каждой из ошибок, как правило, противоречат друг другу. Например, если наш критерий всегда будет отвергать основную гипотезу — $\varphi(\mathbf{X}) \equiv 1$, то вероятность ошибки второго рода равна нулю. И наоборот, всегда принимая основную гипотезу — $\varphi(\mathbf{X}) \equiv 0$, мы никогда не совершим ошибку первого рода.

Конечно, можно привести пример, где есть идеальный критерий. Пусть мы знаем, что наблюдаемая случайная величина распределена равномерно либо на отрезке $[0,1]$ (основная гипотеза), либо на $[2,3]$ (альтернатива). Тогда критической областью будет $S = [2,3]$, и критерий $\varphi(\mathbf{X}) = \mathrm{I}(x_1 \in [2,3])$ будет иметь нулевые вероятности ошибок обоих родов. Причиной тому то, что области значений исследуемых случайных величин не пересекаются. Однако такой пример весьма далёк от реальных ситуаций.

Если мы не можем минимизировать и то и другое, логично зафиксировать один из параметров и оптимизировать оставшийся. Обычно фиксируют уровень значимости, т.е. вероятность ошибки первого рода α.

Рассмотрим так называемые *параметрические гипотезы.* Считаем, что выборка $\mathbf{X}$ взята из некоторого параметрического семейства $\mathcal{F} = \{F_\theta(x),\ \theta \in \Theta \subseteq \mathbb{R}^r\},\ \theta = (\theta_1, \ldots, \theta_r)$.

Основная гипотеза — $H_0\colon \theta \in \Theta_0,\ \Theta_0 \subset \Theta.$

Альтернатива — $H_1\colon \theta \in \Theta_1,\ \Theta_1 = \Theta \setminus \Theta_0.$

Определение. *Мощность критерия*:

$$W(\theta, \varphi) = \mathbb{E}_\theta\, \varphi(\mathbf{X}) = \int\limits_{x \in \mathbb{R}^n} \varphi(x) L(\mathbf{x}, \theta)\, dx.$$[1]

Замечание. Заметим, что если критерий $\varphi(\mathbf{X})$ нерандомизированный (т.е. принимает только значения 0 и 1, однозначно решая, отвергнуть или принять основную гипотезу), то

$$W(\theta, \varphi) = \mathbb{E}_\theta\, \varphi(\mathbf{X}) = 1 \cdot \mathbb{P}_\theta(\mathbf{X} \in S) + 0 \cdot \mathbb{P}_\theta(\mathbf{X} \notin S) = \mathbb{P}_\theta(\mathbf{X} \in S).$$

Если $\theta \in \Theta_0$, то $\mathbb{P}_\theta(\mathbf{X} \in S | H_0) = \alpha(S)$.
Если $\theta \in \Theta_1$, то $\mathbb{P}_\theta(\mathbf{X} \in S | H_1) = 1 - \mathbb{P}_\theta(\mathbf{X} \notin S | H_1) = 1 - \beta(S)$. Т.е. ошибки и первого, и второго рода можно выразить через функцию мощности.

Разберёмся сначала с простыми гипотезами: $H_0\colon \theta = \theta_0, H_1\colon \theta = \theta_1$. Зададим α_0 и будем иметь дело только с такими критериями, где $\alpha_0 \geqslant \alpha(S)$ (т.е. вероятность ошибки первого рода не превосходит величины α_0) и будем решать задачу $\beta(S) \to \min\limits_S$, что равносильно максимизации мощности при $\theta \in \Theta_1$ (в нашем случае $\Theta_1 = \{\theta_1\}$, и мы максимизируем $W(\theta_1, \varphi)$).

Получаем две эквивалентные задачи определения критической области S:

$$\begin{cases} \alpha(S) \leqslant \alpha_0 \\ \beta(S) \to \min\limits_S \end{cases} \iff \begin{cases} W(\theta_0, \varphi) = \mathbb{P}\,\theta_0(\mathbf{X} \in S | H_0) \leqslant \alpha_0 \\ W(\theta_1, \varphi) = \mathbb{P}\,\theta_1(\mathbf{X} \in S | H_1) \to \max\limits_S \end{cases}$$

К сожалению, не всегда удаётся решить такую задачу, используя только нерандомизированные критерии $\varphi(\mathbf{X}) = \mathrm{I}(\mathbf{X} \in S)$. Тяжело вздохнём и обратимся к рандомизированным.

Пусть дана выборка $\mathbf{X} = \big(X_1, \ldots, X_n\big)$, относительно которой выдвинуто две простых параметрических гипотезы: $H_0\colon \theta = \theta_0$ и $H_1\colon \theta = \theta_1$. Без ограничения общности будем предполагать, что существует плотность $f_0(x)$ для функции распределения $F_0(x) = F_{\theta_0}(x)$, соответствующей гипотезе $\theta = \theta_0$, и существует плотность $f_1(x)$ для функции распределения $F_1(x) = F_{\theta_1}(x)$. В дискретном случае все результаты аналогичны.

Если верна гипотеза H_1, то функция правдоподобия выборки X имеет вид:

$$L(\mathbf{X}, \theta_1) = L_1\left(\mathbf{X}\right) = \prod_{i=1}^{n} f_1\left(X_i\right).$$

[1]Напоминаем, что при фиксированном θ функция правдоподобия является вероятностной мерой на выборочном пространстве.

В противном случае

$$L(\mathbf{X}, \theta_0) = L_0(\mathbf{X}) = \prod_{i=1}^{n} f_0(X_i).$$

Замечание. Вероятность ошибки первого и второго рода для рандомизированного критерия будем обозначать $\alpha(\varphi)$ и $\beta(\varphi)$ соответственно.

Посчитаем вероятность ошибок, используя то, что при каждом $x \in \mathbb{R}^n$ значение $\varphi(x)$ — это вероятность отвергнуть H_0.

$$\mathbb{P}_\theta\big(H_0 \text{ отвергнута} | H_0 \text{ верна}\big) = \mathbb{P}\,0\left(\overline{H}_0\right) = \int\limits_{x \in \mathbb{R}^n} \varphi(x) L_0(x)\, dx = \alpha(\varphi)$$

$$\mathbb{P}_\theta\big(H_0 \text{ принята} | H_0 \text{ неверна}\big) = \mathbb{P}\,1\left(H_0\right) = \int\limits_{x \in \mathbb{R}^n} (1 - \varphi(x)) L_1(x)\, dx = \beta(\varphi)$$

Т.е.

$$\begin{aligned} W(\theta_0, \varphi) &= \alpha(\varphi) \\ W(\theta_1, \varphi) &= 1 - \beta(\varphi) \end{aligned}$$

Тогда задача построения статистического критерия будет формулироваться следующим образом:

$$\begin{cases} \alpha(\varphi) \leqslant \alpha_0 \\ \beta(\varphi) \to \min\limits_{\varphi} \end{cases} \iff \begin{cases} W(\theta_0, \varphi) = \mathbb{E}_{\theta_0} \varphi(\mathbf{X}) = \mathbb{E}_0\, \varphi(\mathbf{X}) \leqslant \alpha_0 \\ W(\theta_1, \varphi) = \mathbb{E}_{\theta_1} \varphi(\mathbf{X}) = \mathbb{E}_1\, \varphi(\mathbf{X}) \to \max\limits_{\varphi} \end{cases}$$

Таким образом, задача заключается в том, чтобы найти наиболее мощный критерий, когда вероятность ошибки первого рода не превосходит некоторого заданного порогового значения. Решение сформулированных задач даётся леммой Неймана—Пирсона.

Лемма Неймана—Пирсона. *Пусть $\alpha_0 \in (0, 1)$. Введём отношение функций правдоподобия $l(\mathbf{X}) = \dfrac{L_1(\mathbf{X})}{L_0(\mathbf{X})}$. Тогда при фиксированной вероятности ошибки первого рода α_0 наиболее мощный критерий (сокращенно НМК) име-*

ет вид

$$\varphi^*(x) = \begin{cases} 1, & \text{\textit{если} } l(\mathbf{X}) > C \\ \varepsilon, & \text{\textit{если} } l(\mathbf{X}) = C \\ 0, & \text{\textit{если} } l(\mathbf{X}) < C \end{cases}$$

где константы C и ε являются решениями уравнения «вероятность ошибки первого рода для этого критерия равна α_0»: $\alpha\left(\varphi^*\right) = \alpha_0$.

Замечание. Если для некоторой реализации выборки получится так, что $L_0(\boldsymbol{x}) = 0, L_1(\boldsymbol{x}) > 0$, то просто будем считать, что $l(\boldsymbol{x}) = +\infty > C$, и гипотезу H_0 надо отвергнуть.

Доказательство.

1. Покажем, что константы C и ε могут быть найдены из уравнения $\alpha\left(\varphi^*\right) = \alpha_0$. Заметим, что

$$\alpha(\varphi^*) = P_0\Big(l(\mathbf{X}) > C\Big) + \varepsilon P_0\Big(l(\mathbf{X}) = C\Big)$$

Отношение правдоподобий $l(\mathbf{X})$ — случайная величина. Для удобства обозначим $\eta = l(\mathbf{X})$. Пусть $F_{\eta,H_0}(x)$ — функция распределения этой величины в предположении, что H_0 верна.

Тогда

$$\alpha\left(\varphi^*\right) = 1 - F_{\eta,H_0}(C) + \varepsilon\Big(F_{\eta,H_0}(C) - F_{\eta,H_0}(C-0)\Big)$$

Пусть $g(C) = 1 - F_{\eta,H_0}(C)$, константу C_{α_0} можно выбрать так, чтобы было выполнено неравенство:

$$g(C_{\alpha_0}) \leqslant \alpha_0 \leqslant g(C_{\alpha_0} - 0)$$

Тогда

$$\varepsilon_{\alpha_0} = \begin{cases} 0, & \text{если } g(C_{\alpha_0}) = g(C_{\alpha_0} - 0) \\ \dfrac{\alpha_0 - g\left(C_{\alpha_0}\right)}{g(C_{\alpha_0} - 0) - g(C_{\alpha_0})} \in [0,1], & \text{если } g(C_{\alpha_0}) < g(C_{\alpha_0} - 0) \end{cases}$$

В обоих случаях выполнено равенство:

$$\alpha_0 = g(C_{\alpha_0}) + \varepsilon_{\alpha_0}\Big(g(C_{\alpha_0} - 0) - g(C_{\alpha_0})\Big) = \alpha\left(\varphi^*\right)$$

2. Докажем, что $\varphi^*(x)$ — наиболее мощный критерий.

Выберем любой другой критерий $\tilde{\varphi}(x)$ такой, что $\alpha(\tilde{\varphi}) \leqslant \alpha_0$, и сравним ее с критерием $\varphi^*(x)$. Заметим, что для любого x справедливо неравенство:

$$\big(\varphi^*(x) - \tilde{\varphi}(x)\big)\Big(L_1(x) - c_{\alpha_0} L_0(x)\Big) \geqslant 0$$

Тогда

$$\int\limits_{\mathbb{R}^n} \big(\varphi^*(x) - \tilde{\varphi}(x)\big)\Big(L_1(x) - c_{\alpha_0} L_0(x)\Big)\, dx \geqslant 0$$

Раскроем скобки и преобразуем:

$$\int\limits_{\mathbb{R}^n} \varphi^*(x) L_1(x)\, dx - \int\limits_{\mathbb{R}^n} \tilde{\varphi}(x) L_1(x)\, dx \geqslant$$

$$\geqslant C_{\alpha_0} \left(\int\limits_{\mathbb{R}^n} \varphi^*(x) L_0(x)\, dx - \int\limits_{\mathbb{R}^n} \tilde{\varphi}(x) L_0(x)\, dx \right) \geqslant 0$$

Последнее верно, т.к. $\alpha(\tilde{\varphi}) \leqslant \alpha_0 = \alpha(\varphi^*)$.

Следовательно, $W(\theta_1, \varphi^*) - W(\theta_1, \tilde{\varphi}) \geqslant C_{\alpha_0}\big(\alpha(\varphi^*) - \alpha(\tilde{\varphi})\big)$, откуда получаем неравенство:

$$W(\theta_1, \varphi^*) \geqslant W(\theta_1, \tilde{\varphi})$$

Что и требовалось доказать.

■

Замечание. Все вышесказанное относится к случаю, когда мы проверяем две простые гипотезы: основную $H_0\colon \theta = \theta_0$ и альтернативу $H_1\colon \theta = \theta_1$. В случае сложных гипотез следует зафиксировать некоторые $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$ и применить лемму Неймана—Пирсона уже для простых гипотез. Если в результате построенный критерий не зависит от значения θ_1, то он называется *равномерно наиболее мощным критерием* (сокращённо РНМК).

2.11 Критерии согласия Колмогорова и χ^2

Пусть для наблюдаемого распределения $\mathbb{P}\,\xi$ дана выборка $X_1, \ldots, X_n$, проверяется *гипотеза о виде распределения* $H_0\colon F_\xi = F_0$, где F_0 известна; альтернатива $H_1 : F_\xi \neq F_0$.

Критерий согласия χ^2 Пирсона

Разобьём числовую ось на k промежутков $-\infty = a_0 < a_1 < \ldots < a_k = \infty$, $\Delta_i = (a_{i-1}, a_i]$ и построим статистику $\overline{\chi}^2$:

$$\overline{\chi}^2(\mathbf{X}) = \sum_{i=1}^{k} \frac{(\text{наблюдаемое} - \text{ожидаемое})^2}{\text{ожидаемое}} = \sum_{i=1}^{k} \frac{\left(n_i - np_i^{(0)}\right)^2}{np_i^{(0)}},$$

где n_i — число зафиксированных наблюдений в i-м интервале, $p_i^{(0)} = F_0\left(a_i\right) - F_0\left(a_{i-1}\right)$ — вероятность попадания наблюдения в i-й интервал при выполнении гипотезы H_0, $np_i^{(0)}$, соответственно, ожидаемое число попаданий в i-й интервал.

Формулировка критерия для простой гипотезы:

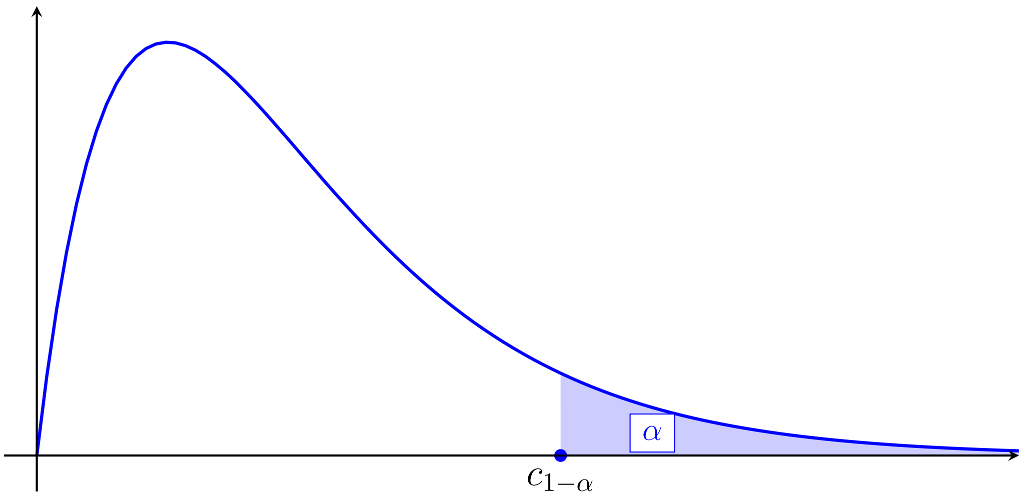
Рассмотрим простую гипотезу $H_0\colon F_\xi(x) = F_0(x)$.

- Если верна гипотеза H_0, то $\overline{\chi}^2\left(\mathbf{X}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \chi^2_{k-1}$, где k — число интервалов разбиения.
- Если верна гипотеза H_1, то $\overline{\chi}^2 \xrightarrow[n\to\infty]{\text{п.н.}} \infty$.

Выберем уровень значимости (т.е. вероятность ошибки первого рода) $\alpha \in (0, 1)$. Область $(\chi^2_{k-1,1-\alpha}, \infty)$, где $\chi^2_{k-1,1-\alpha}$ — квантиль порядка $1-\alpha$ распределения χ^2 с $k-1$ степенями свободы, является критической для гипотезы H_0.

Правило проверки гипотез:

- Если $\overline{\chi}^2\left(\mathbf{X}\right) > \chi^2_{k-1,1-\alpha}$, то H_0 отклоняется;
- Если $\overline{\chi}^2\left(\mathbf{X}\right) \leqslant \chi^2_{k-1,1-\alpha}$, то для отклонения H_0 нет оснований.



Фактически критерий χ^2 проверяет значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот. Рассмотрим его применение на следующем примере.

Пример. При 4040 бросках монеты математик Бюффон получил $n_1 = 2048$ выпадения «герба» и $n_2 = n - n_1 = 1992$ выпадения «решки». Проверим, согласуются ли данные с гипотезой о том, что монета симметрична, т.е. $H_0\colon p = 1/2$. Мы исследуем бернуллиевскую случайную величину $\xi \sim \mathrm{B}_p = \mathrm{Bi}_{1,p}$. Она принимает лишь значения 0 и 1. Выберем промежутки $(-0.5, 0.5)$ и $(0.5, 1.5)$ (или любые другие два, содержащие 0 и 1 соответственно). Согласно предположению, $p = p_1^{(0)} = p_2^{(0)} = 1/2$. Подсчитаем значение статистики $\chi^2_{\mathbf{X}}$:

$$\sum_{i=1}^{k} \frac{(n_i - np_i^{(0)})^2}{np_i^{(0)}} = \frac{(n_1 - np)^2}{np} + \frac{(n_2 - np)^2}{np} = \frac{28^2}{2020} + \frac{28^2}{2020} = \frac{784}{1010} \approx 0.776$$

Положим уровень значимости $\alpha = 0.05$ и найдём квантиль $\chi^2_{k-1,1-\alpha} = \chi^2_{1,0.95} \approx 3.8415$. Сравниваем полученное: $0.776 < 3.8415$. Делаем вывод, что данные не противоречат гипотезе.

Критерий χ^2 для сложной гипотезы.

Рассмотрим сложную гипотезу $H_0\colon F_\xi(x) \in \mathcal{F}_\theta = \big\{F(x,\theta)\colon \theta \in \Theta\big\}$. Как и в простом случае, сгруппируем данные в k интервалов. Теоретические вероятности попадания в интервалы теперь не будут заданы однозначно, а представляют собой некоторые функции от параметра θ. Поэтому статистика имеет вид

$$\overline{\chi^2} = \sum_{i=1}^{n} \frac{\Big(n_i - np_i^{(0)}(\theta)\Big)^2}{np_i^{(0)}(\theta)}$$

Эта статистика зависит от неизвестного параметра; следовательно, непосредственно использовать её для построения критерия пока нельзя — требуется сначала исключить неопределённость, связанную с неизвестным параметром θ. Для этого заменяют θ некоторый оценкой $\tilde{\theta} = \tilde{\theta}(\mathbf{X})$ и получают статистику

$$\overline{\chi^2} = \sum_{i=1}^{n} \frac{\Big(n_i - np_i^{(0)}(\tilde{\theta})\Big)^2}{np_i^{(0)}(\tilde{\theta})}$$

Но, вообще говоря, теперь $p_i^{(0)}(\tilde{\theta})$ являются случайными величинами, и мы не можем утверждать, что распределение статистики $\overline{\chi^2}$ будет стремиться к χ^2_{k-1}. Более того, следует ожидать, что распределение этой статистики (если оно существует) будет зависеть от способа построения оценки $\tilde{\theta}$.

К счастью для нас, английский статистик Рональд Фишер ещё в 1924 году показал, что существуют методы оценивания параметра θ, при котором предельное распределение имеет простой вид, а именно является распределением χ^2_{k-1-r}, где r — размерность оцениваемого параметра. Один из таких методов использует мультиномиальную оценку максимального правдоподобия.

Пример. Следующая задача возникла в связи с бомбардировками Лондона во время Второй мировой войны. Для улучшения организации оборонительных мероприятий, необходимо было понять цель противника. Для этого территорию города условно разделили сеткой из 24 горизонтальных и 24 вертикальных линий на 576 равных участков. В течении некоторого времени в центре организации обороны города собиралась информация о количестве попаданий снарядов в каждый из участков. В итоге были получены следующие данные:

Число попаданий	0	1	2	3	4	5	6	7
Количество участков	229	211	93	35	7	0	0	1

Сформулируем основную гипотезу: стрельба случайна (нет «целевых» участков). В таком случае количество попаданий в участок можно описать распределением Пуассона — оно моделирует число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. Т.е. $H_0\colon F(x) \sim \mathrm{Pois}_\lambda$.[1]

Высчитаем теоретические вероятности:

$$p_i^{(0)} = \mathbb{P}\{S = i\} = \frac{\lambda^i}{i!}e^{-\lambda},$$

где S — число попаданий, $\lambda \approx 0{,}924$ *(мультиномиальная ОМП)*.

Обозначим за n_i количество участков, на которые пришлось i попаданий, и составим новую таблицу для применения критерия.

i	0	1	2	3	4	5	6	7
n_i	229	211	93	35	7	0	0	1
$n_i \cdot p_i^{(0)}$	226,7	211,4	98,5	30,6	7,14	1,33	0,21	0,03
$n_i \cdot \tilde{p}_i^{(0)}$	228,6	211,3	97,6	30,1	8,46			

Прежде чем вычислять статистику $\overline{\chi}^2$, мы объединили 4 последних события с низкими частотами в одно (соответственно, $k = 5$) и пересчитали новые теоретические вероятности $\tilde{p}_i^{(0)}$ и, соответственно, новые ожидаемые значения. В этом случае $\overline{\chi}^2 \approx 1{,}05$. Т.к. $k = 5$, то по таблице распределения χ^2 находим соответствующий уровень значимости $\alpha = 0{,}79$. Гипотеза о низкой точности стрельбы не отклоняется. Посмотрим теперь на квантили распределения $\chi^2_{k-1-r} = \chi^2_3$:

[1]Пример взят из *Лагутин М.Б.* Наглядная математическая статистика: учебное пособие. — 2-е изд., испр. — М.: БИНОМ. Лаборатория знаний, 2009. — стр. 278.

$1-\alpha$	0,025	0,05	0,1	0,2	0,3	0,5	0,9	0,95	0,99
Квантиль χ^2_3	0,216	0,352	0,584	1,005	1,424	2,366	6,2514	7,815	11,345

Даже если мы выберем $\alpha = 0{,}7$ и будем отвергать верную основную гипотезу с вероятностью 0,7, критерий χ^2 всё равно примёт H_0. То есть, соответствие гипотезы с наблюдаемыми данными очень хорошее.

Обратим внимание на необходимость объединения маловероятных промежутков: если оставить $k = 8$, то $\overline{\chi}^2 \approx 32{,}6$, что значительно велико даже на уровне $\alpha = 10^{-5}$. Подобная ошибка критерия χ^2 вероятна на всех выборках с низкочастотными событиями. Проблема решается либо отбрасыванием, либо объединением данных событий.

Рекомендуемые условия применения критерия согласия χ^2 Пирсона — $n \geqslant 50, n_i \geqslant 5 \; \forall i = \overline{1,k}$.

Критерий Колмогорова

Наложим дополнительное условие на исходную задачу проверки гипотезы $H_0\colon F_\xi(x) = F_0(x) - F_0(x) \in C(\mathbb{R})$.

Рассмотрим статистику Колмогорова:

$$D_n\left(\mathbf{X}\right) = \sup_{x \in R} \left|F_n^*(x) - F_0(x)\right|$$

Формулировка критерия:

- Если верна гипотеза H_0, то $D_n\left(\mathbf{X}\right) \xrightarrow[n\to\infty]{\text{п.н.}} 0$;
- Если верна гипотеза H_1, т.е. $F_\xi \equiv G \neq F_0$, то

$$D_n\left(\mathbf{X}\right) \xrightarrow[n\to\infty]{\text{п.н.}} \sup_{x \in R} \left|G(x) - F_0(x)\right| > 0$$

Лемма. *Если гипотеза H_0 верна, и $F_0(x) \in C(\mathbb{R})$, то распределение статистики $D_n = \sup\limits_{x \in R} \left|F_n^*(x) - F_0(x)\right|$ не зависит от наблюдаемого распределения.*

При больших n применяется асимптотический подход.

Теорема Колмогорова. *Если гипотеза H_0 верна, и $F_0(x) \in C(\mathbb{R})$, то имеет место сходимость:*

$$P\left\{\sqrt{n} D_n\left(\mathbf{X}\right) \leqslant z\right\} \xrightarrow[n\to\infty]{} K(z) = 1 + 2\sum_{m=1}^{\infty} (-1)^m e^{-2m^2 z^2}$$

Находим константу $d_{1-\alpha}$ как решение уравнения $K\left(d_{1-\alpha}\right) = 1 - \alpha$.

Правило проверки гипотез:

- Если $\sqrt{n}D_n(\mathbf{X}) \in (d_{1-\alpha}, \infty)$, то гипотеза H_0 отвергается;
- Если $\sqrt{n}D_n(\mathbf{X}) \notin (d_{1-\alpha}, \infty)$, то гипотеза H_0 принимается.

Пример. Среди 100 студентов ВМК была проведена контрольная по ТВиМС.

Количество решённых задач	1	2	3	4	5
Частота	18	16	26	22	18

На уровне значимости $\alpha = 0{,}2$ с помощью критерия Колмогорова определить, подчиняются ли данные выборки на интервале $[0, 5]$ равномерному закону распределения случайной величины.

Запишем теоретическую функцию распределения:

$$F_0(x) = \begin{cases} 0, & x < 0 \\ x/5, & 0 \leqslant x \leqslant 5 \\ 1, & x > 5 \end{cases}$$

Составим следующую таблицу:

x_i	$F(x_i)$	n_i	$F_n^*(x_i)$	$\lvert F_0(x_i) - F_n^*(x_i)\rvert$
1	0,2	18	0,18	0,02
2	0,4	16	0,34	0,06
3	0,6	26	0,6	0
4	0,8	22	0,82	0,02
5	1	18	1	0

Отсюда $D_n(x) = \sup\limits_{x \in R} |F_n^*) - F_0(x)| = 0{,}06$, $\sqrt{n}D_n(x) = 0{,}6$, что меньше критического значения 0,65 функции Колмогорова при уровне значимости $\alpha = 0{,}2$, следовательно, гипотеза о равномерном распределении принимается.

2.12 Статистические выводы о параметрах нормального распределения. Распределения χ^2 и Стьюдента. Теорема Фишера

Определение. Говорят, что случайная величина ξ имеет *гамма-распределение* с параметрами $\lambda > 0,\ \alpha > 0$ (или $\xi \sim \Gamma_{\lambda,\alpha}$), если ξ имеет следующую плотность распределения:

$$p_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ c \cdot x^{\alpha-1}e^{-\lambda x}, & x > 0, \end{cases}$$

где постоянная c вычисляется из свойства нормировки плотности:

$$1 = \int\limits_{-\infty}^{\infty} p_\xi(x)dx = c\int\limits_{0}^{\infty} x^{\alpha-1}e^{-\lambda x}dx = \frac{c}{\lambda^\alpha}\int\limits_{0}^{\infty}(\lambda x)^{\alpha-1}e^{-\lambda x}d(\lambda x) = \frac{c}{\lambda^\alpha}\Gamma(\alpha),$$

откуда $c = \lambda^\alpha/\Gamma(\alpha)$.

Лемма. *Пусть* $\xi_1, \ldots, \xi_n$ *независимы, и* $\xi_i \sim \Gamma_{\lambda,\alpha_i}, i = \overline{1,n}$. *Тогда их сумма* $S_n = \xi_1 + \ldots + \xi_n \sim \Gamma_{\lambda,\alpha_1+\ldots+\alpha_n}$.

Доказательство. Найдём для начала характеристическую функцию гамма-распределения:

$$\int\limits_{-\infty}^{\infty} p_\xi(x)e^{itx}\,dx = \int\limits_{-\infty}^{\infty}\frac{\lambda^\alpha\,x^{\alpha-1}}{\Gamma(\alpha)}\,e^{-\lambda x}\,e^{itx}\,dx = \left|\begin{array}{c} y = -(\lambda - it)x \\ dy = -(\lambda - it)dx \end{array}\right| =$$

$$\int\limits_{-\infty}^{\infty}\frac{\lambda^\alpha\,y^{\alpha-1}}{(\lambda-it)^{\alpha-1}\,\Gamma(\alpha)}\,e^{-y}\frac{1}{\lambda-it}\,dy = \frac{\lambda^\alpha}{(\lambda-it)^\alpha}\int\limits_{-\infty}^{\infty}\frac{y^{\alpha-1}}{\Gamma(\alpha)}\,e^{-y}\,dy = \left(\frac{1}{1-\frac{it}{\lambda}}\right)^\alpha.$$

Теперь вспомним, что если ξ и η независимы, то $f_{\xi+\eta}(t) = f_\xi(t)f_\eta(t)$. Тогда, если у нас есть независимые случайные величины $\xi_1, \ldots, \xi_n$, $\xi_i \sim \Gamma(\lambda, \alpha_i)$, то характеристическая функция суммы выражаются следующим образом:

$$\left(\frac{1}{1-\frac{it}{\lambda}}\right)_1^\alpha \cdot \ldots \cdot \left(\frac{1}{1-\frac{it}{\lambda}}\right)_n^\alpha = \left(\frac{1}{1-\frac{it}{\lambda}}\right)^{\alpha_1+\ldots+\alpha_n}$$

По теореме Леви о непрерывности заключаем, что сумма имеет распределение $\Gamma(\lambda, \alpha_1 + \ldots + \alpha_n)$. ■

Лемма. *Если* $\xi \sim \mathrm{N}_{0,1}$, *то* $\xi^2 \sim \Gamma_{1/2,1/2}$.

Доказательство. Найдём функцию распределения случайной величины $\eta = \xi^2$:

$$\mathbb{P}(\xi^2 < x) = \mathbb{P}(|\xi| < \sqrt{x}) = 2\ \mathbb{P}(0 < x < \sqrt{x}) =$$

$$2\int\limits_{0}^{\sqrt{x}}\frac{1}{\sqrt{2\pi}}\,e^{u^2/2}\,du = \left|\begin{array}{cc} z = u^2 & u = \sqrt{z} \\ dz = 2udu & du = \frac{dz}{2\sqrt{z}} \\ u = 0 \implies z = 0 & u = \sqrt{x} \implies z = x \end{array}\right| =$$

$$= 2\int\limits_{0}^{x}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z}\frac{dz}{2\sqrt{z}} = \int\limits_{0}^{x}\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}z^{\frac{1}{2}-1}}{\sqrt{\pi}}e^{-z/2}dz.$$

Но это в точности функция гамма-распределения с параметрами $\lambda = \frac{1}{2}, \alpha = \frac{1}{2}$ (ведь $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$). ■

Следствие. *Если $\xi_1, \ldots, \xi_k$ независимы и $\xi_i \sim \mathrm{N}_{0,1}$, то случайная величина $\chi^2 = \xi_1^2 + \ldots + \xi_k^2 \sim \Gamma_{1/2,k/2}$*

Определение. Распределение суммы k квадратов независимых случайных величин со стандартным нормальным распределением называется *распределением хи-квадрат* с k степенями свободы (обозначение: χ_k^2).

Плотность распределения χ_k^2 имеет вид

$$f(y) = \begin{cases} \dfrac{1}{2^{x/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}, & \text{если } x > 0 \\ 0, & \text{если } x \leqslant 0 \end{cases}$$

Замечание. $\chi_2^2 = \Gamma_{1/2,1} = \mathrm{Exp}_{1/2}$.

Свойства распределения χ^2.

1. *Если случайные величины $\xi_1 \sim \chi_k^2$ и $\xi_2 \sim \chi_m^2$ независимы, то их сумма $\xi_1 + \xi_1 \sim \chi_{k+m}^2$;*

2. $\mathbb{E}\,\chi^2 = k, \quad \mathbb{D}\,\chi^2 = 2k.$

3. *Пусть дана последовательность случайных величин χ_n^2. Тогда при $n \to \infty$:*
$$\frac{\chi_n^2}{n} \xrightarrow[n\to\infty]{p} 1, \quad \frac{\chi_n^2 - n}{\sqrt{2n}} \xrightarrow[n\to\infty]{d} \mathrm{N}_{0,1}$$

4. *Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы и $\xi_i \sim \mathrm{N}_{a,\sigma^2}$. Тогда*
$$\sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma}\right)^2 \sim \chi_k^2.$$

Определение. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и $\xi_i \sim \mathrm{N}_{0,1}$. Распределение случайной величины

$$t_k = \frac{\xi_0}{\sqrt{\frac{\xi_1^2+\ldots+\xi_k^2}{k}}} = \frac{\xi_0}{\sqrt{x_k^2/k}}$$

называется *распределением Стьюдента* (*t-распределением* с k степенями свободы ($St(k)$ или T_k).

Плотность распределения T_k имеет вид

$$f_k(y) = \frac{\Gamma\big((k+1)/2\big)}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{y^2}{k}\right)^{-(k+1)/2}$$

Свойства распределения Стьюдента.

1. *Распределение Стьюдента симметрично, т.е. если $t_k \sim \mathrm{T}_k$, то $-t_k \sim \mathrm{T}_k$.*

2. $\mathrm{T}_k \xrightarrow[k\to\infty]{} \mathrm{N}_{0,1}$.

3. *У распределения Стьюдента T_k существуют только моменты порядка $m < k$, при этом все существующие моменты нечётного порядка равны нулю.*

Теорема Фишера. *Пусть случайные величины $X_1, \ldots, X_n$ независимы и $X_i \sim \mathrm{N}_{a,\sigma^2}$. Тогда:*

1. $\sqrt{n}\frac{\overline{X}-a}{\sigma} \sim \mathrm{N}_{0,1}$

2. $\frac{(n-1)S_0^2}{\sigma^2} = \sum\limits_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$

3. *Случайные величины $\overline{X}$ и S_0^2 независимы.*

Следствие. *Пусть случайные величины $X_1, \ldots, X_n$ независимы и $X_i \sim \mathrm{N}_{a,\sigma^2}$. Тогда:*

1. $\sqrt{n}\frac{\overline{X}-a}{\sigma} \sim \mathrm{N}_{0,1}$ *(для a при известном σ^2)*

2. $\sum\limits_{i=1}^{n} \left(\frac{X_i - a}{\sigma}\right)^2 \sim \chi^2_n$ *(для σ^2 при известном a)*

3. $\frac{(n-1)S_0^2}{\sigma^2} \sim \chi^2_{n-1}$ *(для σ^2 при неизвестном a)*

4. $\sqrt{n}\frac{\overline{X}-a}{S_0} \sim \mathrm{T}_{n-1}$ *(для a при неизвестном σ^2)*

Статистические выводы о параметрах нормального распределения

Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения N_{a,σ^2}. Построим точные доверительные интервалы (ДИ) с уровнем доверия α для параметров нормального распределения, используя следствие из теоремы Фишера.

1. ДИ для a при известном σ^2:

$$\mathbb{P}\left(\overline{X} - \frac{\tau\sigma}{\sqrt{n}} < a < \overline{X} + \frac{\tau\sigma}{\sqrt{n}}\right) = \alpha, \text{ где } \varphi_{0,1}(\tau) = \frac{1+\alpha}{2}$$

2. ДИ для σ^2 при известном a:

$$\frac{nS_1^2}{\sigma^2} \sim \chi^2_n, \text{ где } S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$$

Пусть g_1 и g_2 — квантили распределения χ^2_n уровней $\frac{1-\alpha}{2}$ и $\frac{1+\alpha}{2}$ соответственно. Тогда

$$\frac{(n-1)S_0^2}{\sigma^2} \sim \chi^2_{n-1}, \text{ где } S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

3. ДИ для σ^2 при неизвестном a:

$$\frac{(n-1)S_0^2}{\sigma^2} \sim \chi^2_{n-1}, \text{ где } S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Пусть g_1 и g_2 — квантили распределения χ^2_{n-1} уровней $\frac{1-\alpha}{2}$ и $\frac{1+\alpha}{2}$ соответственно. Тогда

$$\alpha = \mathbb{P}\left(g_1 < \frac{(n-1)S_0^2}{\sigma^2} < g_2\right) = \mathbb{P}\left(\frac{(n-1)S_0^2}{g_2} < \sigma^2 < \frac{(n-1)S_0^2}{g_1}\right)$$

4. ДИ для a при неизвестном σ^2:

$$\sqrt{n}\frac{\overline{X} - a}{S_0} \sim \mathrm{T}_{n-1}$$

Пусть c — квантиль распределения T_{n-1} уровня $\frac{1-\alpha}{2}$. Распределение Стьюдента симметрично, поэтому

$$\alpha = \mathbb{P}\left(-c < \sqrt{n}\frac{\overline{X} - a}{S_0} < c\right) = \mathbb{P}\left(\overline{X} - \frac{cS_0}{\sqrt{n}} < a < \overline{X} + \frac{cS_0}{\sqrt{n}}\right)$$

Дополнительные главы теории вероятностей

A.1 Усиленный закон больших чисел

Усиленный закон больших чисел в форме Колмогорова. *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots,$ — независимые одинаково распределённые случайные величины. Тогда*

1. *Если существует $\mathbb{E}\,\xi_1 = a$, то $\mathbb{P}(\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i = a) = 1$.*

   Иными словами, $\frac{1}{n}\sum\limits_{i=1}^{n} \xi_i \xrightarrow{\text{п.н.}} a$.

2. *Если существует $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i = a$, то существует $\mathbb{E}\,\xi_1 = a$.*

A.2 Обобщённое неравенство Чебышёва

Обобщённое неравенство Чебышёва. *Пусть функция g не убывает и неотрицательна на $\mathbb{R}$. Если $\mathbb{E}\,g(\xi) < +\infty$, то для любого $x \in \mathbb{R}$*

$$\mathbb{P}(\xi \geqslant x) \leqslant \frac{\mathbb{E}\,g(\xi)}{g(x)}.$$

Доказательство. Заметим, что $\mathbb{P}(\xi \geqslant x) \leqslant \mathbb{P}\big(g(\xi) \geqslant g(x)\big)$, поскольку функция g не убывает. Оценим последнюю вероятность по неравенству Маркова, которое можно применять в силу неотрицательности $g(x)$

$$\mathbb{P}\Big(g(\xi) \geqslant g(x)\Big) \leqslant \frac{\mathbb{E}\,g(\xi)}{g(x)}$$

■

Используя эту теорему, можно получить экспоненциально убывающую оценку (в то время как оценки по неравествам Чебышёва и Маркова убывают по степенному закону).

Таблицы основных распределений

Таблица B.1: Дискретные распределения.

Распределение	$\mathbb{P}(\xi = k)$	$\mathbb{E}\,\xi$	$\mathbb{D}\,\xi$	$\varphi(t)$
Бернулли B_p $p \in (0;1)$	$\mathbb{P}(\xi = 1) = p$ $\mathbb{P}(\xi = 0) = q$	p	pq	$q + pe^{it}$
Биномиальное $\mathrm{Bi}_{n,p}$ $p \in (0;1),\ n = 1, 2, \ldots$	$C_n^k\, p^k (1-p)^{n-k}$	np	npq	$(q + pe^{it})^n$
Пуассона Pois_λ $\lambda > 0$	$\dfrac{\lambda^k}{k!} \cdot e^{-\lambda}$	λ	λ	$e^{\lambda(e^{it}-1)}$
Геометрическое Geom_p $p \in (0;1),\ k = 1, 2, \ldots$	pq^{k-1}	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{p}{1 - qe^{it}}$
Отрицательное биномиальное $\mathrm{NB}_{r,p}$ $p \in (0;1),\ r = 1, 2, \ldots$	$C_{r-1+k}^k\, p^r q^{k-1}$	$\dfrac{r}{p}$	$\dfrac{rq}{p^2}$	$\left(\dfrac{p}{1 - qe^{it}}\right)^r$

Таблица B.2: Абсолютно непрерывные распределения.

Распределение	$f_\xi(x)$	$\mathbb{E}\,\xi$	$\mathbb{D}\,\xi$	$\varphi(t)$
Равномерное $\mathrm{U}_{[a,b]}$ $a < b$	$\dfrac{1}{b-a} \cdot \mathrm{I}(x \in [a;b])$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{itb} - e^{ita}}{it(b-a)}$
Показательное Exp_λ $\lambda > 0$	$\lambda e^{-\lambda x} \cdot \mathrm{I}(x > 0)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\dfrac{1}{1 - it/\lambda}$
Гамма $\Gamma_{\lambda,\alpha}$ $\alpha > 0, \lambda > 0$	$\dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot \mathrm{I}(x > 0)$	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$	$\left(\dfrac{1}{1 - it/\lambda}\right)^\alpha$
Нормальное N_{a,σ^2} $a \in \mathbb{R}, \sigma > 0$	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}$	a	σ^2	$e^{ait - \sigma^2 t^2/2}$
Коши $\mathrm{C}_{a,\sigma}$ $a \in \mathbb{R}, \sigma > 0$	$\dfrac{1}{\pi} \cdot \dfrac{\sigma}{\sigma^2 + (x-a)^2}$	-	-	$e^{ait - \sigma\|t\|}$

Таблица B.3: Фишеровская информация, содержащаяся в выборке размера 1.

Модель	$\mathrm{N}_{\theta,\sigma^2}$	$\mathrm{N}_{\mu,\theta^2}$	$\Gamma_{\theta,\alpha}$	$\mathrm{C}_{\theta,\sigma^2}$	$\mathrm{Bi}_{k,\theta}$	Pois_θ	$\mathrm{NB}_{r,\theta}$
$i_1(\theta)$	$\dfrac{1}{\sigma^2}$	$\dfrac{2}{\theta^2}$	$\dfrac{\alpha}{\theta^2}$	$\dfrac{1}{2}$	$\dfrac{k}{\theta(1-\theta)}$	$\dfrac{1}{\theta}$	$\dfrac{r}{\theta(1-\theta)^2}$

Таблица B.4: Эффективные оценки в регулярных моделях.

Модель	$\tau(\theta)$	$\tau^*(\mathbf{X})$	$\mathbb{D}\,\tau^*(\mathbf{X})$
$\mathrm{N}_{\theta,\sigma^2}$	θ	$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$	$\dfrac{\sigma^2}{n}$
$\mathrm{N}_{\mu,\theta^2}$	θ^2	$S^2 = \frac{1}{n}\sum\limits_{i=1}^{n} (X_i-\mu)^2$	$\dfrac{2\theta^4}{n}$
$\Gamma_{\theta,\alpha}$	$\dfrac{1}{\theta}$	$\dfrac{\overline{X}}{\alpha}$	$\dfrac{1}{\alpha n\theta^2}$
$\mathrm{Bi}_{k,\theta}$	θ	$\dfrac{\overline{X}}{k}$	$\dfrac{\theta(1-\theta)}{kn}$
Pois_θ	θ	$\overline{X}$	$\dfrac{\theta}{n}$
$\mathrm{NB}_{r,\theta}$	$\dfrac{\theta}{1-\theta}$	$\dfrac{\overline{X}}{r}$	$\dfrac{\theta}{rn(1-\theta)^2}$

Литература

Жуковский М. Е. Теория вероятностей. — Москва, 2017. — 61 с.

Ульянов В. В. Курс лекций по теории вероятности и математической статистике. — Москва. — 143 с.

Ушаков В. Г. Теория вероятностей и математическая статистика. — Москва, 2002. — 72 с.

Чернова Н. И. Теория вероятностей: Учеб. пособие. — Новосибирск : Новосиб. гос.ун-т., 2007. — 160 с.

Чернова Н. И. Математическая статистика: Учеб. пособие. — 2-е изд. — Новосибирск : Новосиб. гос.ун-т., РИЦ НГУ, 2014. — 150 с.