Московский государственный университет
им. М.В. Ломоносова
Факультет вычислительной математики и
кибернетики

Теория вероятностей
Математическая статистика

Москва
2020

Данное учебное пособие составлено на основе экзаменационных билетов 2020 года по курсу «Теория вероятностей и математическая статистика» факультета ВМК МГУ 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 (математическая статистика) — Рыгин А. С.
- Рецензент — 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 исправления в репозитории проекта: `https://github.com/MandaloreUltimate/stat-book`.

Оглавление

Глава 1

Математическая статистика

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

Определение. *Статистическая структура* — совокупность $(\Omega, \mathcal{A}, \mathcal{P})$, где Ω —множество элементарных исходов, $\mathcal{A}$ — σ-алгебра событий, $\mathcal{P}$ — семейство вероятностных мер, определённых на $\mathcal{A}$, параметризованное одно- или многомерным числовым параметром: $\mathcal{P} = (\mathcal{P}_\theta : \theta \in \Theta \subset R^m)$.

Определение. *Выборка* $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $\mathbf{X}(\omega_0) = (X_1(\omega_0), \ldots, X_n(\omega_0)) = (x_1, \ldots, x_n)$ — *реализация выборки* на элементарном исходе ω_0.

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

Определение. *Вариационный ряд* — выборка $X_1, \ldots, X_n$, упорядоченная по возрастанию на каждом элементарном исходе.

Вариационный ряд строится следующим образом:

$$X_{(1)}(\omega) = \min(X_1(\omega), \ldots, X_n(\omega)); X_{(n)}(\omega) = \max(X_1(\omega), \ldots, X_n(\omega));$$
$$X_{(k)}(\omega) = \{\forall \omega \in \Omega \Rightarrow \exists\ m \leqslant i_1, \ldots, i_{k-1}, i_k, i_{k+1}, \ldots, i_n \leqslant n, i_j \neq i_m\ (j \neq m) :$$
$$X_{(k)}(\omega) = X_{i_k}(\omega); X_{i_1}(\omega), \ldots, X_{i_{k-1}}(\omega) \leqslant X_{i_k}(\omega); X_{i_{k+1}}(\omega), \ldots, X_{i_n}(\omega) > X_{i_k}(\omega)\},$$
$$k = \overline{2, n-1}$$

Элемент $X_{(k)}$ — *k-я порядковая статистика.*

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

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

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

Эмпирическая функция распределенния строится по вариационному ряду следующим образом:

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

Пример. Найдём эмпирические функции распределения для крайних порядковых статистик.

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

Свойства эмпирической функции распределения.

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

2. *Для любого $y \in \mathbb{R}$:*

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

   2) *$\mathbb{D}F_n^*(y) = \dfrac{F(y)(1 - F(y))}{n} \leqslant \dfrac{1}{4n}$*

   3) *$\sqrt{n}(F_n^*(y) - F(y)) \Rightarrow \mathbf{N}(0, (1 - F(y))F(y))$, т.е. $F_n^*(y)$ — асимптотически нормальная оценка для $F(y)$.*

   4) *$nF_n^*(y) \sim \mathbf{B}(n, F(y))$*

   5) *$F_n^*(y) \overset{п.н.}{\to} F(y)$*

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:

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

Следовательно, применим ЗБЧ в форме Хинчина:

$$F_n^*(y) = \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\mathrm{I}(X_1 < y) = F(y)$$

2. Заметим:

$$\mathrm{I}(X_1 < y) \sim \mathbf{Bi}(F(y)) \Rightarrow \mathbb{E}\mathrm{I}(X_1 < y) = F(y)$$
$$\mathbb{D}\mathrm{I}(X_1 < y) = F(y)(1 - F(y))$$

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

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

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

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

Значения $F(y)$ принадлежат отрезку $[0, 1]$, а значит, произведение $F(y)(1 - F(y)) \leqslant \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{1}{4}$. (Читатель может взять производную по $F(y)$ и убедиться, что $\frac{1}{2}$ — точка максимума). А значит, $\mathbb{D}F_n^* \leqslant \dfrac{1}{4n}$.

Замечание. Пользуясь полученной оценкой на дисперсию и неравенством Чебышёва, можно показать, что эмпирическая функция распределения сходится к истинной по вероятности:

$$\mathbb{P}\left(|F_n^*(y) - F(y)| \geqslant \varepsilon\right) \ \leqslant \ \frac{\mathbb{D}F_n^*(y)}{\varepsilon^2} \ \leqslant \ \frac{1}{4n\varepsilon^2} \xrightarrow[n\to\infty]{} 0 \ \ \forall\, y \in \mathbb{R}.$$

Заметим так же, что ввиду 5-го свойства это замечание бесполезно.

3. Применим ЦПТ:

$$\sqrt{n}\left(F_n^*(y) - F(y)\right) = \sqrt{n}\left(\frac{\sum \mathrm{I}(X_i < y)}{n} - F(y)\right) =$$

$$= \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y) - nF(y)}{\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y) - n\mathbb{E}\mathrm{I}(X_1 < y)}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow \mathbf{N}(0, \mathbb{D}\mathrm{I}(X_1 < y)) = \mathbf{N}(0, (1 - F(y))F(y))$$

4. Следует из устойчивости по суммированию биномиального распределения. Поскольку $\mathrm{I}\left(X_i < y\right)$ независимы и имеют распределение Бернулли $\mathbf{Bi}(F(y))$, то их сумма

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

имеет биномиальное распределение $\mathbf{B}(n, F(y))$.

5. Выберем произвольный $y \in \mathbb{R}$. $\xi_i = \mathrm{I}(X_i < y)$ независимы, одинаково распределены и $\exists\ \mathbb{E}\xi_i = F(y)$. Тогда можно применить усиленный закон больших чисел в форме Колмогорова: $\mathbb{P}\left(\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i = F(y)\right) = 1$. Но это то же самое, что $\mathbb{P}\left(\lim\limits_{n\to\infty} F_n^*(y) = F(y)\right) = 1$, а это в точности определение сходимости почти наверное.

■

1.2 Выборочные моменты. Их свойства

В параграфе 2.1 мы предположили, что все случайные величины выборки $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $g(\xi^*)$:

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

Определение. *Выборочная дисперсия:*

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

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:*

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

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

Рассмотрим статистику $T(\mathbf{X})$, оценивающую значение некоторой функции параметра $\tau(\theta)$. Введем несколько определений.

Определение. Статистика (или оценка) называется *несмещённой*, если $\mathbb{E}_\theta T(\mathbf{X}) = \tau(\theta) \quad \forall\, \theta \in \Theta$.

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

Определение. Статистика называется *состоятельной*, если $T_n(\mathbf{X}) \xrightarrow{\mathbb{P}_\theta} \theta \quad \forall\, \tau(\theta) \in \Theta$ при $n \to \infty$, где n — объём выборки $\mathbf{X} = (X_1, \ldots, X_n)$. Иными словами, статистика называется состоятельной, если её значение сходится к истинному значению параметра с ростом объёма выборки.

Определение. Статистика называется *асимптотически нормальной*, если существуют такие $a_n(\theta), b_n(\theta)$, что $\dfrac{T_n(\mathbf{X}) - a_n(\theta)}{b_n(\theta)} \Rightarrow \mathbf{N}(0,1)$. Иными словами, оценка называется асимптотически нормальной, если с ростом объёма выборки распределение статистики (статистика, являясь измеримой функцией от выборки, сама является случайной величиной) слабо стремится (или, что то же самое, стремится по распределению) к нормальному.

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

1. *Если $\mathbb{E}|X_1| < \infty$, то $\mathbb{E}\overline{X} = \mathbb{E}X_1 = a$*
2. *Если $\mathbb{E}|X_1| < \infty$, то $\overline{X} \xrightarrow{\mathbb{P}} \mathbb{E}X_1 = a$ при $n \to \infty$.*
3. *Если $\mathbb{D}X_1 < \infty,\ \mathbb{D}X_1 \neq 0$, то $\sqrt{n}(\overline{X} - \mathbb{E}X_1) \Rightarrow \mathbf{N}(0, \mathbb{D}X_1)$.*

Доказательство.

1. $\mathbb{E}\overline{X} = \frac{1}{n}(\mathbb{E}X_1 + \ldots + \mathbb{E}X_n) = \frac{1}{n} \cdot n\mathbb{E}X_1 = \mathbb{E}X_1 = a$
2. Из ЗБЧ в форме Хинчина:
$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1 = a$$
3. Из ЦПТ:
$$\sqrt{n}\left(\overline{X} - \mathbb{E}X_1\right) = \frac{\sum\limits_{i=1}^{n} X_i - n\mathbb{E}X_1}{\sqrt{n}} \Rightarrow \mathbf{N}(0, \mathbb{D}X_1)$$

■

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* $\mathbb{D}X_1 < \infty$.

1. *Выборочные дисперсии* S^2 *и* S_0^2 *являются состоятельными оценками для истинной дисперсии:*

$$S^2 \xrightarrow{\mathbb{P}} \mathbb{D}X_1 = \sigma^2, \quad S_0^2 \xrightarrow{\mathbb{P}} \mathbb{D}X_1 = \sigma^2$$

2. *Величина* S^2 *— смещённая оценка дисперсии, а* S_0^2 *—– несмещённая:*

$$\mathbb{E}S^2 = \frac{n-1}{n}\mathbb{D}X_1 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad \mathbb{E}S_0^2 = \mathbb{D}X_1 = \sigma^2$$

3. *Если* $0 \neq \mathbb{D}(X_1 - \mathbb{E}X_1)^2 < \infty$, *то* S^2 *и* S_0^2 *являются асимптотически нормальными оценками истинной дисперсии:*

$$\sqrt{n}\left(S^2 - \mathbb{D}X_1\right) \Rightarrow \mathbf{N}(0, \mathbb{D}(X_1 - \mathbb{E}X_1)^2)$$

Доказательство.

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

   Используя состоятельность первого и второго выборочных моментов и свойства сходимости по вероятности, получаем:

$$S^2 = \overline{X^2} - (\overline{X})^2 \xrightarrow{\mathbb{P}} \mathbb{E}X_1^2 - \left(\mathbb{E}X_1\right)^2 = \sigma^2$$

$$\frac{n}{n-1} \underset{n\to\infty}{\longrightarrow} 1 \Rightarrow S_0^2 = \frac{n}{n-1}S^2 \xrightarrow{\mathbb{P}} \sigma^2$$

2. Используя несмещённость первого и второго выборочных моментов:

$$\mathbb{E}S^2 = \mathbb{E}\left(\overline{X^2} - (\overline{X})^2\right) = \mathbb{E}\overline{X^2} - \mathbb{E}(\overline{X})^2 = \mathbb{E}X_1^2 - \mathbb{E}(\overline{X})^2 =$$
$$= \mathbb{E}X_1^2 - \left((\mathbb{E}\overline{X})^2 + \mathbb{D}\overline{X}\right) = \mathbb{E}X_1^2 - \left(\mathbb{E}X_1\right)^2 - \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) =$$
$$= \mathbb{D}X_1 - \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \sigma^2 - \frac{1}{n^2}n\mathbb{D}X_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

   Откуда следует:

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

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

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

$$\sqrt{n}(S^2 - \sigma^2) = \sqrt{n}(\overline{Y^2} - (\overline{Y})^2 - \sigma^2) = \sqrt{n}t(\overline{Y^2} - \mathbb{E}Y_1^2) - \sqrt{n}(\overline{Y})^2 =$$

$$= \frac{\sum\limits_{i=1}^{n} Y_i^2 - n\mathbb{E}Y_1^2}{\sqrt{n}} - \overline{Y}\cdot\sqrt{n}\overline{Y} \Rightarrow \mathbf{N}(0, \mathbb{D}(X_1 - a)^2),$$

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

■

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

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

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

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

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

Оценки также могут вводиться и для функций $\tau(\theta)$ параметра θ; они обладают всеми аналогичными свойствами.

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. В отсутствие этого свойства оценка 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 $\mathbb{E}X$ могут выступать $\mathbb{E}X_1$ или $\mathbb{E}\overline{\mathbf{X}}$.

2. Несмещённые оценки могут не существовать.

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

   $$\mathbb{E}T(\mathbf{X}) = \sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta} \Rightarrow \sum_{x=0}^{\infty} T(x) \frac{\theta^{x+1}}{x!} = \sum_{r=0}^{\infty} \frac{\theta^r}{r!} \Rightarrow T(x) \equiv \frac{1}{\theta}$$

   Т.к. полученная статистика зависит от θ, искомой несмещённой оценки для $\tau(\theta)$ не существует.

3. Несмещённые оценки могут существовать, но быть бессмысленными.

   Пример. Дано отрицательное биноминальное распределение $\mathbf{NB}(1, 1-\theta) \equiv \mathbf{Geom}(1-\theta)$, над которым произведено одно наблюдение. Найти несмещённую оценку для параметра θ.

   $$\sum_{x=0}^{\infty} T(x)\theta^x = \frac{\theta}{1-\theta} = \sum_{r=1}^{\infty} \theta^r$$
   $$T(x) = \begin{cases} 0, & \text{если } x = 0 \\ 1, & \text{если } x \geqslant 1 \end{cases}$$

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

4. Состоятельные оценки не единственны.

   К примеру, выборочная дисперсия S^2 и несмещённая выборочная дисперсия S_0^2 являются состоятельными оценками теоретической дисперсии.

5. Состоятельные оценки могут быть смещёнными.

   Как было показано ранее, выборочная дисперсия является состоятельной, но смещённой оценкой теоретической дисперсии.

Рассмотрим несмещённые оценки $T(\mathbf{X})$ параметра θ, для которых существует дисперсия: $\mathbb{E}(T(\mathbf{X}) - \theta)^2 = \mathbb{D}T(\mathbf{X})$. Тогда сравнивать две разные несмещённые оценки $T_1(\mathbf{X})$ и $T_2(\mathbf{X})$, для которых $\mathbb{E}T_1 = \mathbb{E}T_2 = \theta$, можно по их дисперсиям, однако в таком случае неравенство $\mathbb{D}T_1 < \mathbb{D}T_2$ должно выполняться при любом θ. Введём понятие оптимальной оценки.

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

1. $T(\mathbf{X})$ — несмещённая.
2. $T(\mathbf{X})$ имеет равномерно минимальную дисперсию, т.е. для любой другой **несмещённой** оценки $T_1(\mathbf{X})$ параметра θ: $\mathbb{D}T(\mathbf{X}) \leqslant \mathbb{D}T_1(\mathbf{X})\ \forall X$.

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: $\mathbb{E}T_1(\mathbf{X}) = \mathbb{E}T_2(\mathbf{X}) = T(\mathbf{X})$, а а в силу того, что они имеют равномерно минимальную дисперсию: $\mathbb{D}T_1(\mathbf{X}) = \mathbb{D}T_2(\mathbf{X})\ \forall\theta$.

Введём новую статистику:

$$T_3(\mathbf{X}) = \frac{T_1(\mathbf{X}) + T_2(\mathbf{X})}{2} \tag{1.1}$$

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

Имеем также:

$$\mathbb{D}T_3(\mathbf{X}) = \frac{\mathbb{D}\left(T_1(\mathbf{X}) + T_2(\mathbf{X})\right)}{4} = \frac{\mathbb{D}T_1(\mathbf{X}) + \mathbb{D}T_2(\mathbf{X}) + 2\operatorname{cov}\left(T_1(\mathbf{X})T_2(\mathbf{X})\right)}{4}$$

В силу свойства

$$\mathbb{E}\xi^2 < \infty, \mathbb{E}\eta^2 < \infty \Rightarrow |\operatorname{cov}(\xi,\eta)| = |\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)| \leqslant \sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta},$$

где равенство достигается тогда и только тогда, когда $\xi = a\eta + b$, получаем:

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

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, справедливо равенство, достигаемое при следующих условиях:

$$T_1(\mathbf{X}) = aT_2(\mathbf{X}) + b \Rightarrow \mathbb{E}T_1(\mathbf{X}) = a\mathbb{E}T_2(\mathbf{X}) + b \Leftrightarrow$$
$$\Leftrightarrow \theta = a\theta + b\ \forall\theta \Rightarrow a = 1, b = 0$$

■

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

В зависимости от типа 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:

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

Таким образом, смысл функции правдоподобия — «вероятность»[2] попасть в заданную точку при соответствующем параметре θ в дискретном случае; для абсолютно непрерывного аналогично — вероятность попасть в куб с центром в $x_1, \ldots, x_n$ и сторонами $dx_1, \ldots, dx_n$.

Определение. *Достаточная статистика* — статистика $T(\mathbf{X})$ такая, что $\forall t \in \mathbb{R}^m$, [3] $\forall B \in \mathfrak{B}(\mathbb{R}^n)$ условное распределение $\mathbb{P}(X_1, \ldots, X_n \in B | T = t)$ не

[2]Строго говоря, функция правдоподобия не является вероятностной мерой над θ, хотя бы потому, что $\int_\Theta L(x, \theta) d\theta \neq 1$. Термин «вероятность» применён здесь в переносном смысле.

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

зависит от параметра θ.

Иными словами, если значение статистики T известно и фиксировано, то даже знание её распределения больше не даёт никакой информации о параметре; достаточно лишь вычислить T по выборке.

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

$$L(\mathbf{X}_1, \dots, X_n, \theta) \overset{п.н.}{=} h(\mathbf{X}) \cdot \Psi(T, \theta)$$

Доказательство. Рассмотрим только дискретный случай.

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

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

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

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

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 $\mathbb{E}\varphi(T) = 0 \ \forall \theta \in \Theta$, 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 распределения $X_{(n)}$ для $U[0, \theta]$ - это $f(t) = \dfrac{nt^{n-1}}{\theta^n} \mathrm{I}(0 \leqslant t \leqslant \theta)$.

$$\mathbb{E}\varphi(X_{(n)}) = \int\limits_0^\theta \varphi(t)\frac{nt^{n-1}}{\theta^n}dt = 0 \quad \Rightarrow \quad \int\limits_0^\theta \varphi(t)t^{n-1}dt = 0 \quad \Rightarrow$$

$$\{\text{продифференцируем по } \theta\} \quad \Rightarrow \quad \varphi(\theta)\theta^{n-1} = 0 \quad \Rightarrow \quad \varphi(T) \equiv 0 \ \ \forall\, T > 0$$

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

1.5 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 $\varphi(\theta) = \int\limits_{\mathbf{R}^n} T(x)L(x,\theta)\mu(dx) < \infty$, 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

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

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

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,

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

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

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

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

Посчитаем дисперсию функции вклада:

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

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

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

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

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

Определение. Пусть $\mathbf{X} = (X_1,\ldots,X_n)$ — 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 $X_1,\ldots,X_n$ — 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:*

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

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

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

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

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

Заметим, что

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

Откуда следует:

$$\int U(x,\theta)L(x,\theta)\mu(dx) = 0 \Leftrightarrow \mathbb{E}U(X,\theta) = 0$$
$$\int T(x)U(x,\theta)L(x,\theta)\mu(dx) = \tau'(\theta) \Leftrightarrow \mathbb{E}T(\mathbf{X})U(X,\theta) = \tau'(\theta)$$

Вычитая из первого равенства, умноженного на $\tau(\theta)$, второе, получаем:

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

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

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

Из неравенства Коши-Буняковского:

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

что равносильно п.1 теоремы:

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

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

$$T(\mathbf{X}) = \varphi(\theta)U(X,\theta) + \psi(\theta) \Rightarrow \tau(\theta) = \psi(\theta) \Rightarrow a_n(\theta) = \varphi(\theta)$$

■

Рассмотрим некоторый класс оценок $K = \left\{\hat{\theta}\left(\mathbf{X}\right)\right\}$ параметра θ.

Определение. Говорят, что оценка $\theta^*\left(\mathbf{X}\right) \in K$ является эффективной оценкой параметра θ в классе K, если для любой другой оценки $\hat{\theta} \in K$ имеет место неравенство:

$$E\left(\theta^* - \theta\right)^2 \leqslant \mathbb{E}(\hat{\theta} - \theta)^2\ \forall\theta \in \Theta$$

Обозначим класс несмещённых оценок:

$$K_0 = \left\{\hat{\theta}\left(\mathbf{X}\right) : E\hat{\theta} = \theta, \forall\theta \in \Theta\right\}$$

Оценка, эффективная в K_0 называется просто *эффективной*.

Для оценки $\theta^* \in K_0$ по определению дисперсии

$$\mathbb{E}\left(\theta^* - \theta\right)^2 = \mathbb{E}\left(\theta^* - \mathbb{E}\theta^*\right)^2 = \mathbb{D}\theta^*$$

Замечание. В качестве критерия эффективности можно использовать неравенство Рао—Крамера — оно обращается в равенство тогда и только тогда, когда оценка является эффективной.

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 того, что неравенство Рао—Крамера должно обращаться в равенство).

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

Теорема Рао—Блекуэлла—Колмогорова. *Если оптимальная оценка параметра θ существует, то она является функцией от достаточной статистики.*

Доказательство. В доказательстве используются следующие свойства условного матожидания:

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

1. Пусть $T(\mathbf{X})$ — достаточная статистика, $T_1(\mathbf{X})$ — несмещённая оценка параметра θ, т.е. $\mathbb{E}T_1(\mathbf{X}) = \theta$. Рассмотрим функцию $H(T) = \mathbb{E}\left(T_1|T\right)$. Тогда из первого свойства следует:

$$\mathbb{E}H(T) = \mathbb{E}\left(\mathbb{E}\left(T_1|T\right)\right) = \mathbb{E}T_1 = \theta \Rightarrow H(T) \text{ — несмещённая оценка } \theta$$

2. Докажем равномерную минимальность её дисперсии:

$$\begin{aligned}\mathbb{E}((T_1 - H(T))(H(T) - \theta) = \mathbb{E}(\mathbb{E}((T_1 - H(T))(H(T) - \theta)|T)) =\\ = \mathbb{E}((H(T) - H(T))(H(T) - \theta)) = 0\end{aligned}$$

Тогда из свойств условного матожидания

$$\mathbb{D}\left(T_1\right) = \mathbb{E}\left(T_1 - \theta\right)^2 = \mathbb{E}\left(T_1 - H(T) + H(T) - \theta\right)^2 =$$

$$= \mathbb{E}\left(T_1 - H(T)\right)^2 + \mathbb{D}(H(T)) \geqslant \mathbb{D}(H(T))$$

Таким образом, $H(T)$ — оптимальная оценка θ.

$\blacksquare$

Теорема Колмогорова. *Если $T(\mathbf{X})$ — полная достаточная статистика, то она является оптимальной оценкой своего математического ожидания.*

Доказательство. Докажем, что $T(\mathbf{X})$ является единственной несмещённой оценкой для $\mathbb{E}T(\mathbf{X})$. Тогда $T(\mathbf{X})$ будет оптимальной оценкой. Предположим, что $T_1(\mathbf{X})$ — оптимальная оценка для $\mathbb{E}T(\mathbf{X})$. Из теоремы Рао—Блекуэлла—Колмогорова получаем, что $T_1 = H(T)$ и $\mathbb{E}T_1 = \mathbb{E}T$. Тогда:

$$\mathbb{E}\underbrace{\left(T(\mathbf{X}) - H(T(\mathbf{X}))\right)}_{\varphi(T)} = 0$$

Из условия полноты $T(\mathbf{X})$ следует, что $\varphi(T) = 0$ с вероятностью 1, т.е. $T = H(T)$ с вероятностью 1. $\blacksquare$

1.7 Метод моментов. Свойства оценок, полученных методом моментов

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

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

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:

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

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

$$\mathrm{E}X_1 = \frac{\theta}{2}, \quad \theta = 2\mathrm{E}X_1, \quad \theta_1^* = 2\bar{X}$$

Найдём оценку метода моментов k по k-му моменту:

$$\mathrm{E}X_1^k = \int_0^\theta y^k \frac{1}{\theta} dy = \frac{\theta^k}{k+1}, \quad \theta = \sqrt[k]{(k+1)\mathrm{E}X_1^k}, \quad \theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}$$

Утверждение. *Пусть $\theta^* = h^{-1}(\overline{g(\mathbf{X})})$ — оценка параметра θ, полученная методом моментов, причём функция h^{-1} непрерывна. Тогда оценка θ^* состоятельна.*

Доказательство. По ЗБЧ Хинчина имеем:

$$\overline{g(\mathbf{X})} = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right) \xrightarrow{\mathbb{P}} \mathbb{E}g\left(X_1\right) = h(\theta)$$

Ввиду непрерывности функции h^{-1}:

$$\theta^* = h^{-1}(\overline{g(\mathbf{X})}) \xrightarrow{\mathbb{P}} h^{-1}\left(\mathbb{E}g\left(X_1\right)\right) = h^{-1}(h(\theta)) = \theta$$

■

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:

$$\frac{\sqrt{n}(\theta^* - \theta)}{\sigma(\theta)} \Rightarrow \mathbf{N}(0, 1).$$

Лемма. *Пусть функция $g(y)$ такова, что $0 \neq \mathbb{D}g\left(X_1\right) < \infty$. Тогда статистика $\overline{g(\mathbf{X})}$ является асимптотически нормальной оценкой для $\mathbb{E}g\left(X_1\right)$ с коэффициентом $\sigma^2(\theta) = \mathbb{D}g\left(X_1\right)$:*

$$\sqrt{n}\frac{\overline{g(\mathbf{X})} - \mathbb{E}g\left(X_1\right)}{\sqrt{\mathbb{D}g\left(X_1\right)}} \Rightarrow \mathbf{N}(0; 1)$$

Доказательство. Следует непосредственно из ЦПТ. ■

Замечание. Следующая теорема утверждает асимптотическую нормальность оценок вида

$$\theta^* = H(\overline{g(\mathbf{X})}) = H\left(\frac{g\left(X_1\right) + \ldots + g\left(X_n\right)}{n}\right)$$

которые обычно получаются при использовании метода моментов, при этом всегда $\theta = H\left(\mathbb{E}g\left(X_1\right)\right)$.

Утверждение. *Пусть функция* $g(y)$ *такова, что* $0 \neq \mathbb{D}g\left(X_1\right) < \infty$, *функция* $H(y)$ *дифференцируема в точке* $a = \mathbb{E}g\left(X_1\right)$ *и её производная в этой точке* $H'(a) = H'(y)|_{y=a}$ *отлична от нуля. Тогда оценка* $\theta^* = H(\overline{g(\mathbf{X})})$ *является асимптотически нормальной оценкой для параметра* $\theta = H\left(\mathbb{E}g\left(X_1\right)\right) = H(a)$ *с коэффициентом асимптотической нормальности* $\sigma^2(\theta) = \left(H'(a)\right)^2 \cdot Dg\left(X_1\right)$.

Доказательство. Согласно ЗБЧ последовательность $\overline{g(\mathbf{X})}$ стремится к $a = \mathbb{E}g\left(X_1\right)$ 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 веро- ятности сохраняется под действием непрерывной функции, получим, что $G(\overline{g(\mathbf{X})}) \xrightarrow{\mathbb{P}} G(a) = H'(a)$.

Заметим также, что по вышеприведённой лемме величина $\sqrt{n}(\overline{g(\mathbf{X})} - a)$ слабо сходится к нормальному распределению $\mathbf{N}(0, \mathbb{D}g(X_1))$: Пусть ξ — случайная величина из этого распределения. Тогда

$$\sqrt{n}(H(\overline{g(\mathbf{X})}) - H(a)) = \sqrt{n}(\overline{g(\mathbf{X})} - a) \cdot G(\overline{g(\mathbf{X})}) \Rightarrow \xi \cdot H'(a)$$

Мы использовали следующее свойство слабой сходимости: если $\xi_n \Rightarrow \xi$ и $\eta_n \xrightarrow{\mathbb{P}} c = \text{const}$, то $\xi_n \eta_n \Rightarrow c\xi$. Но распределение случайной величины $\xi \cdot H'(a)$ есть $\mathbf{N}(0, (H'(a))^2 \cdot \mathbb{D}g(X_1))$, откуда следует

$$\sigma^2(\theta) = \left(H'(a)\right)^2 \cdot \mathbb{D}g\left(X_1\right)$$

∎

1.8 Метод максимального правдоподобия. Свойства оценок максимального правдоподобия

Определение. *Оценка максимального правдоподобия* $\hat{\theta}$ *параметра* θ — точка параметрического множества Θ, в которой функция правдоподобия $L(\mathbf{X}, \theta)$ при заданном X достигает максимума, т.е.:

$$L(\boldsymbol{x}, \hat{\theta}) = \sup_{\theta \in \Theta} L(\boldsymbol{x}, \theta)$$

Замечание. Поскольку функция $\ln y$ монотонна, то точки максимума функций $L(\mathbf{X},\theta)$ и $lnL(\mathbf{X},\theta)$ 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 $\hat{\theta}$ удовлетворяет уравнению:

$$\frac{\partial \ln L(\mathbf{x},\theta)}{\partial \theta} = 0$$

Если θ — векторный параметр: $\theta = (\theta_1, \dots, \theta_n)$, то это уравнение заменяется системой уравнений:

$$\frac{\partial \ln L(\mathbf{x},\theta)}{\partial \theta_i} = 0,\ \ i = \overline{1,n}$$

Утверждение. *Если существует эффективная оценка $T(\mathbf{X})$ скалярного параметра θ, то она совпадает с оценкой максимального правдоподобия.*

Доказательство. Если оценка $T(\mathbf{X})$ скалярного параметра θ эффективна, то в неравенстве Рао—Крамера достигается равенство:

$$U(X,\theta) = \frac{\partial \ln L(\mathbf{x},\theta)}{\partial \theta} = \frac{T(\mathbf{X}) - \theta}{a_n(\theta)}$$

■

Утверждение. *Если $T(\mathbf{X})$ достаточная статистика, а оценка максимального правдоподобия $\hat{\theta}$ 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 $g(T(\mathbf{X}),\theta)$ по θ, Следовательно $\hat{\theta}$ есть функция от $T(\mathbf{X})$. ■

Определение. *Асимптотически эффективная оценка* параметра $\tau(\theta)$ — оценка τ^*:

$$D\tau^* \cdot \frac{i_n(\theta)}{(\tau'(\theta))^2} \xrightarrow[n\to\infty]{} 1,\ \text{где}\ i_n(\theta) = \mathbb{E}\left(\frac{\partial \ln L(X,\theta)}{\partial \theta}\right)^2 = \mathbb{E}(U^2(X,\theta))$$

Утверждение. *Пусть выполнены следующие условия:*

1. *Функция правдоподобия $L(\mathbf{X}, \theta)$ удовлетворяет условиям регулярности для первых двух производных;*

2. *$\exists!\ \theta^*$ — оценка максимального правдоподобия для всех θ, которая достигается во внутренней точке Θ.*

Тогда оценка θ^*:

1. *асимптотически несмещена*

2. *состоятельна*

3. *асимптотически эффективна*

4. *асимптотически нормальна*

1.9 Интервальное оценивание. Методы центральной статистики и использования точечной оценки

Определение. *Доверительный интервал* для параметра θ с коэффициентом доверия $0 \leqslant \alpha \leqslant 1$ — интервал $(T_1(\mathbf{X}), T_2(\mathbf{X}))$, т.ч. $\mathbb{P}_\theta(T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})) \geqslant \alpha$.

Пример. Пусть $X_1, \ldots, X_n$ — выборка из $\mathbf{N}(\theta, 1)$. Тогда

$$\theta^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathbf{N}\left(\theta, \frac{1}{n}\right) \Rightarrow (\overline{X} - \theta)\sqrt{n} \sim \mathbf{N}(0; 1)$$

Для величины, имеющей стандартное нормальное распределение, строим доверительный интервал, т.е. находим такое $t_{\alpha/2}$, что

$$\mathbb{P}_\theta\left(|(\overline{X} - \theta)\sqrt{n}| < t_{\alpha/2}\right) = \alpha$$

Решаем уравнение относительно θ и получаем

$$\mathbb{P}_\theta\left(\overline{X} - \frac{t_{\alpha/2}}{\sqrt{n}} < \theta < \overline{X} + \frac{t_{\alpha/2}}{\sqrt{n}}\right) = \alpha$$

Определение. *Центральная статистика* — функция $G(X, \theta)$, т.ч.:

1. $G(X, \theta)$ непрерывна и строго монотонна по θ при любом фиксированном X.

2. $\mathbb{P}_\theta(G(X,\theta) < t) = F(t)$ непрерывна и не зависит от θ.

Замечание. Формально определённая выше величина не является статистикой, т.к. зависит от неизвестного параметра θ.

Построение доверительного интервала с помощью центральной статистики:

1. Зафиксируем $\alpha_1, \alpha_2 \in \mathbf{R}$, т.ч.

$$\mathbb{P}_\theta(\alpha_1 \leqslant G(X,\theta) \leqslant \alpha_2) = \alpha \ \forall\theta \Leftrightarrow F_G(\alpha_2) - F_G(\alpha_1) = \alpha$$

2. Пусть $G(X,\theta)$ возрастает. Из условий

$$\begin{cases} G(X,\theta) \leqslant \alpha_2 \\ G(X,\theta) \geqslant \alpha_1 \end{cases}$$

находятся статистики

$$\begin{cases} T_2(\mathbf{X}) : G(X, T_2(\mathbf{X})) = \alpha_2 \\ T_1(\mathbf{X}) : G(X, T_1(\mathbf{X})) = \alpha_1 \end{cases} \Leftrightarrow T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})$$

откуда $\mathbb{P}_\theta\left(T_1(\mathbf{X}) \leqslant \theta \leqslant T_2(\mathbf{X})\right) \geqslant \alpha \ \forall\theta$.

Определение. *Центральный доверительный предел* для параметра θ с коэффициентом доверия $0 \leqslant \alpha \leqslant 1$ — интервал $(T_1(\mathbf{X}), T_2(\mathbf{X}))$, т.ч.

$$\mathbb{P}_\theta\left(T_1(\mathbf{X}) > \theta\right) = \frac{1-\alpha}{2}$$
$$\mathbb{P}_\theta\left(T_2(\mathbf{X}) < \theta\right) = \frac{1-\alpha}{2}$$

Построение доверительного интервала с помощью точечной оценки:

1. Пусть $T(\mathbf{X})$ — точечная оценка θ. Обозначим $H(t,\theta) = \mathbb{P}_\theta(T(\mathbf{X}) < t)$. Предположим, что $H(t,\theta)$ — непрерывная и строго монотонная функция θ при любом фиксированном t. В этом случае

$$\begin{cases} \mathbb{P}_\theta\left(T(\mathbf{X}) > a_1(\theta)\right) = \dfrac{1-\alpha}{2} \\ \mathbb{P}_\theta\left(T(\mathbf{X}) < \alpha_2(\theta)\right) = \dfrac{1-\alpha}{2} \end{cases} \Leftrightarrow \begin{cases} 1 - H(\alpha_1(\theta),\theta) = \dfrac{1-\alpha}{2} \\ H(\alpha_2(\theta),\theta) = \dfrac{1-\alpha}{2} \end{cases}$$

2. Рассмотрим вспомогательную лемму.

   Лемма. *Если $H(t,\theta)$ возрастает по θ, то $\alpha_1(\theta)$ и $\alpha_2(\theta)$ убывают. Если же $H(t,\theta)$ убывает по θ, то $\alpha_1(\theta)$ и $\alpha_2(\theta)$ возрастают.*

Доказательство. Пусть $H(t,\theta)$ возрастает. Предположим, что $\theta_1 < \theta_2 \Rightarrow \alpha_2(\theta_1) \leqslant \alpha_2(\theta_2)$ и рассмотрим $a_2(\theta)$, учитывая, что $H(t,\theta)$, как и всякая функция распределения, неубывает по первому аргументу:

$$\frac{1-a}{2} = H\left(\alpha_2\left(\theta_1\right)\theta_1\right) < H\left(\alpha_2\left(\theta_1\right)\theta_2\right) \leqslant H\left(\alpha_2\left(\theta_2\right)\theta_2\right) = \frac{1-\alpha}{2}$$

Полученное противоречие завершает доказательство. ∎

3. Из леммы следует, что для любого θ

$$\begin{aligned}
\alpha_1(\theta) < T(\mathbf{X}) \Leftrightarrow \theta > \varphi_1(T(\mathbf{X})) \Rightarrow \mathbb{P}_\theta(\theta > \varphi_1(T(\mathbf{X}))) &= \frac{1-\alpha}{2} \\
\alpha_2(\theta) > T(\mathbf{X}) \Leftrightarrow \theta < \varphi_2(T(\mathbf{X})) \Rightarrow \mathbb{P}_\theta(\theta < \varphi_2(T(\mathbf{X}))) &= \frac{1-\alpha}{2} \\
\Rightarrow P_\theta(\underbrace{\varphi_2(T(\mathbf{X}))}_{T_1(\mathbf{X})} \leqslant \theta \leqslant \underbrace{\varphi_1(T(\mathbf{X}))}_{T_2(\mathbf{X})}) &= \alpha
\end{aligned}$$

1.10 Проверка гипотез. Лемма Неймана—Пирсона

Определение. *Гипотеза H* — любое предположение о распределении наблюдаемой случайной величины: $H = \{\mathcal{P} = \mathcal{P}_1\}$ или $H = \{\mathcal{P} \in \mathbb{F}\}$, где $\mathbb{F}$ — некоторое подмножество в множестве всех распределений. Гипотеза называется *простой* в первом случае, *сложной* во втором. Если гипотез всего две, то одну из них принято называть *основной*, а другую — *альтернативой*.

Замечание. Типичные задачи проверки гипотез:

1. Гипотезы о виде распределения: как правило, это проверка принадлежности некоторому параметрически заданному семейству распределений. В этом случае гипотеза называется *простой*, если она состоит из *одного* значения параметра, и *сложной* иначе;

2. Гипотезы о проверке однородности выборки: дано несколько выборок; основная гипотеза состоит в том, что эти выборки извлечены из одного распределения;

3. Гипотеза независимости: по выборке $(X_1, Y_1), \ldots, (X_n, Y_n)$ из n независимых наблюдений пары случайных величин проверяется гипотеза $H_1 = \{X_i \text{ и } Y_i \text{ независимы}\}$ при альтернативе $H_1 = \{\text{ предположение неверно }\}$. Обе гипотезы являются сложными;

4. Гипотеза случайности: в эксперименте наблюдаются n случайных величин $X_1, \ldots, X_n$ и проверяется сложная гипотеза $H_1 = \{X_1, \ldots, X_n$ независимы и одинаково распределены$\}$

Пусть дана выборка $X_1, \ldots, X_n$, относительно распределения которой выдвинуты две простые гипотезы H_0 и H_1.

Определение. Критерий — это статистика $\varphi(\mathbf{X})$ (т.е. измеримая функция от выборки) со значениями из $[0, 1]$. Трактуется как "вероятность"отвергнуть H_0.

Выборка $\mathbf{X} = (X_1, \ldots, X_n)$ объёма n — точка в пространстве $\mathbb{R}^n$. Выделим множество $S \subset \mathbb{R}^n$ — *критическую область* для гипотезы H_0. В этом случае критерий можно сформулировать следующим образом:

- $\mathbf{X} \in S \Rightarrow$ отвергаем H_0, принимаем H_1;
- $\mathbf{X} \notin S \Rightarrow$ отвергаем H_1, принимаем H_0;

Определение. Говорят, что произошла *ошибка 1-го рода*, если критерий отверг верную гипотезу H_0. Вероятность ошибки 1-го рода (или *уровень значимости критерия*):

$$\alpha(S) = \mathbb{P}\left(\mathbf{X} \in S | H_0\right) = \mathbb{P}_0\left\{\mathbf{X} \in S\right\}$$

Аналогично вероятность ошибки 2-го рода:

$$\beta(S) = \mathbb{P}\left(\mathbf{X} \notin S | H_1\right) = \mathbb{P}_1\left(\mathbf{X} \notin S\right)$$

Определение. *Мощность критерия*:

$$\gamma(S) = 1 - \beta(S) = \mathbb{P}_1\left(\mathbf{X} \in S\right)$$

Истинная гипотеза	Результат принятия решения	
	H_0 отклонена	H_0 принята
H_0	α	$1 - \alpha$
H_1	$1 - \beta$	β

Если $\gamma(S) < \alpha(S)$, то попасть в S при условии истинности гипотезы H_1 труднее, чем при условии истинности гипотезы H_0, т.е. S — критическая область скорее для H_1. Следовательно, неравенство должно иметь вид $\gamma(S) > \alpha(S)$.

Определение. Критерий называется *несмещённым*, если выполняется условие

$$\alpha(S) \leqslant \gamma(S) = 1 - \beta(S)$$

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 дополнительно будем решать задачу $\beta(S) \to \min\limits_{S}$.

Получаем две эквивалентные задачи определения критической области S:

$$\begin{cases} \alpha_0 \geqslant \alpha(S) \\ \beta(S) \to \min\limits_{S} \end{cases} \Leftrightarrow \begin{cases} \alpha_0 \geqslant \alpha(S) \\ \gamma(S) \to \max\limits_{S} \end{cases}$$

Задачи в такой постановке не всегда решаемы, так как требуется ответить точно «да» или «нет». Такие статистические критерии называются *нерандомизированными критериями.*

Пример. Рассмотрим *критическую функцию* $\varphi(x) = I\{x \in S\}$. Тогда критерий примет вид:

- Если $\varphi\left(\mathbf{X}\right) = 1$, тогда отвергаем гипотезу H_0, принимаем H_1.
- Если $\varphi\left(\mathbf{X}\right) = 0$, тогда отвергаем гипотезу H_1, принимаем H_0.

Пример. Рассмотрим другую критическую функцию $\varphi(x) = P\left\{\overline{H}_0 | \mathbf{X} = x\right\}$. В этом случае $\varphi\left(\mathbf{X}\right) \in [0;1]$ — условная вероятность отклонения гипотезы H_0. При таком определении $\varphi(x)$ приходим к *рандомизированному критерию*, то есть, критерию, который при некоторых значениях s может не давать ответа «да» или «нет» в отношении истинности гипотезы H_0. Тогда формулировка критерия следующая:

- с вероятностью $1 - \varphi\left(\mathbf{X}\right)$ следует принимать гипотезу H_0;
- с вероятностью $\varphi\left(\mathbf{X}\right)$ следует отвергнуть гипотезу H_0.

Замечание. При использовании введенного обозначения вероятность ошибки первого рода, вероятность ошибки второго рода и мощность критерия будем обозначать: $\alpha(\varphi)$, $\beta(\varphi)$ и $\gamma(\varphi) = 1 - \beta(\varphi)$ соответственно.

Без ограничения общности будем предполагать, что существует плотность $f_0(x)$ для функции распределения $F_0(x)$, и существует плотность $f_1(x)$ для функции распределения $F_1(x)$. В дискретном случае все результаты аналогичны.

Если верна гипотеза H_1, то функция правдоподобия выборки X имеет вид:

$$L_1\left(\mathbf{X}\right) = \prod_{i=1}^{n} f_1\left(X_i\right)$$

Для рандомизированного критерия получаем

$$P_0\left(\overline{H}_0\right) = \int\limits_{\mathbb{R}^n} \varphi(x) L_0(x) \mu^n(dx) = \alpha(\varphi)$$

$$P_1(H_0) = \int\limits_{\mathbb{R}^n} (1-\varphi(x)) L_1(x) \mu^n(dx) = \beta(\varphi)$$

$$\gamma(\varphi) = \int\limits_{\mathbb{R}^n} \varphi(x) L_1(x) \mu^n(dx), \quad \gamma(\varphi) = 1 - \beta(\varphi)$$

Тогда задача построения статистического критерия сводится к нахождению критической функции $\varphi(x)$ и будет формулироваться следующим образом:

$$\begin{cases} \alpha_0 \geqslant \alpha(\varphi) \\ \beta(\varphi) \to \min\limits_{\varphi} \end{cases} \Leftrightarrow \begin{cases} \alpha_0 \geqslant \alpha(\varphi) \\ \gamma(\varphi) \to \max\limits_{\varphi} \end{cases}$$

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 $\alpha_0 \in (0;1)$, тогда при фиксированной вероятности ошибки первого рода α_0 наиболее мощный критерий имеет критическую функцию φ^* вида*

$$\varphi^*(x) = \begin{cases} 1, & \text{если } L_1(x) > cL_0(x) \\ \varepsilon, & \text{если } L_1(x) = cL_0(x) \\ 0, & \text{если } L_1(x) < cL_0(x) \end{cases}$$

где $L_j(x) = \prod_{i=1}^n f_j(x_i)$ соответствует гипотезе $H_j, j = \overline{1,2}$, константы c и ε являются решениями уравнения $\alpha(\varphi^*) = \alpha_0$.

Доказательство.

1. Покажем, что константы c и ε могут быть найдены из уравнения $\alpha(\varphi^*) = \alpha_0$. Заметим, что

$$\alpha(\varphi^*) = P_0(L_1(\mathbf{X}) > cL_0(\mathbf{X})) + \varepsilon P_0(L_1(\mathbf{X}) = cL_0(\mathbf{X})) = \\ = P_0\left(\frac{L_1(\mathbf{X})}{L_0(\mathbf{X})} > c\right) + \varepsilon P_0\left(\frac{L_1(\mathbf{X})}{L_0(\mathbf{X})} = c\right)$$

Если предположить, что $L_0(\mathbf{X}) = 0$, то

$$P_0\{L_0(\mathbf{X}) = 0\} = \int\limits_{\{x: L_0(x)=0\}} L_0(x)\mu(dx) = 0$$

и, следовательно, вышеприведённое равенство корректно. Поэтому рассмотрим случайную величину $\eta(\mathbf{X}) = \dfrac{L_1(\mathbf{X})}{L_0(\mathbf{X})}$

Положим $F_{H_0,\eta}(t) = P\{\eta \leqslant t\}$, тогда

$$\alpha\left(\varphi^*\right) = 1 - F_{H_0,\eta}(c) + \varepsilon\left(F_{H_0,\eta}(c) - F_{H_0,\eta}(c-0)\right)$$

Пусть $g(c) = 1 - F_{H_0,\eta}(c)$, 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

$$\varepsilon_{\alpha_0} = \begin{cases} 0, & \text{если } g\left(c_{\alpha_0}\right) = g\left(c_{\alpha_0} - 0\right) \\ \dfrac{\alpha_0 - g\left(c_{\alpha_0}\right)}{g\left(c_{\alpha_0} - 0\right) - g\left(c_{\alpha_0}\right)} \in [0;1], & \text{если } g\left(c_{\alpha_0}\right) < g\left(c_{\alpha_0} - 0\right) \end{cases}$$

В обоих случаях выполнено равенство:

$$\alpha_0 = g\left(c_{\alpha_0}\right) + \varepsilon_{\alpha_0}\left(g\left(c_{\alpha_0} - 0\right) - g\left(c_{\alpha_0}\right)\right) = \alpha\left(\varphi^*\right)$$

2. Докажем, что $\varphi^*(x)$ — критическая функция наиболее мощного критерия.

   Выберем любую другую критическую функцию $\tilde{\varphi}(x)$ такую, что $\alpha(\tilde{\varphi}) \leqslant \alpha_0$, и сравним ее с критической функцией $\varphi^*(x)$. Заметим, что для любого x справедливо неравенство:

$$\left(\varphi^*(x) - \tilde{\varphi}(x)\right)\left(L_1(x) - c_{\alpha_0} L_0(x)\right) \geqslant 0$$

   Тогда

$$\int\limits_{\mathbb{R}^n} \left(\varphi^*(x) - \tilde{\varphi}(x)\right)\left(L_1(x) - c_{\alpha_0} L_0(x)\right)\mu^n(dx) \geqslant 0$$

   Раскроем скобки и преобразуем:

$$\int\limits_{\mathbb{R}^n} \varphi^*(x) L_1(x) \mu^n(dx) - \int\limits_{\mathbb{R}^n} \tilde{\varphi}(x) L_1(x) \mu^n(dx) \geqslant$$
$$\geqslant c_{\alpha_0}\left(\int\limits_{\mathbb{R}^n} \varphi^*(x) L_0(x) \mu^n(dx) - \int\limits_{\mathbb{R}^n} \tilde{\varphi}(x) L_0(x) \mu^n(dx)\right) \geqslant 0$$

Следовательно, $\gamma(\varphi^*) - \gamma(\tilde{\varphi}) \geqslant c_{\alpha_0}(\alpha(\varphi^*) - \alpha(\tilde{\varphi}))$, откуда получаем неравенство:

$$\gamma(\varphi^*) \geqslant \gamma(\tilde{\varphi})$$

■

1.11 Критерии согласия Колмогорова и χ^2

Пусть для наблюдаемого распределения $\mathbb{P}_\xi$ дана выборка $X_1, \ldots, X_n$, проверяется *гипотеза согласия* $H_0 : F_\xi = F_0$, где F_0 известна; альтернатива $H_1 : F_\xi \neq F_0$.

Критерий χ^2

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 $\overline{\chi}^2$:

$$\overline{\chi}^2(\mathbf{X}) = \sum_{i=1}^{k} \frac{\left(n_i - np_i^{(0)}\right)^2}{np_i^{(0)}},$$

где n_i — число зафиксированных наблюдений в i-м интервале, $p_i^{(0)} = F_0(a_i) - F_0(a_{i-1})$ — вероятность попадания наблюдения в i-й интервал при выполнении гипотезы H_0, $np_i^{(0)}$, соответственно, ожидаемое число попаданий в i-й интервал.

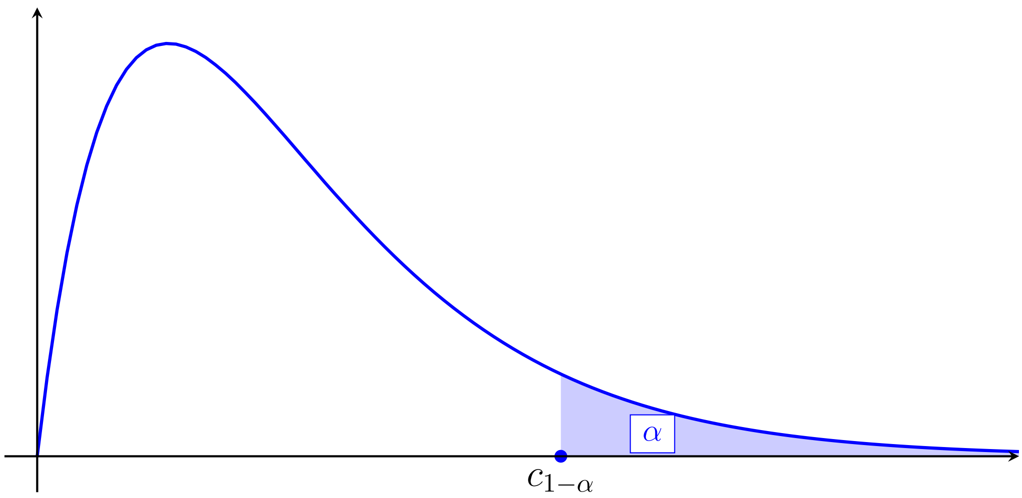
Формулировка критерия:

- Если верна гипотеза H_0, то $\overline{\chi}^2(\mathbf{X}) \xrightarrow[n\to\infty]{\mathrm{d}} \zeta$, где ζ подчиняется распределению χ^2 с $k-1-r$ степенями свободы (k — число интервалов разбиения, r — число параметров предполагаемого закона распределения);
- Если верна гипотеза H_1, то $\overline{\chi}^2 \xrightarrow[n\to\infty]{\text{п.н.}} \infty$.

Выберем вероятность $\alpha \in (0; 1)$. Область $(C(k-1, 1-\alpha), \infty)$, где $C(k-1, 1-\alpha)$ — квантиль порядка $1-\alpha$ распределения χ^2 с $k-1-r$ степенями свободы, является критической для гипотезы H_0.

Правило проверки гипотез:

- Если $\overline{\chi}^2(\mathbf{X}) > C(r-1, 1-\alpha)$, то H_0 отклоняется;
- Если $\overline{\chi}^2(\mathbf{X}) \leqslant C(r-1, 1-\alpha)$, 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. 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

Гипотеза H_0: стрельба случайна (нет «целевых» участков).

Высчитаем теоретические вероятности по закону редких событий (распределение Пуассона):

$$p_i^{(0)} = \mathbb{P}\{S = i\} = \frac{\lambda^i}{i!}e^{-\lambda}, \text{ где } S \text{ — число попаданий}, \ \lambda = \overline{X} \approx 0,932$$

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 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.

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 (*коррекция Йетса*).

Критерий Колмогорова

Наложим дополнительное условие на исходную задачу: $F_0(x) \in C(\mathbb{R})$.

Рассмотрим статистику Колмогорова:

$$D_n(\mathbf{X}) = \sup_{x\in R} |F_n^*(x) - F_0(x)|$$

Формулировка критерия:

- Если верна гипотеза H_0, то $D_n(\mathbf{X}) \xrightarrow[n\to\infty]{\text{п.н.}} 0$;
- Если верна гипотеза H_1, т.е. $F_\xi \equiv G \neq F_0$, то

$$D_n(\mathbf{X}) \xrightarrow[n\to\infty]{\text{п.н.}} \sup_{x\in R} |G(x) - F_0(x)| > 0$$

Лемма. *Если гипотеза H_0 верна, и $F_0(x) \in C(\mathbb{R})$, то распределение статистики $D_n = \sup_{x\in R} |F_n^*(x) - F_0(x)|$ 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:*

$$P\left\{\sqrt{n}D_n(\mathbf{X}) \leqslant z\right\} \xrightarrow[n\to\infty]{} K(z) = 1 + 2\sum_{m=1}^{\infty}(-1)^m e^{-2m^2z^2}$$

Находим константу $d_{1-\alpha}$ как решение уравнения $K(d_{1-\alpha}) = 1 - \alpha$.

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. Приведена таблица результатов исследования при $n = 100$:

Количество предметов	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.

1.12 Статистические выводы о параметрах нормального распределения. Распределения χ^2 и Стьюдента. Теорема Фишера

Определение. Говорят, что случайная величина ξ имеет *гамма-распределение* с параметрами $\alpha > 0$, $\lambda > 0$ $(\mathbf{\Gamma}(\alpha, \lambda))$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \begin{cases} 0, & \text{если } x \leqslant 0 \\ c \cdot x^{\lambda-1} e^{-\alpha x}, & \text{если } x > 0 \end{cases},$$

где постоянная c вычисляется из свойства нормировки плотности:

$$1 = \int\limits_{-\infty}^{\infty} f_\xi(x) dx = c \int\limits_0^\infty x^{\lambda-1} e^{-\alpha x} dx = \frac{c}{\alpha^\lambda} \int\limits_0^\infty (\alpha x)^{\lambda-1} e^{-\alpha x} d(\alpha x) = \frac{c}{\alpha^\lambda} \Gamma(\lambda),$$

откуда $c = \alpha^\lambda / \Gamma(\lambda)$.

Лемма. *Пусть $\xi_1, \ldots, \xi_n$ независимы, и $\xi_i \sim \mathbf{\Gamma}(a, \lambda_i), i = \overline{1,n}$. Тогда их сумма $S_n = \xi_1 + \ldots + \xi_n \sim \mathbf{\Gamma}(a, \lambda_1 + \ldots + \lambda_n)$*

Лемма. *Если $\xi \sim \mathbf{N}(0,1)$, то $\xi^2 \sim \mathbf{\Gamma}(1/2, 1/2)$.*

Следствие. *Если $\xi_1, \ldots, \xi_k$ независимы и $\xi_i \sim \mathbf{N}(0,1)$, то случайная величина $\chi^2 = \xi_1^2 + \ldots + \xi_k^2 \sim \mathbf{\Gamma}(1/2, k/2)$*

Определение. Распределение суммы k квадратов независимых случайных величин со стандартным нормальным распределением называется *распределением хи-квадрат* с k степенями свободы ($\chi^2(k)$).

Плотность распределения $\chi^2(k)$ имеет вид

$$f(y) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)} y^{\frac{k}{2}-1} e^{-y/2}, & \text{если } y > 0 \\ 0, & \text{если } y \leqslant 0 \end{cases}$$

Замечание. $\chi^2(2) = \mathbf{\Gamma}(1/2, 1) = \mathbf{Exp}(1/2)$

Свойства распределения χ^2.

1. *Если случайные величины $\xi_1 \sim \chi^2(k)$ и $\xi_2 \sim \chi^2(m)$ независимы, то их сумма $\xi_1 + \xi_1 \sim \chi^2(k+m)$;*
2. $\mathbb{E}\chi^2 = k, \quad \mathbb{D}\chi^2 = 2k$
3. *Пусть дана последовательность случайных величин χ_n^2. Тогда при $n \to \infty$:*
$$\frac{\chi_n^2}{n} \xrightarrow{\mathbb{P}} 1, \quad \frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow \mathrm{N}_{0,1}$$
4. *Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы и $\xi_i \sim \mathbf{N}(a, \sigma^2)$. Тогда*
$$\sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma}\right)^2 \sim \chi^2(k)$$

Определение. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и $\xi_i \sim \mathbf{N}(0,1)$. 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 ($\mathbf{T}(k)$).

Плотность распределения $\mathbf{T}(k)$ имеет вид

$$f_k(y) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{y^2}{k}\right)^{-(k+1)/2}$$

Свойства распределения Стьюдента.

1. *Распределение Стьюдента симметрично, т.е. если* $t_k \sim \mathbf{T}(k)$, *то* $-t_k \sim \mathbf{T}(k)$.

2. $\mathbf{T}(k) \Rightarrow \mathbf{N}(0,1)$ *при* $k \to \infty$.

3. *У распределения Стьюдента* $\mathbf{T}(k)$ *существуют только моменты порядка* $m < k$, *при этом все существующие моменты нечётного порядка равны нулю.*

Теорема Фишера. *Пусть случайные величины* $\xi_1, \ldots, \xi_n$ *независимы и* $\xi_i \sim \mathbf{N}(a, \sigma^2)$. *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* $\xi_1, \ldots, \xi_n$ *независимы и* $\xi_i \sim \mathbf{N}(a, \sigma^2)$. *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 $\chi^2(n)$ 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 $\chi^2(n-1)$ 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 $\mathrm{T}(n-1)$ уровня $\frac{1-\alpha}{2}$. Распределение Стьюдента симметрично, поэтому

$$\alpha = \mathbb{P}\left(-c < \sqrt{n}\frac{\overline{X} - a}{S_0} < c\right) = \mathbb{P}\left(\overline{X} - \frac{cS_0}{\sqrt{n}} < a < \overline{X} + \frac{cS_0}{\sqrt{n}}\right)$$

Глава 2

Теория вероятностей

2.1 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. $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$; (здесь $\overline{A} = \Omega \setminus A$ — *дополнение* к A)
3. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ (по индукции: $A_1, A_2, \ldots, A_n \in \mathcal{A} \Rightarrow \bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$).

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

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 $A_1, A_2, \ldots, A_n, A_{n+1} = \varnothing, A_{n+2} = \varnothing, \ldots \in \mathcal{A}$. По определению σ-алгебры: $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F} \Rightarrow \bigcup\limits_{i=1}^{n} A_i \in \mathcal{F}$, следовательно, выполнен третий пункт определения алгебры.

Определение. *Случайное событие* A — элемент σ-алгебры $\mathcal{F}$. $A = \varnothing$ — *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 тогда и только тогда, когда происходят или A, или B, или оба вместе.
- *Пересечение* $A \cap B$ (или AB) — происходит тогда и только тогда, когда происходят и A и B вместе. Если $AB = \varnothing$, то события A и B называются *несовместными.*
- *Разность* $A \setminus B$ — происходит тогда и только тогда, когда происходит A и не происходит B.
- *Симметрическая разность* $A \triangle B$ — происходит тогда и только тогда, когда 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 $\sigma(K) = \{\varnothing, \Omega, A, \overline{A}\}$.

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

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

   (*счётная аддитивность — вероятность счётного объединения несовместных событий равна сумме их вероятностей*)

Замечание. Из счётной аддитивности, как уже было доказано выше, следует и конечная аддитивность.

Свойства вероятности.

1. $\mathbb{P}(\varnothing) = 0$*;*

2. $A, B \in \mathcal{F}, B \subset A \Rightarrow \mathbb{P}(A) \geqslant \mathbb{P}(B)$ *(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. $\forall\, A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots :\ \bigcap\limits_{n=1}^{\infty} A_n = A \quad \Rightarrow$

$\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ *(непрерывность).*

Доказательство.

1. Рассмотрим последовательность событий $A_1 = \Omega, A_2 = \varnothing, \ldots, A_n = \varnothing, \ldots$:
$$\bigcup_{i=1}^{\infty} A_i = \Omega \Rightarrow \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \mathbb{P}(\Omega) = 1.$$
При этом $A_iA_j = \varnothing$ $(i \neq j)$, следовательно, по пункту 3 определения вероятности: $\sum\limits_{i=2}^{\infty} \mathbb{P}(\varnothing) = 0 \Rightarrow \mathbb{P}(\varnothing) = 0$.

2. $B \subset A \Rightarrow A = (A \setminus B) \cup B$. 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 $C_l = A_l \setminus A_{l+1} \Rightarrow C_l \cap C_m = \varnothing$). 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:

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

Определение. Тройка $(\Omega, \mathcal{F}, \mathbb{P})$ — *вероятностное пространство.*

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:

- $\Omega = [0,1] \neq \varnothing$
- $\mathfrak{B}_{[0,1]}$ — σ-алгебра над $\Omega = [0,1]$
- $\lambda(\Omega) = \lambda_{[0,1]} = 1$, $\forall A \in \mathfrak{B}_{[0,1]}$ $\lambda(A) \geqslant 0$ 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 $\mathbb{P}(\{\omega_i\}) = p_i$. Вероятностное пространство, определённое таким образом, — *дискретное вероятностное пространство.* Тогда для любого события $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$ его вероятность $\mathbb{P}(A) = \sum\limits_{j=1}^{k} p_{i_j}$.

Определение. *Классическое определение вероятности*:

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

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

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

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

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 $A_1, A_2, \ldots A_n, \ldots$; $A_i \cap A_j = \varnothing$ $(i \neq j)$. Тогда:

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

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

$\blacksquare$

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 $\forall k \in \overline{2, n}$ и $\forall i_1, \ldots, i_k \colon 1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ 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 $AB = \varnothing \Rightarrow \mathbb{P}(AB) = 0$. Но $\mathbb{P}(A)\mathbb{P}(B) = 0 \cdot \mathbb{P}(B) = 0 \Rightarrow \mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

   Если же $\mathbb{P}(A) = 0$, то $AB \subset A \Rightarrow \mathbb{P}(AB) \leqslant \mathbb{P}(A) = 0$. 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

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

   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 $A \subset B$: $\mathbb{P}(AB) = \mathbb{P}(A)$, 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 вероятностное пространство, в котором всего 4 различных элементарных исхода: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Пусть $\mathcal{F}$ — множество всех подмножеств Ω, $\mathbb{P}(\{\omega_i\}) = \dfrac{1}{4}$, $i = \overline{1, 4}$.

Рассмотрим три события

$$A_1 = \{\omega_1, \omega_4\},\ A_2 = \{\omega_2, \omega_4\},\ A_3 = \{\omega_3, \omega_4\}$$

Их пересечения имеют вид:

$$A_1A_2 = A_2A_3 = A_3A_1 = \{\omega_4\},\ A_1A_2A_3 = \{\omega_4\}$$

Докажем, что события A_1, A_2, A_3 не являются независимыми в совокупности:

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{2},\ \mathbb{P}(A_1A_2) = \mathbb{P}(A_2A_3) = \mathbb{P}(A_3A_1) = \frac{1}{4},$$
$$\mathbb{P}(A_1A_2A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

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* $\Leftrightarrow \forall\ \delta_1, \delta_2, \ldots \delta_n \in \{0, 1\}$ *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:*

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

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

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

■

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:*

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

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

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

■

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.

2.3 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 $f : \mathbb{R} \to \mathbb{R}$:

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

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 $D : \mathbb{R} \to \{0, 1\}$

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

является борелевской.

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

Определение. Функция $\xi\colon \Omega \mapsto \mathbb{R}^{(n)}$ называется *измеримой относительно σ-алгебры $\mathcal{F}$*, если полный прообраз борелевского множества B лежит в σ-алгебре $\mathcal{F}$, т.е.

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

Замечание. Борелевская функция — это функция, измеримая относительно борелевской σ-алгебры.

Случайные величины

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

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

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

$\Omega = [0; 1], \mathcal{F} = \{\varnothing, \Omega\}$ — 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 $A_1 = (-\infty, \frac{1}{3})$ — его прообразом является множество $(\frac{1}{2}, 1]$. Но это множество не входит в $\mathcal{F}$, а значит, $\xi(\omega)$ неизмерима.

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

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, $\mathbb{P}(\xi < \frac{1}{3})$. Ведь $\mathbb{P}(\xi < \frac{1}{3}) = \mathbb{P}(\xi \in A_1) = \mathbb{P}(\omega \in (\frac{1}{2}, 1])$, но множество $(\frac{1}{2}, 1] \notin \mathcal{F}$, 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 сигма-алгебре, то есть

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

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

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

Но функция g по предположению борелевская, а значит, 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 прооборaз произвольного борелесвкого множества принадлежит σ-алгебре. Значит, $\eta = g(\xi)$ — случайная величина. ∎

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

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

Доказательство.

$\Leftarrow$ 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 и случайной величины ξ:

$$\xi^{-1}\left(\bigcup_\alpha A_\alpha\right) = \bigcup_\alpha \xi^{-1}(A_\alpha), \quad \xi^{-1}(\overline{A}) = \overline{\xi^{-1}(A)},$$
$$\xi^{-1}\left(\bigcap_\alpha A_\alpha\right) = \bigcap_\alpha \xi^{-1}(A_\alpha),$$

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

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

∎

Следствие. *ξ — случайная величина $\Leftrightarrow \forall x \in \mathbb{R}\colon \{\omega\colon \xi(\omega) < x\} \in \mathcal{F}$. 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. $\mathcal{F}_\xi \subset \mathcal{F}$.

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. Можно рассматривать распределение как композицию отображений. Если $\xi : \Omega \mapsto \mathbb{R}$, то полный прообраз — это отображение $\xi^{-1} : \mathfrak{B} \mapsto \mathcal{F}$. В свою очередь, $\mathbb{P} : \mathcal{F} \mapsto \mathbb{R}$. Тогда

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

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

Функция распределения, её свойства

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

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

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. *$\forall x\ 0 \leqslant F_\xi(x) \leqslant 1$;*

2. *$F_\xi(x)$ монотонно неубывает (т.е. $x_1 < x_2 \Rightarrow F_\xi(x_1) \leqslant F_\xi(x_2)\ \forall x_1, x_2$);*

3. $\lim\limits_{x\to+\infty} F_\xi(x) = 1, \lim\limits_{x\to-\infty} F_\xi(x) = 0$*;*

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

Доказательство.

1. Следует из свойств вероятности.

2. $x_1 < x_2 \Rightarrow \{\xi < x_1\} \subseteq \{\xi < x_2\}$. 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 $B_n = \{\xi < -n\}$, $B_{n+1} = \{\xi < -(n+1)\} \subseteq B_n = \{\xi < -n\}$ $\forall n \ \geqslant 1$:

$$\bigcap_{j=1}^{\infty} B_j = \{\omega \colon \xi(\omega) < x, \forall x \in \mathbb{R}\} \Rightarrow \bigcap_{j=1}^{\infty} B_j = \varnothing.$$

   $F_\xi(-n) = \mathbb{P}(B_n) \xrightarrow[n\to+\infty]{} \mathbb{P}(B) = 0$ (в силу непрерывности вероятностной меры)

   Отсюда следует: $F_\xi(n) \xrightarrow[n\to+\infty]{} 1 \Leftrightarrow 1 - F_\xi(n) = \mathbb{P}(\xi \geqslant n) \xrightarrow[n\to+\infty]{} 0$.

4. Докажем, что $F_\xi(x_0 - \frac{1}{n}) \xrightarrow[n\to+\infty]{} F_\xi(x_0)$, что равносильно

$$F_\xi(x_0) - F_\xi\left(x_0 - \frac{1}{n}\right) = \mathbb{P}(\xi < x_0) - \mathbb{P}\left(\xi < x_0 - \frac{1}{n}\right) =$$
$$= \mathbb{P}\left(x_0 - \frac{1}{n} \leqslant \xi < x_0\right) \xrightarrow[n\to+\infty]{} 0$$

   Сходимость выполняется в силу непрерывности вероятностной меры.

■

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 $f(\omega) \colon \Omega \mapsto \mathbb{R}$?

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

2. *$f_\xi(x) \overset{п.н.}{\geqslant} 0\ \forall x$ (неотрицательность);*

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

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

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

■

Замечание. Заметим, что любая функция распределения дифференцируема почти всюду, поэтому 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. Например, далее мы увидим, что функция распределения сингулярного распределения непрерывна и дифференцируема почти всюду, однако плотности у этого распределения нет, так как производная функции распределения почти всюду равна нулю.

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

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

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

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

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

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

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

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*

$$F_\xi(x) = p_1 F_{ac}(x) + p_2 F_s(x) + p_3 F_d(x).$$

2.5 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 $\xi : \Omega \mapsto \mathbb{R}$. 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) сходящегося ряда

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

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

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

Математическое ожидание 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 распределения), то точка $\mathbb{E}\xi$ будет координатой «центра тяжести» прямой.

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

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:

$$\mathbb{E}(\xi + \eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x + y) f(x, y) dx dy \tag{2.1}$$

2. *Математическое ожидание линейно:*

$$\mathbb{E}(a\xi + b) = a\mathbb{E}\xi + b \quad \forall a, b \in \mathbb{R} \tag{2.2}$$

3. $\mathbb{E}(\xi + \eta) = \mathbb{E}\xi + \mathbb{E}\eta$, *при условии, что* $\mathbb{E}\xi$ *и* $\mathbb{E}\eta$ *существуют.*

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* $\mathbb{E}\xi \leqslant \mathbb{E}\eta$.

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

5. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.*

   Замечание. *Обратное неверно: из равенства* $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$ *не следует независимость величин* ξ *и* η.

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

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

8. $\mathbb{P}(A) = \mathbb{E}\left[\mathrm{I}_A(\omega)\right]$.

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) и теоремой о совместном распределении:

$$\mathbb{E}(\xi + \eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x+y)f(x,y)dxdy =$$
$$= \int\limits_{-\infty}^{\infty} xdx \int\limits_{-\infty}^{\infty} f(x,y)dy + \int\limits_{-\infty}^{\infty} ydy \int\limits_{-\infty}^{\infty} f(x,y)dx =$$
$$= \int\limits_{-\infty}^{\infty} xf_\xi(x)dx + \int\limits_{-\infty}^{\infty} yf_\eta(y)dy = \mathbb{E}\xi + \mathbb{E}\eta.$$

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) заменим сложение умножением и плотность совместного распределения произведением плотностей (это возможно в силу независимости случайных величин):

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

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

аналогичного неравенства для интегралов (для непрерывного случая).

7. Это свойство мы докажем, заранее предполагая, что ξ имеет дискретное распределение с неотрицательными значениями $a_k \geqslant 0$. 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 $a_k = 0$.

8. Это верно в силу определений функции-индикатора и матожидания.

9. Начнём с 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

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

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

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*.

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

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 5 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

Замечание. Во всех свойствах предполагается существование вторых моментов случайных величин. Тогда (в силу вышеописанной теоремы) существуют и сами матожидания.

1. Дисперсия может быть вычислена по формуле: $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$.

   Доказательство. 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. 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: $\mathbb{D}\xi \geqslant 0$.

   Доказательство. Пусть $a = \mathbb{E}\xi$. Дисперсия есть математичекое ожидание неотрицательной случайной величины $(\xi - a)^2$, откуда (и из свойства 5 матожидания) следует неотрицательность дисперсии. ■

4. Дисперсия обращается в нуль лишь для вырожденного распределения: если $\mathbb{D}\xi = 0$, то $\xi \overset{\text{п.н.}}{=} \mathrm{const}$, и наоборот.

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

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

6. Если ξ и η независимы, то $\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta$.

   Доказательство. Действительно, применяя свойство (6) матожидания, получим:

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

∎

Замечание. Обратное, аналогично замечанию к свойству (6) матожидания, неверно.

Следствие. *Если ξ и η независимы, то $\mathbb{D}(\xi - \eta) = \mathbb{D}\xi + \mathbb{D}\eta$.*

Доказательство. Из свойств (6) и (2) получим:

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

∎

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

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

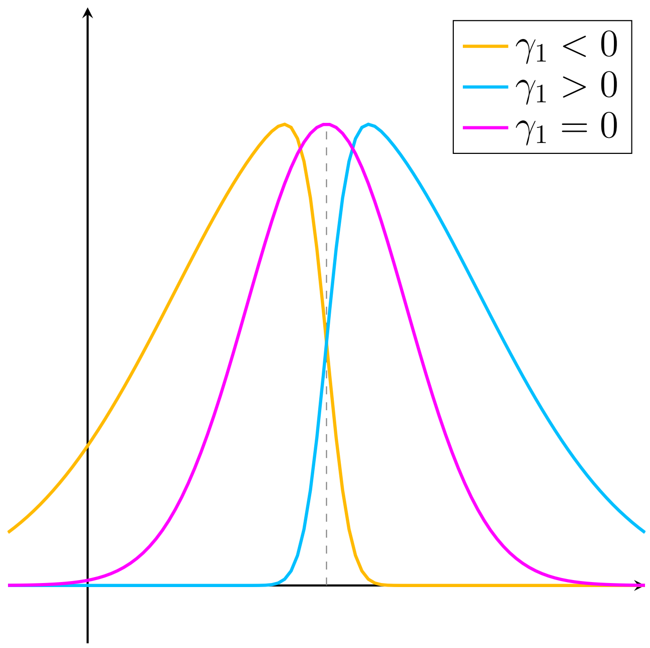
Замечание. В последнем равенстве величина $\mathbb{E}(\xi\eta) - \mathbb{E}\xi\mathbb{E}\eta$ есть *ковариация* случайных величин ξ и η — $\operatorname{cov}(\xi, \eta)$.

Другие числовые характеристики

Определение. *Коэффициент асимметрии* случайной величины ξ:

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

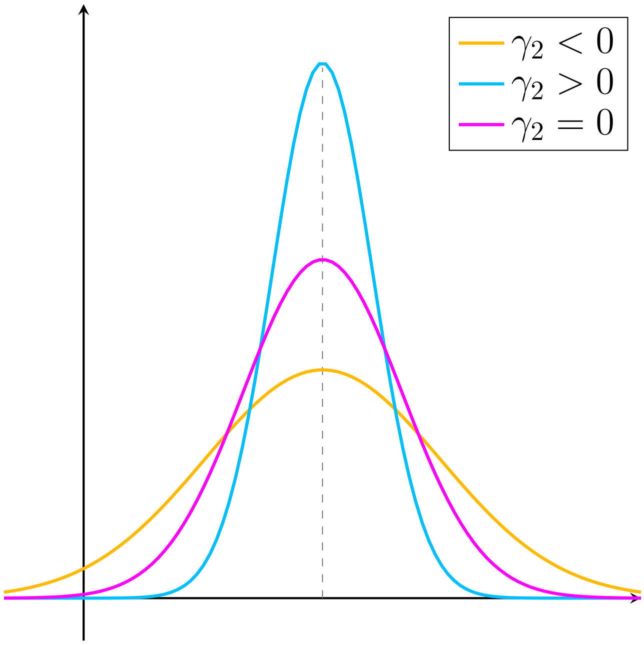
Характеризует «скошенность» графика плотности распределения:



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.

2.6 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

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

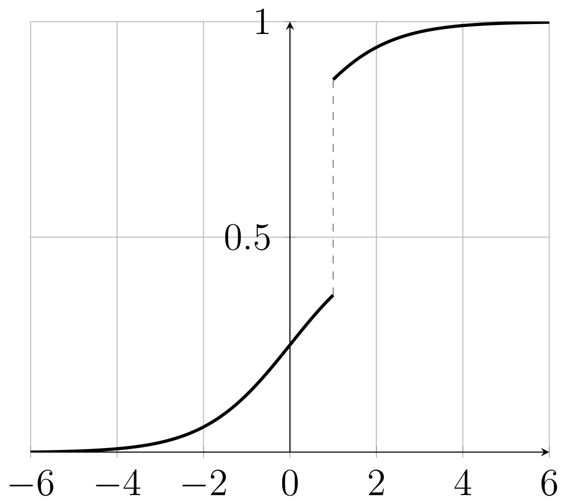
Замечание. Медиана распределения всегда существует, но может быть не единственна, к примеру, в случае дискретного распределения.

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

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

Замечание. Если функция распределения F непрерывна и строго монотонна, то *квантилем* уровня (порядка) 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 её отклонения:*

$$\mathbb{E}|\xi - \operatorname{Med}\xi| = \min_a \mathbb{E}|\xi - a|;$$

2. *Отклонение медианы случайной величины ξ от её математического ожидания $\mathbb{E}\xi$ 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

$$\begin{aligned} |\eta - c| - |\eta| &= c, \quad \eta < 0 \\ |\eta - c| - |\eta| &\geqslant -c, \quad \eta \geqslant 0. \end{aligned}$$

Тогда

$$\begin{aligned} \mathbb{E}\big(|\eta - c| - |\eta|\big) &= \mathbb{E}\big((|\eta - c| - |\eta|) \cdot \mathrm{I}(\eta < 0) + (|\eta - c| - |\eta|) \cdot \mathrm{I}(\eta \geqslant 0)\big) \\ \mathbb{E}\big(|\eta - c| - |\eta|\big) &\geqslant c\, \mathbb{P}(\eta < 0) - c\, \mathbb{P}(\eta \geqslant 0). \end{aligned}$$

Так как $\operatorname{Med}\eta = 0$, то $\mathbb{P}(\eta \leqslant 0) = \mathbb{P}(\eta \geqslant 0) = \frac{1}{2}$. Отсюда вытекает

$$\mathbb{E}\big(|\eta - c| - |\eta|\big) \geqslant 0.$$

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:

$$\begin{aligned} |\mathbb{E}\xi - \operatorname{Med}\xi| = |\mathbb{E}\left[\operatorname{Med}\xi - \xi\right]| &\leqslant \{\text{шестое мат. ожидания}\} \\ \leqslant \mathbb{E}|\operatorname{Med}\xi - \xi| &\leqslant \{\text{первое свойство медианы}\} \\ \leqslant \mathbb{E}|\mathbb{E}\xi - \xi| &\leqslant \{\text{неравенство Йенсена}\} \\ &\leqslant \sqrt{\mathbb{E}|\mathbb{E}\xi - \xi|^2} = \sqrt{\mathbb{D}\xi} = \sigma. \end{aligned}$$

■

Интерквантильный размах

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.

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

Определение. *Схема Бернулли* — последовательность независимых испытаний, в каждом из которых возможны лишь два исхода — «успех» и «неудача», при этом успех в каждом испытании происходит с одной и той же вероятностью $p \in (0;1)$, а неудача — с вероятностью $q = 1 - p$.

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*

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

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 $p^k q^{n-k}$. ∎

Определение. Пусть $\xi_1, \ldots, \xi_n$ — последовательность независимых случайных величин, имеющих одинаковое распределение Бернулли с параметром p ($\mathbf{Bi}(p)$), 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 ($\mathbf{B}(n,p)$).

Числовые характеристики $\mathbf{Bi}(p)$

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 $\mathbf{B}(n, p)$

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 $np_n \underset{n\to+\infty}{\longrightarrow} \lambda$. Тогда*

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

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

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

■

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

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:

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

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

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

Рассмотрим 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 $\mathbb{P}(\tau = k) = pq^{k-1}$.

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 будем принимать значения из множества $\{0, 1, 2, \ldots\}$.

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 мат. ожидание 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 $\lambda > 0$ ($\mathbf{Exp}(\lambda)$), 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 $\mathbf{Exp}(\lambda)$

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

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

В последнем равенстве была использована формула для гамма-функции:

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

1. Математическое ожидание:
$$\mathbb{E}\xi = \frac{1}{\lambda}$$

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

Утверждение. *Пусть проводится n обобщённых испытаний Бернулли (т.е. вероятность успеха испытания зависит от n) с вероятностью успеха p_n, $\xi \sim \mathbf{Geom}(p_n)$, $np_n \xrightarrow[n\to+\infty]{} \lambda > 0$. Тогда распределение случайной величины $\frac{\xi}{n}$ сходится к показательному с параметром λ при $n \to +\infty$.*

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

$$F_n(x) = \mathbb{P}\left(\frac{\xi_n}{n} \leqslant x\right) = \mathbb{P}\left(\xi_n \leqslant nx\right) = \mathbb{P}\left(\xi_n \leq \lfloor nx \rfloor\right) = 1 - (1-p_n)^{\lfloor nx \rfloor}$$

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, $nx - 1 < \lfloor nx \rfloor \leqslant nx$. и таким образом верно $(1-p_n)^{\lfloor nx \rfloor} \xrightarrow[n\to+\infty]{} e^{-\lambda x}$. Следовательно, $F_n(x) \xrightarrow[n\to+\infty]{} 1 - e^{-\lambda x}$, что есть функция показательного распределения. ■

Теорема Реньи. *Пусть даны случайная величина $N \sim \mathbf{Geom}(p)$, $\xi_1, \xi_2, \ldots$ —*

независимые одинаково распределённые случайные величины, $\xi_i \geqslant 0$ *и* $0 < a = \mathbb{E}\xi < \infty$, $S_n = \sum\limits_{i=1}^{N} \xi_i$. *Тогда*

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

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

Если $b^2 = \mathbb{E}\xi_i^2$, *тогда*

$$\sup_x \left|\mathbb{P}\left(\frac{p}{a} S_N < x\right) - G(x)\right| \leqslant \frac{pb^2}{(1-p)a^2}$$

2.9 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. *Если выполнено условие локальной теоремы и* C *- произвольная положительная константа, то равномерно по* a *и* b *из отрезка* $[-C, C]$ *(пусть* $b \geqslant a$*)*

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

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

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

Матожидание и дисперсия нормального распределения

Найдем матожидание и дисперсию для *стандартного* нормального распределения, т.е. для нормального распределения с параметрами $\alpha = 0$ и $\sigma^2 = 1$:

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

Далее,

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

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

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

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

2.10 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 $(\xi_1, \dots, \xi_n)$ — функция $\mathbb{P} : \mathfrak{B}(\mathbb{R}^n) \to \mathbb{R}^n$:

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

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

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

Замечание. Для функции совместного распределения выполняются свойства, аналогичные одномерному случаю. При этом частные функции распределения восстанавливаются по совместной следующим образом:

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

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

$$\mathbb{P}\left((\xi_1, \xi_2) \in B\right) = \iint\limits_B f_{\xi_1, \xi_2}(x, y) dx dy$$

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 $B_1, \ldots, B_n \in \mathfrak{B}(\mathbb{R})$:

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

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

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

В частности, в случае дискретного совместного распределения:

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

В случае абсолютно непрерывного:

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

Формула свёртки

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

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

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

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 $g\left(\xi_1, \xi_2\right)$ 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 $v\in(-\infty,x-u)\Leftrightarrow t\in(-\infty,x), dv=dt$. 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

Рассмотрим случайные величины ξ и η. Дисперсия их суммы в общем случае равна

$$\mathbb{D}(\xi+\mathrm{n})=\mathbb{D}\xi+\mathbb{D}\eta+2(\mathbb{E}(\xi\eta)-\mathbb{E}\xi\mathbb{E}\eta)$$

Определение. Величина $\mathbb{E}(\xi\eta)-\mathbb{E}\xi\mathbb{E}\eta=\operatorname{cov}(\xi,\eta)$ называется *ковариацией* случайных величин ξ и η.

Если ξ и η независимы, то $\operatorname{cov}(\xi,\eta)=0$. Обратное, вообще говоря, неверно.

Пример. Рассмотрим $\xi\sim\mathbf{U}[0;1]$, случайные величины $\eta_1=\cos\xi$ и $\eta_2=\sin\xi$.

1. Докажем некореллированность данных случайных величин.

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

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

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. $\operatorname{cov}(a\xi + b, \eta) = a\operatorname{cov}(\xi, \eta)$*, если* $a, b \in \mathbb{R}$*;*
4. $\operatorname{cov}(\eta + \zeta, \xi) = \operatorname{cov}(\eta, \xi) + \operatorname{cov}(\zeta, \xi)$*;*
5. $\operatorname{cov}^2(\xi, \eta) \leqslant \mathbb{D}\xi\,\mathbb{D}\eta$*,*
   $\operatorname{cov}^2(\xi, \eta) = \mathbb{D}\xi\,\mathbb{D}\eta \;\Leftrightarrow\; \xi \overset{\text{п.н.}}{=} a\eta + b,\; a, b \in \mathbb{R}$*;*
   (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 никак, 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 $|\rho(X, Y)| = 1$, то с вероятностью один X и Y линейно выражаются друг через друга. То есть,*

$$|\rho(X, Y)| = 1 \Longrightarrow \exists\, b \neq 0,\, c \in \mathbb{R} : \mathbb{P}(X - bY = c) = 1$$

При этом знак коэффициента b 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. $\mathbb{E}\xi_c = \mathbb{E}\eta_c = 0$:

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

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

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

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

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

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

$$(\mathbb{E}(\xi_c\eta_c))^2 - \mathbb{E}\xi_c^2\mathbb{E}\eta_c^2 \leqslant 0 \Longleftrightarrow |\mathbb{E}(\xi_c\eta_c)| \leqslant \sqrt{\mathbb{E}\xi_c^2\mathbb{E}\eta_c^2} \Rightarrow$$
$$\Rightarrow |\operatorname{cov}(\xi_c, \eta_c)| \leqslant \sqrt{\mathbb{D}\xi_c\mathbb{D}\eta_c}$$

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 может обращаться в равенство при некотором $a = b$. Но это равенство означает, что равна нулю $\mathbb{D}(\xi - b\eta)$, а это сразу говорит о том, что с вероятностью один $\xi - b\eta$ равна константе. Обозначим эту константу за c 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

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

■

2.11 Виды сходимости последовательностей случайных величин

Определение. Последовательность случайных величин $\{\xi_n\}$ *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 $\{\xi_n\}$ *сходится по вероятности* к случайной величине ξ ($\xi_n \xrightarrow{\text{p}} \xi$), если

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

Определение. Последовательность случайных величин $\{\xi_n\}$ *сходится в среднем* к случайной величине ξ ($\xi_n \xrightarrow{(\text{r})} \xi$), если

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

Определение. Последовательность случайных величин $\{\xi_n\}$ *сходится по распределению* к случайной величине ξ ($\xi_n \xrightarrow{\text{d}} \xi$), 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 $\{\xi_n\}$ *слабо сходится* к случайной величине ξ ($\xi_n \overset{\mathrm{w}}{\Rightarrow} \xi$), 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:*

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

Доказательство.

(r) $\Rightarrow$ p Из обобщённого неравенства Чебышёва:

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

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

$$\xi_n = \begin{cases} 0, & \frac{1}{n} \leqslant \omega \leqslant 1; \\ \sqrt[r]{n}, & 0 \leqslant \omega \leqslant \frac{1}{n}. \end{cases} \Rightarrow p_1 = 1 - \frac{1}{n},\ p_2 = \frac{1}{n}$$

$$\mathbb{P}(|\xi_n| > \varepsilon) \xrightarrow[n\to+\infty]{} 0, \text{ однако } \mathbb{E}|\xi_n|^r = 1.$$

п.н. $\Rightarrow$ p Ограничимся для простоты случаем, когда $\xi_n(\omega) \to \xi(\omega)$ для любого ω. Зафикисируем $\omega \in \Omega$. По определению предела, $\xi_n(\omega) \xrightarrow[n\to+\infty]{} \xi(\omega)$, если для всякого $\varepsilon > 0$ найдётся $N = N(\omega, \varepsilon) \geqslant 0$ такое, что для всех $n > N$ выполняется неравенство $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$.

Событие $A = \{n > N(\omega, \varepsilon)\}$ влечёт событие $B = \{|\xi_n(\omega) - \xi(\omega)| < \varepsilon\}$. Тогда

$$1 \geqslant \mathbb{P}(B) \geqslant \mathbb{P}(A) = \mathbb{P}(N(\omega, \varepsilon) < n) = F_{N(\varepsilon,\omega)}(n) \xrightarrow[n\to+\infty]{} 1.$$

по свойству функции распределения. Таким образом, было получено, что $\mathbb{P}(B) \to 1$, т.е. $\xi_n \xrightarrow{\mathrm{p}} \xi$.

п.н. $\not\Leftarrow$ p Положим $\xi_{2^k} = \mathrm{I}\left(\left[0, \frac{1}{2^k}\right]\right), \xi_{2^k+p} = \mathrm{I}\left(\left[\frac{p}{2^k}, \frac{p+1}{2^k}\right]\right), 1 \leqslant p < 2^k$. Тогда $\xi_n \xrightarrow{\mathrm{p}} 0$, т.к. $\mathbb{P}(\xi_n > 0) \leqslant$ длина отрезка в индикаторе $\leqslant \frac{2}{n} \to 0$, но $\xi_n \overset{\text{п.н.}}{\not\to} 0$, т.к. $\forall\omega \quad \exists$ бесконечно много n, таких что $\xi_n(\omega) = 1$.

(r) $\not\Rightarrow$ п.н. См. предыдущий пример.

(r) $\not\Leftarrow$ п.н. $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P}$ - равномерное распределение. Определим для $k \geqslant 1 \quad \xi_k = 2^{k-1}\mathrm{I}\left(\left[0, \frac{1}{2^{k-1}}\right]\right)$. Тогда $\forall k \quad \mathbb{E}\xi_k = 1$, но $\xi = \mathrm{I}(\omega = 0)$.

p $\Rightarrow$ w Пусть f - ограниченная и непрерывная функция, $|f| \leqslant C$. Зафиксируем $\varepsilon > 0$. Т.к. $\mathbb{P}(|\xi| = \infty) = 0$, то $\exists N, \exists \delta$:

1. $\mathbb{P}(|\xi| > N) \leqslant \frac{\varepsilon}{6C}$, т.к. $\mathbb{P}(\xi = \infty) = 0$.
2. $\mathbb{P}(|\xi_n - \xi| > \delta) \leqslant \frac{\varepsilon}{6C}$ из сходимости по вероятности (при достаточно больших n).
3. $\forall x, y |x| < N, |x - y| < \delta |f(x) - f(y)| \leqslant \frac{\varepsilon}{3}$, т.к. f равномерно непрерывна на отрезке $[-N, N]$.

Рассмотрим следующие события:

$$A_1 = \{|\xi_n - \xi| \leqslant \delta\} \cap \{|\xi| < N\}$$
$$A_2 = \{|\xi_n - \xi| \leqslant \delta\} \cap \{|\xi| \geqslant N\}$$
$$A_3 = \{|\xi_n - \xi| > \delta\}$$

Эти события образуют разбиение $\Omega = A_1 \sqcup A_2 \sqcup A_3$.

Оценим $|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi)|$:

$$|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi)| = |\mathbb{E}(f(\xi_n) - f(\xi))| \leqslant \mathbb{E}|f(\xi_n) - f(\xi)| =$$
$$= \mathbb{E}|f(\xi_n) - f(\xi)|(\mathrm{I}_{A_1} + \mathrm{I}_{A_2} + \mathrm{I}_{A_3}) \leq$$
$$\leqslant \frac{\varepsilon}{3}\mathbb{P}(A_1) + 2C(\mathbb{P}(A_2) + \mathbb{P}(A_3)) \leqslant \frac{\varepsilon}{3} + 2C\left(\frac{\varepsilon}{6C} + \frac{\varepsilon}{6C}\right) = \varepsilon,$$

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

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

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.

$$\mathbb{P}\left(|\xi_n - \xi| > \frac{1}{3}\right) = \frac{1}{2} \not\to 0 \text{ при } n \to \infty.$$

p $\not\Leftarrow$ w Пусть $\Omega = \{\omega_1, \omega_2\}, \mathbb{P}(\{\omega_i\}) = \frac{1}{2}$.

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 $\forall n \quad |\xi_n - \xi| = 2 \Rightarrow \xi_n \stackrel{\text{p}}{\nrightarrow} \xi$.

■

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, найденные авторами, либо требуют знания теории меры, либо являются слишком кринжовыми, чтобы включать их в это учебное пособие.

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

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

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

Доказательство.

$$\mathrm{I}(A) \sim \mathbf{Bi}(p),\ p = \mathbb{P}(\mathrm{I}(A) = 1) = \mathbb{P}(A) = \mathbb{E}\mathrm{I}(A)$$

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

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

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

Следствие (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 $\mathbb{P}(\xi \geqslant x) \leqslant \mathbb{P}(g(\xi) \geqslant g(x))$, поскольку функция g не убывает. Оценим последнюю вероятность по неравенству Маркова, которое можно применять в силу неотрицательности g

$$\mathbb{P}(g(\xi) \geqslant g(x)) \leqslant \frac{\mathbb{E}g(\xi)}{g(x)}$$

$\blacksquare$

Следствие (Неравенство Чебышёва)**.** *Если $\mathbb{D}\xi$ существует, то для любого $x > 0$*

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

Доказательство. Для $x > 0$ неравенство $|\xi - \mathbb{E}\xi| \geqslant x \Leftrightarrow (\xi - \mathbb{E}\xi)^2 \geqslant x^2$, поэтому

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geqslant x) = \mathbb{P}\left((\xi - \mathbb{E}\xi)^2 \geqslant x^2\right) \leqslant \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^2}{x^2} = \frac{\mathbb{D}\xi}{x^2}$$

$\blacksquare$

Определение. В неравенстве Чебышёва в качестве x можно брать любое положительное число. Если взять в качестве x величину 3σ, где σ — стандартное отклонение (то есть именно корень из дисперсии), то получится

$$\mathbb{P}(|X - \mathbb{E}X| > 3\sigma) \leqslant \frac{\mathbb{D}X}{9\mathbb{D}X} = \frac{1}{9} \Leftrightarrow \mathbb{P}(|X - \mathbb{E}X| \leqslant 3\sigma) \geqslant 1 - \frac{1}{9} = \frac{8}{9}$$

Это соотношение называется *правилом трёх сигм.*

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

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

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

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mathbb{E}\xi_1 + \ldots + \mathbb{E}\xi_n}{n} \xrightarrow{\text{p}} 0 \text{ при } n \to +\infty$$

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 $\operatorname{cov}(\xi_i, \xi_j)$ по свойству ковариации обратились в нуль при $i \neq j$. ■

2.13 Характеристические функции и их свойства

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

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

где интеграл справа называется *интегралом Фурье-Стильтьеса.*

Для абсолютно непрерывного распределения характеристическая функция имеет вид

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

Для дискретного, соответственно

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

Пример. Характеристическая функция случайной величины $\xi \sim \mathbf{N}(0; 1)$:

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

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

Свойства характеристической функции.

1. *Характеристическая функция существует для любой случайной величины ξ.*
2. $\forall\ \xi,\ \forall\ a, b \in \mathbb{R}$: $\varphi_{a\xi+b}(t) = e^{itb}\varphi_\xi(at)$
3. $|\varphi_\xi(t)| = |\mathbb{E}e^{it\xi}| \leqslant 1,\ \varphi_\xi(0) = 1,\ \overline{\varphi_\xi(t)} = \varphi_\xi(-t) = \varphi_{-\xi}(t)\ \forall t \in \mathbb{R}$

   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:*

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

   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_ξ. Функция распределения восставливаются по характеристической функции с помощью обратных преобразований 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:

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

8. $\xi_n \Rightarrow \xi \Leftrightarrow \varphi_{\xi_n}(t) \to \varphi_\xi(t)$ *(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. $\varphi_{a\xi+b}(t) = \mathbb{E}e^{it(a\xi+b)} = e^{itb}\mathbb{E}e^{ita\xi} = e^{itb}\varphi_\xi(ta)$

3. Неравенство доказано в пункте 1, первое равенство 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. $\varphi_{\xi+\eta}(t) = \mathbb{E}e^{it(\xi+\eta)} = \mathbb{E}e^{it\xi}\mathbb{E}e^{it\eta} = \varphi_\xi(t)\varphi_\eta(t)$

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

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

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

Теперь выберем R настолько большим, чтобы $\mathbb{P}(|X|>R) < \frac{\varepsilon}{4}$. 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.[2] Тогда мы получим, что

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

[2]Достаточно обозначить интеграл $\int\limits_{\mathbb{R}} x^m e^{itx} dF(x)$, $m = \overline{0, k-1}$ за $G(t)$ и расписать его производную по t по определению, воспользовавшись тем, что $\int\limits_{\mathbb{R}} o(\Delta t) dF(x) = o(\Delta t) \int\limits_{\mathbb{R}} dF(x) = o(\Delta t)$.

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

■

2.14 Закон больших чисел в форме Хинчина. Усиленный закон больших чисел в форме Колмогорова

Закон больших чисел в форме Хинчина.

Для любой последовательности $\xi_1, \xi_2, \ldots$ независимых и одинаково распределённых случайных величин с конечным первым моментом $E|\xi_1| < \infty$ 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 $\xi_n \Rightarrow c = const$, то $\xi_n \xrightarrow{p} c$*

Доказательство. Пусть $\xi_n \Rightarrow c$, 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 $\mathbb{P}(|\xi_n - c| < \varepsilon) \to 1$:

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

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

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

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

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:

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

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

$$\varphi_{S_n/n}(t) = \left(1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right)\right)^n \to e^{ita}$$

$\blacksquare$

Усиленный закон больших чисел в форме Колмогорова. *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots,$ — независимые одинаково распределённые случайные величины (сокращённо н.о.р.с.в.). 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, $\frac{1}{n}\sum\limits_{i=1}^{n} \xi_i \xrightarrow{п.н.} a$.

2. *Если существует $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i = a$, то существует $\mathbb{E}\xi_1 = a$.*

2.15 Центральная предельная теорема

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

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

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

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

В силу разложения характеристической функции

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

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 $\mathbf{N}(0,1)$. В силу теорем о непрерывном соответсвии между функциями распределения и характеристическими функциями центральная предельная теорема доказана. ■

2.16 Условное математическое ожидание

Пусть ξ и η — две случайные величины на некотором вероятностном пространстве, причём $E|\xi| < \infty$; $L = L(\eta)$ — множество, в котором собраны все случайные величины, имеющие вид $\zeta = g(\eta)$, где $g(x)$ — произвольная борелевская функция. Скалярным произведением двух случайных величин φ и ζ назовём $(\varphi, \zeta) = \mathbb{E}(\varphi \cdot \zeta)$, если это математическое ожидание существует.

Условное математическое ожидание $\mathbb{E}(\xi|\eta)$ случайной величины ξ относительно η можно представлять себе как результат ортогонального проектирования случайной величины ξ на пространство L.

Результат проектирования — такая случайная величина $\mathbb{E}(\xi|\eta) = \widehat{\xi}$, для которой выполнено основное и единственное свойство ортопроекции: её разность с ξ ортогональна всем элементам L. Ортогональность означает, что для любой $g(\eta) \in L$ обращается в нуль (если вообще существует) скалярное произведение $(\xi - \widehat{\xi}, g(\eta))$, т. е.

$$\mathbb{E}((\xi - \widehat{\xi})g(\eta)) = 0 \text{ или } \mathbb{E}(\xi g(\eta)) = \mathbb{E}(\widehat{\xi} g(\eta))$$

Это свойство называют *тождеством ортопроекции.* Чтобы матожидание существовало всегда, достаточно брать лишь ограниченные функции $g(y)$.

Определение. Пусть $\mathbb{E}|\xi| < \infty$, $L = L(\eta)$ — множество всех борелевских функций от случайной величины η. Условным математическим ожиданием $\mathbb{E}(\xi|\eta)$ называется случайная величина $\widehat{\xi} \in L$, удовлетворяющая тождеству ортопроекции.

Свойства условного математического ожидания.

1. *Все свойства математического ожидания, как, например, линейность, сохраняются, однако борелевские функции от случайной величины η выносятся из-под знака математического ожидания как постоянные.*

2. *Пусть $\mathbb{E}\xi^2 < \infty$. Тогда расстояние от ξ до её ортопроекции $\widehat{\xi} = \mathbb{E}(\xi|\eta)$ является наименьшим из расстояний от ξ до всех «точек» множества L:*
$$\min_{g(\eta)\in L} \mathbb{E}(\xi - g(\eta))^2 = \mathbb{E}(\xi - \widehat{\xi})^2$$

3. *Если $f(\eta) \in L$ такова, что $\mathbb{E}|f(\eta) \cdot \xi| < \infty$, то*
$$\mathbb{E}(f(\eta) \cdot \xi|\eta) \overset{\text{п.н.}}{=} f(\eta) \cdot \mathbb{E}(\xi|\eta)$$

   Доказательство. *Рассмотрим только случай, когда $f(\eta)$ ограничена. Проверим, что $\zeta = f(\eta) \cdot \mathbb{E}(\xi|\eta)$ удовлетворяет тождеству ортопроекции: для любой ограниченной $g(\eta) \in L$*
$$\mathbb{E}(f(\eta)\xi \cdot g(\eta)) = \mathbb{E}(\zeta \cdot g(\eta))$$

   Обозначим $h(\eta) = f(\eta)g(\eta) \in L$. Эта функция ограничена, поэтому
$$\mathbb{E}(\xi f(\eta) \cdot g(\eta)) = \mathbb{E}(\xi h(\eta)) = \mathbb{E}(\widehat{\xi} h(\eta)) = \mathbb{E}(\zeta \cdot g(\eta))$$

$\blacksquare$

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

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

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

Вычисление условного матожидания

Возьмём функцию $h(y)=\mathbb{E}(\xi\,|\,\eta=y)$, для которой $\mathbb{E}(\xi|\eta)=h(\eta)$ и рассмотрим, что такое условное математическое ожидание относительно события $\{\eta=y\}$ для двух случаев: когда обе случайные величины имеют дискретное распределение и когда их совместное распределение абсолютно непрерывно.

1. Пусть ξ принимает значения $a_1,a_2,\ldots$, а η — значения $b_1,b_2,\ldots$ Тогда $h(\eta)$ может принимать только значения $h\left(b_1\right),h\left(b_2\right),\ldots$, где

   $$h(y)=\sum_i a_i\mathbb{P}\left(\xi=a_i|\eta=y\right)$$

   Иначе говоря, при каждом фиксированном y начение $h(y)$ определяется как математическое ожидание дискретного распределения со значениями a_i и вероятностями $\mathbb{P}\left(\xi=a_i|\eta=y\right)$. Такое распределение называется *условным распределением случайной величины ξ при условии $\eta=y$.*

2. Во втором случае пусть $f_{\xi,\eta}(x,y)$ — плотность совместного распределения, $f_\eta(y)$ — плотность распределения величины η.Тогда положим

   $$h(y)=\int\limits_{\mathbb{R}}x\frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}dx$$

   При фиксированном y число $h(y)$ есть математическое ожидание абсолютно непрерывного распределения с плотностью распределения

   $$f(x|y)=\frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}=\frac{f_{\xi,\eta}(x,y)}{\int\limits_{\mathbb{R}}f_{\xi,\eta}(x,y)dx}$$

   Такое распределение называется *условным распределением величины ξ при условии $\eta=y$*, а функция $f(x|y)$ — *условной плотностью.*

Убедимся формально (скажем, в абсолютно непрерывном случае), что определённая выше $h(\eta)$ удовлетворяет тождеству ортопроекции и, следовательно, является условным матожиданием $\mathbb{E}(\xi|\eta)$. Для любой $g(\eta) \in L$ (такой, что соответствующее математическое ожидание существует) левая часть тождества ортопроекции равна

$$\mathbb{E}(\xi \cdot g(\eta)) = \iint_{\mathbb{R}^2} xg(y)f_{\xi,\eta}(x,y)dxdy$$

Правая часть равна

$$\mathbb{E}(h(\eta)g(\eta)) = \int\limits_{\mathbb{R}} h(y)g(y)f_\eta(y)dy = \iint_{\mathbb{R}\mathbb{R}} x\frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}dx \cdot g(y)f_\eta(y)dy$$

Сокращая $f_\eta(y)$, получаем равенство левой и правой частей.

Литература

[1] Н.И.Чернова. Теория вероятностей, 2007

[2] Н.И.Чернова. Математическая статистика, 2014

[3] В.Г. Ушаков. Теория вероятностей и математическая статистика, 2014

[4] В.В. Ульянов. Конспект лекций по теории вероятностей и математической статистике

[5] А. Марков. Теория вероятностей, 2017