Содержание

1 Основы теории вероятности.

Вероятностное пространство.

Определение 1. Пусть Ω — множество, тогда $\mathfrak{A} \in 2^{\Omega}$ называется **алгеброй**, если

1. $\Omega \in A$.
2. $\forall A \in \mathfrak{A}\ \overline{A} \in \mathfrak{A}$. Здесь и далее $\overline{A} = \Omega \setminus A$.
3. $\forall A, B \in \mathfrak{A}\ A \cup B \in \mathfrak{A}$.

При этом Ω называется **множеством элементов событий**, $\mathfrak{A}$ — **набор событий**, $A \in \mathfrak{A}$ — **событие**, $A \cup B = A + B$ — **сумма событий**, $\overline{A}$ — **противоположное событие**, $A \cap B = AB$ — **произведение событий**.

Определение 2. Алгебра является **сигма-алгеброй**, если она замкнута относительно объединения счётного количества своих элементов.

Определение 3. Пусть $\mathfrak{A}$ — сигма-алгебра на Ω. Пусть $P\colon \mathfrak{A} \to [0; +\infty)$ и

1. $P(\Omega) = 1$.
2. Если $\{A_i\}_{i=1}^{\infty} \subset \mathfrak{A}$ и $\forall A_i A_j = \varnothing$ то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Тогда $(\Omega; \mathfrak{A}; P)$ называется **вероятностным пространством**.

Определение 4. Пара событий называется **несовместной**, если их пересечение пусто. Набор событий **несовместен**, если они попарно несовместны.

Определение 5. Пусть $A \subset 2^{\Omega}$ — алгебра. Тогда минимальная по включению сигма-алгебра $\sigma(A) \supset A$ называется **минимальной сигма-алгеброй**.

Утверждение. *Таковая существует.*

Доказательство. Хотя бы одна такая существует (2^{Ω}), причём если пересечь сколько угодно сигма-алгебр, то получится искомая сигма-алгебра. □

Определение 6. Пусть $\mathfrak{A}$ — алгебра на Ω, $P\colon \mathfrak{A} \to [0; +\infty)$ и

- $P(\Omega) = 1$.
- Если $\{A_i\}_{i=1}^{\infty} \subset \mathfrak{A}$ и $\bigsqcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$, то

$$P\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Тогда $(\Omega; \mathfrak{A}; P)$ — **вероятностное пространство в широком смысле**.

Теорема 1 (О продолжении меры)**.** *Пусть $(\Omega; \mathfrak{A}; P)$ — вероятностное пространство в широком смысле. Тогда существует единственная функция вероятности $Q\colon \sigma(\mathfrak{A}) \to [0 : +\infty)$, такое что $Q\Big|_{\mathfrak{A}} \equiv P$.*
Без доказательства.

Замечание. Эта теорема позволяет нам сказать, например, что мы хотим задать вероятность на отрезках.

Определение 7. Борелевская сигма-алгебра — минимальная σ-алгебра, которая содержит все открытые множества.

Пример. Дискретное вероятностное пространство: $\Omega = \{\omega_i\}_{i=1}^{N}$, $A = 2^{\Omega}$, $P(\{\omega_i\}) = p_i$, $\sum p_i = 1$. Тогда $P(A)$ — сумма вероятностей элементов A.

Пример. Геометрическая вероятность: $\Omega \subset \mathbb{R}^n$, измеримо по Лебегу, $\mu A < +\infty$, $\mathfrak{A}$ состоит из измеримых по Лебегу множеств, $P(A) = \frac{\mu A}{\mu \Omega}$. Обычно при этом $\mathbb{R}^n$ не более чем трёхмерно.

Свойства вероятности.

Свойство 7.1.

$$\forall A, B \in \mathfrak{A} \ A \subset B \Rightarrow P(A) \leqslant P(B)$$

Доказательство. Понятно, что $B \setminus A \in \mathfrak{A}$. Тогда

$$B = A \sqcup (B \setminus A) \Rightarrow P(B) = P(A) + P(B \setminus A) \geqslant P(A)$$

□

Следствие 0.1.

$$\forall A \in \mathfrak{A} \; P(A) \leqslant 1$$

Свойство 7.2.

$$P(A) = 1 - P(\overline{A})$$

Свойство 7.3.

$$P(A + B) = P(A) + P(B) - P(AB)$$

Доказательство.

$$B = (B \setminus AB) \sqcup AB \Rightarrow P(B) = P(B \setminus AB) + P(AB)$$

Тогда

$$P(A + B) = P(A) + P(B \setminus AB) = P(A) + P(B) - P(AB)$$

□

Утверждение (Формула включений-исключений)**.**

$$P(A_1 + \cdots + A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,j=1\\ i<j}}^{n} P(A_iA_j) + \sum_{\substack{i,j,k=1\\ i<j<k}}^{n} P(A_iA_jA_k) - \cdots + (-1)^n P(A_1 \cdots A_n)$$

Доказательство. Мне лень это писать, докажите сами по индукции. □

Утверждение.

$$P\left(\bigcup_i A_i\right) \leqslant \sum_i P(A_i)$$

Доказательство. Пусть $B_1 = A_1$, $B_2 = A_2\overline{A_1}$, $B_3 = A_3\overline{A_1 \cup A_2}$ и так далее. Тогда

$$\bigcup_i A_i = \bigsqcup_i B_i$$

При этом $B_i \subset A_i$, а значит

$$\sum_i P(A_i) \geqslant \sum_i P(B_i)$$

□

Теорема 2. *Пусть $(\Omega; \mathfrak{A}; P)$ — вероятностное пространство. Тогда следующие утверждения равносильны:*

1. *P счётно-аддитивна.*
2. *P конечно-аддитивна и $\forall \{B_i\}_{i=1}^{\infty} : B_{i+1} \subset B_i$, $B = \bigcap\limits_{i=1}^{\infty} B_i$ $\lim\limits_{n\to\infty} P(B_i) = P(B)$ (непрерывность сверху).*
3. *P конечно-аддитивна и $\forall \{C_i\}_{i=1}^{\infty} : C_{i+1} \supset B_i$, $C = \bigcup\limits_{i=1}^{\infty} C_i$ $\lim\limits_{n\to\infty} P(C_i) = P(C)$ (непрерывность снизу).*

Доказательство. Равносильность двух непрерывностей тривиально из формул де Моргана.
Докажем, что из 1 следует 2. Конечная аддитивность есть, докажем непрерывность сверху. Пусть $A_1 = B_1\overline{B_2}$, $A_2 = B_2\overline{B_3}$ и так далее. Очевидно, A_i несовместны. Также очевидно, что A_i несовместны с B. Также заметим, что

$$B_n = B \sqcup \bigsqcup_{i=n+1}^{\infty} A_i$$

Отсюда $P(B_n) = P(B) + \sum\limits_{i=n+1}^{\infty} P(A_i)$, а справа остаток (очевидно, сходящегося) ряда, который стремится к нулю при $n \to \infty$.
Теперь из 2 докажем 1. Рассмотрим $\{A_i\}_{i=1}^{\infty}$ несовместные. Очевидно,

$$\sum_{i=1}^{\infty} P(A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i)$$

А ещё мы знаем, что

$$\bigsqcup_{i=1}^{\infty} A_i = \bigsqcup_{i=1}^{n} A_i \sqcup \bigsqcup_{i=n+1}^{\infty} A_i$$

То есть

$$\lim_{n\to\infty} P\left(\bigsqcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty}\left(P\left(\bigsqcup_{i=1}^{\infty} A_i\right) - P\left(\bigsqcup_{i=n+1}^{\infty} A_i\right)\right) = P\left(\bigsqcup_{i=1}^{\infty} A_i\right) - \lim_{n\to\infty} P\left(\bigsqcup_{i=n+1}^{\infty} A_i\right)$$

Второе слагаемое — ноль но непрерывности меры, а отсюда счётная аддитивность. □

Условная вероятность.

Замечание. Пусть $|\Omega| = n$, $|A| = k$, $|B| = m$, $|AB| = l$. Если мы знаем, что B произошло, как узнать вероятность того, что произошло A? Ну, это

$$\frac{l}{m} = \frac{l/n}{m/n} = \frac{P(AB)}{P(B)}$$

Определение 8. Пусть $(\Omega; \mathfrak{A}; P)$ — вероятностное пространство, $B \in \mathfrak{A}$, $P(B) > 0$. Тогда **условной вероятностью** A при условии B называется

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

Также обозначается $P_B(A)$.

Свойство 8.1. *Несложно проверить, что условная вероятность является вероятностью.*

Утверждение (Произведение вероятностей)**.** *Несложно по определению проверить*

$$P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$$

Теорема 3 (Формула полной вероятности)**.** *Пусть $A \in \mathfrak{A}$, $B_i \in \mathfrak{A}$ несовместны, $A \subset \bigsqcup\limits_{i=1}^{n} B_i$ (обычно объединение равно Ω), и $\forall i \in [0:n]$ $P(B_i) > 0$. Тогда*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Доказательство.

$$P(A) = P\left(A \cap \bigsqcup_{i=1}^{n} B_i\right) = P\left(\bigsqcup_{i=1}^{n} A \cap B_i\right)$$

Всё. □

Теорема 4 (Формула Байеса)**.** *Пусть $A, B \in \mathfrak{A}$, $P(A) > 0$, $P(B) > 0$. Тогда*

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

Доказательство. Очевидно из определения. □

Определение 9. События $A, B \in \mathfrak{A}$ называются **независимыми**, если

$$P(AB) = P(A)P(B)$$

Определение 10. Говорят, что $A_1, A_2, \dots, A_n \in \mathfrak{A}$ **независимы в совокупности**, если $P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)$

Свойство 10.1. *Несложно проверить, что независимость событий A, B равносильна $P(A|B) = P(A)$.*

Свойство 10.2. *Независимые в совокупности события попарно независимы. Обратное неверно.*

Определение 11. Пусть у нас есть два вероятностных пространства: $(\Omega_1, \mathfrak{A}_1; P_1)$ и $(\Omega_2, \mathfrak{A}_2; P_2)$. Рассмотрим вот такое вероятностное пространство: $(\Omega, \mathfrak{A}; P)$, где $\Omega = \Omega_1 \times \Omega_2$, $\mathfrak{A}$ — минимальная σ-алгебра, включающая в себя $\mathfrak{A}_1 \times \mathfrak{A}_2$,

$$P((A_1; A_2)) = P_1(A_1)P_2(A_2)$$

Тогда $(\Omega_1, \mathfrak{A}_1; P_1)$ и $(\Omega_2, \mathfrak{A}_2; P_2)$ — независимые испытания.

Схема Бернулли.

Пример. Схема Бернулли: $\Omega_1 = \{0; 1\}$, $\mathfrak{A}_1 = 2^{\Omega_1}$, $P_1(1) = p$, $P_1(0) = 1 - p = q$. Хочется рассмотреть эту штуку в степени n (то есть n одинаковых независимых испытаний). Тогда что у нас получается для $\omega \in \Omega = \Omega_1^n$?

$$P(\omega) = \sum_{i=1}^{n} P_i(\omega_i) = p^{\sum \omega_i} q^{n - \sum \omega_i}$$

Посчитаем тут такую вероятность: пусть S_n — количество успехов в n испытаниях? Посчитаем вероятность того, что $S_n = k$? Очевидно, оно равно $\binom{n}{k} p^k q^{n-k}$.

Утверждение. *Пусть k^* — наиболее вероятное число успехов в Бернуллиевских испытаниях. Тогда*

$$k^* = \begin{cases} p(n-1) \text{ или } p(n-1)+1 & p(n-1) \in \mathbb{N} \\ \lceil p(n-1) - 1 \rceil & p(n-1) \notin \mathbb{N} \end{cases}$$

Доказательство. Давайте рассмотрим вот такое частное:

$$\frac{P(S_n = k+1)}{P(S_n = k)}$$

Чему оно равно?

$$\frac{P(S_n = k+1)}{P(S_n = k)} = \frac{\binom{n}{k+1} p^{k+1} q^{n-k-1}}{\binom{n}{k} p^k q^{n-k}} = \frac{p}{q} \cdot \frac{n-k}{k+1}$$

Нам хочется оценить, больше это чем 1 или меньше (это позволит нам найти K^*). Ну,

$$\frac{p}{q} \cdot \frac{n-k}{k+1} > 1 \Leftrightarrow p(n-k) > q(k+1) \Leftrightarrow pn - pk > k = pk + 1 - p \Leftrightarrow pn > k + 1 - p \Leftrightarrow pn + p - 1 > k$$

То есть возрастание достигается при $k < p(n-1) - 1$, а иначе убывание. Тогда где экстремум? Рассмотрим $k = p(n-1) - 1$. Если это целое число, то там $P(S_n = k+1) = P(S_n = k)$, и это самое k даёт значение больше остальных. То есть $k^* = p(n-1) - 1$ или $k^* = p(n-1)$.
А что если оно не целое? То надо куда-то округлить. А именно вверх, потому что тогда оно больше, чем следующее, а предыдущее меньше его. □

Пример. Пусть $n = 10000$, $p = \frac{1}{10000}$. Давайте посчитаем $P(S_n > 3)$. Ну, это

$$1 - P(S_n \leqslant 3) = 1 - q^{10000} - 10000pq^{10000-1} - \binom{10000}{2} p^2 q^{10000-2} - \binom{10000}{3} p^3 q^{10000-3}$$

Фиг мы такое посчитаем.

Пример. Или если взять $p = q = 0.5$, то при $n = 5 \cdot 10^3$ мы не сможем нормально посчитать $P(S_n = 2349)$.

Замечание. Ну и как такое считать?

Теорема 5 (Теорема Пуассона)**.** *Пусть у нас есть несколько схем Бернулли. В первой одно испытание и вероятность успеха p_1, во второй — 2 и вероятность успеха p_2, в n-ной n испытаний и вероятность p_n. Пусть $np_n \underset{n\to\infty}{\longrightarrow} \lambda > 0$. Тогда*

$$P(S_n = k) \underset{n\to\infty}{\longrightarrow} e^{-\lambda} \frac{\lambda^k}{k!}$$

Доказательство. Известно,

$$P(S_n = k) = \frac{1}{k!} n(n-1)\cdots(n-k+1) p_n^k (1-p_n)^{n-k}$$

Известно, что

$$np_n = \lambda + o(1) \Rightarrow p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$$

Тогда

$$P(S_n = k) = \frac{1}{k!} \cancelto{1}{n(n-1)\cdots(n-k+1)\frac{1}{n^k}} \cancelto{\lambda^k}{(\lambda + o(1))^k} \frac{\cancelto{e^{-\lambda}}{\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^n}}{\cancelto{1}{\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k}} \underset{n\to\infty}{\longrightarrow} e^{-\lambda} \frac{\lambda^k}{k!}$$

□

Лемма 1. *Пусть $p \in (0;1)$, $H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}$. Пусть $p^* = \frac{k}{n}$. Пусть $k \to +\infty$, $n-k \to +\infty$. Тогда*

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} \exp(-nH(p^*))$$

Доказательство. Мы знаем формулу Стирлинга

$$n! \sim \sqrt{2\pi n} n^n e^{-n}$$

Тогда

$$P(S_n = k) = \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi \underbrace{k}_{np^*}} k^k e^{-k} \sqrt{2\pi \underbrace{(n-k)}_{n(1-p^*)}} (n-k)^{n-k} e^{-n+k}} p^k (1-p)^{n-k} =$$

$$= \frac{n^n p^k (1-p)^{n-k}}{\sqrt{2\pi n p^*(1-p^*)} k^k (n-k)^{n-k}} = \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} \exp \underbrace{\ln \frac{n^n p^k (1-p)^{n-k}}{k^k (n-k)^{n-k}}}_{L}$$

При этом

$$L = \ln \frac{n^n p^k (1-p)^{n-k}}{k^k (n-k)^{n-k}} = \ln \frac{n^n p^k (1-p)^n (n-k)^k}{k^k (n-k)^n (1-p)^k} =$$

$$= \ln \left(\underbrace{\frac{n^n}{(n-k)^n}}_{(1-p^*)^{-n}} (1-p)^n \right) + \ln \frac{p^k (n-k)^k}{(np^*)^k (1-p)^k} =$$

$$= n \ln \frac{1-p}{1-p^*} + k \ln \frac{p}{p^*} + k \ln \frac{(n-k)}{n(1-p)} = n \ln \frac{1-p}{1-p^*} + k \ln \frac{p}{p^*} + k \ln \frac{1-p^*}{1-p} =$$

$$= -(n-k) \ln \frac{1-p^*}{1-p} - k \ln \frac{p^*}{p} = -n \underbrace{\left(p^* \ln \frac{p^*}{p} + (1-p^*) \ln \frac{1-p^*}{1-p} \right)}_{H(p^*)}$$

Это ли не то, что нам надо? □

Лемма 2.

$$H(x) = \frac{(x-p)^2}{2p(1-p)} + O((x-p)^3)$$

Доказательство.

$$H'(x) = \ln \frac{x}{p} + x \cdot \frac{p}{x} \cdot \frac{1}{p} - \ln \frac{1-x}{1-p} - 1 = \ln \frac{x}{p} - \ln \frac{1-x}{1-p}$$

$$H''(x) = \frac{1}{x} + \frac{1}{1-x}$$

Тогда $H'(p) = 0$, $H''(p) = \frac{1}{p(1-p)}$. По Тейлору получаем искомое. □

Теорема 6 (Локальная теорема Муавра — Лапласа)**.** *Пусть* $p \in (0;1)$, $H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}$*.*
Пусть $p^* = \frac{k}{n}$*. Пусть* $k \to +\infty$, $n-k \to +\infty$*.*
Пусть $k = np = p(n^{2/3})$*. Тогда*

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)} \right)$$

Доказательство. Известно

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp(-nH(p^*))$$

Отсюда

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp(-n\frac{(p^*-p)^2}{2p(1-p)} + n \cdot O((p^*-p)^3))$$

Заметим, что $\frac{k}{n} - p = o(n^{-1/3})$. Тогда

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-n\frac{(p-k/n)^2}{2p(1-p)} + O(n(k/n-p)^3) \right) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-n\frac{(np-k)^2}{2p(1-p)n^2} + o(1) \right)$$

Что и требовалось доказать. □

Теорема 7 (Интегральная теорема Муавра — Лапласа)**.** *Пусть*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} \, \mathrm{d}t$$

Далее мы будем называть эту функцию функцией стандартного нормального распределения. Тогда

$$\sup_{-\infty<x_1<x_2<+\infty} \left| P\left(x_1 \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant x_2\right) (\Phi(x_2) - \Phi(x_1)) \right| \underset{n\to\infty}{\longrightarrow} 0$$

Иными словами

$$P\left(x_1 \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant x_2\right) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{x_1}^{x_2} e^{-\frac{t^2}{2}}\ \mathrm{d}t$$

Пока без доказательства.

Замечание. Оценка теоремы Пуассона.
Обычно в задачах np_n не стремится, а просто равно $\lambda > 0$. Тогда

$$\sum_{k=0}^{\infty} \left| P(S_n = k) - e^{-\lambda}\frac{\lambda^k}{k!} \leqslant \frac{\lambda}{n} \right| \leqslant \frac{2\lambda}{n}\min\{2;\lambda\}$$

Оценка локальной теоремы Лапласа. Если $|p^* - p| \leqslant \frac{1}{2}mn\min\{p;q\}$, то

$$P(S_n = k) = \frac{1}{\sqrt{2\pi np(1-p)}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)(1+\varepsilon(k;n))$$

Где

$$\varepsilon(k;n) = \exp\left(\theta\frac{|k-np|^3}{3n^2p^2q^2} + \frac{1}{npq}\left(\frac{1}{6} + |k-np|\right)\right) \qquad |\theta| < 1$$

Оценка интегральной теоремы Лапласа.

$$\sum_x \left| P\left(\frac{S_n - np}{\sqrt{npq}} \leqslant x\right) - \Phi(x) \right| \leqslant \frac{p^2+q^2}{\sqrt{npq}}$$

Пример. Пусть у нас есть два узла связи на 2000 пользователей в каждом. И у нас есть канал связи, который пропускает N. Хочется минимизировать N, но так, чтобы вероятность перегрузки была меньше $\frac{1}{100}$. Будем предполагать, что люди пользуются данным каналом связи в течение двух минут из одного часа, то есть каждый пользователь может пользоваться каналом в данный момент с вероятностью $p = \frac{1}{30}$.
Ну так и что мы хотим по сути? Мы хотим $P(S_{2000} > N) < \frac{1}{100}$, что равносильно $P(S_{2000} \leqslant N) \geqslant \frac{99}{100}$.
Используем Пуассона: $np \approx 6.67$.

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

Это мы хрен посчитаем, но, короче, получится $N = 87$.
А если применить интегральную теорему Муавра — Лапласа, то получим мы

$$N = \left\lceil q_{\frac{99}{100}}\sqrt{npq} + np \right\rceil = 86$$

Где $q_{\frac{99}{100}}$ — такое число, что $\Phi(q_{\frac{99}{100}}) = \frac{99}{100}$.

Определение 12. Если $\alpha \in (0;1)$ и $\Phi(q_\alpha) = \alpha$, то q_α называется **квантилем порядка** α.

2 Случайные величины.

2.1 Одномерные случайные величины.

Распределение случайных величин, функция распределения случайных величин.

Определение 1. Борелевская сигма-алгебра — это минимальная сигма-алгебра, содержащая все открытые множества.

Определение 2. Пусть $\Omega; \mathfrak{A}$ — множество с сигма-алгеброй. Тогда такое $\xi\colon \Omega \to \mathbb{R}$, что $\forall B \in \mathfrak{B}\ \xi^{-1}(B) \in \mathfrak{A}$, называется **случайной величиной**.

Определение 3. Пусть $(\Omega; \mathfrak{A}; P)$ — вероятностное пространство, ξ — случайная величина. Тогда распределение ξ — функция

$$P_\xi\colon \substack{\mathfrak{B}\to\mathbb{R}\\ B\mapsto P(\{\omega\mid\xi(\omega)\in B\})}$$

Замечание. $P(\{\omega \mid \xi(\omega) \in B\})$ обозначается $P(\xi \in B)$.

Свойство 3.1. *P_ξ — вероятность на $(\mathbb{R}; \mathfrak{B})$.*

Определение 4. Пусть ξ — случайная величина. Тогда

$$F_\xi(t) = P(\xi \leqslant t)$$

называется **функцией распределения** ξ

Свойство 4.1. *Очевидно, функция распределения нестрого возрастает.*

Свойство 4.2. *Не менее очевидно, $F_\xi(+\infty) = 1$, $F_\xi(-\infty) = 0$;*

Свойство 4.3. *Функция распределения непрерывна справа.*

Доказательство. Возьмём $F(t + \varepsilon_n) - F(t)$. Она равна $P(t < \xi \leqslant t + \varepsilon_n)$. При $\varepsilon_n \to 0$, получим, что аргумент P стремится к $\varnothing$, а значит $P(t < \xi \leqslant t + \varepsilon_n)$ стремится к нулю. □

Лемма 1. *Пусть $P\colon \mathfrak{B} \to \mathbb{R}$ — некоторая функция. Тогда*

$$\forall \{B_n\}_{n=1}^\infty \subset \mathfrak{B} : B_{n+1} \subset B_n\ P(B_n) \underset{n\to\infty}{\to} P\left(\bigcap_{n=1}^\infty B_n\right) \Leftrightarrow$$

$$\Leftrightarrow \forall \{C_n\}_{n=1}^\infty \subset \mathfrak{B} : C_{n+1} \subset C_n, \bigcap_{n=1}^\infty C_n = \varnothing\ P(C_n) \underset{n\to\infty}{\to} 0$$

Доказательство. Следствие слева направо очевидно. Наоборот. Пусть

$$\bigcap_{n=1}^\infty B_n = B$$

Возьмём $C_n = B_n\overline{B}$. Тогда, очевидно, C_n подходят под условие справа, а значит $P(C_n) \longrightarrow 0$.
Также несложно заметить, что $P(B_n) = P(C_n) + P(B)$, а отсюда получим $P(B_n) \longrightarrow P(B)$. □

Теорема 1. *Пусть F — монотонно возрастающая непрерывная слева функция, равная нулю в $-\infty$ и единице в $+\infty$. Тогда существует вероятностное пространство и случайная величина в нём, что F — её функция распределения.*

Доказательство. Пусть $\Omega = \mathbb{R}$, $\mathfrak{A}$ — алгебра, состоящая из множеств вида $\bigsqcup\limits_{k=1}^n (a_k; b_k]$ или $(-\infty; b)$ или $(a; \infty)$ или $\mathbb{R}$.

$$P\left(\bigsqcup_{k=1}^n (a_k; b_k]\right) = \sum_{k=1}^n F(b_k) - F(a_k)$$

$$P((-\infty; b)) = F(b) \qquad P((a; +\infty)) = 1 - F(a) \qquad P(\mathbb{R}) = 1$$

Получим вероятностное пространство в широком смысле (разве что непрерывность сверху надо проверить). Ну, проверим её, используя лемму. Пусть, не умаляя общности, $A_n = (a_{n,1}; a_{n,2}]$, $A_{n+1} \subset A_n$

и пересечение всех пусто.
Из непрерывности F справа следует, что существует $B_n = (b_{n,1}; b_{n,2}]$, где $\operatorname{Cl} B_n \subset A_n$ и $P(A_n) - P(B_n) \leqslant \varepsilon 2^{-n}$. Тогда пересечение всех B_n также пусто.
Предположим, что существует M, такое что $\forall n\ A_n \in [-M; M]$. $[-M; M]$ — компакт, следовательно. Заметим, что

$$[-M; M] = \bigcup_{k=1}^{\infty} [-M; M] \setminus \operatorname{Cl} B_n$$

Справа — открытое покрытие компакта, значит из него можно вытащить конечное подпокрытие, то есть пересечение какого-то конечного числа B_n пусто. Пусть это пересечение от 1 до n_0. Тогда

$$P(A_{n_0}) = P(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k) + P\left(\bigcap_{k=1}^{n_0} B_k\right)$$

Отсюда $P\left(\bigcap\limits_{k=1}^{n_0} B_k\right) = 0$. А

$$P(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k) = P(\bigcap_{k=1}^{n_0} A_{n_0} \setminus B_k) \leqslant P(\bigcap_{k=1}^{n_0} A_k \setminus B_k) \leqslant \sum_{k=1} P(A_n) - P(B_k) \leqslant \varepsilon \sum k = 1^{n_0} 2^{-n} < \varepsilon$$

Если же мы не находимся в промежутке $[-M; M]$, то можно указать такие M_1 и M_2, что $P((-\infty; M_1) \cup (M_2; +\infty)) < \frac{\varepsilon}{2}$. Тогда

$$P(A_n) = P(A_n[M_1; M_2]) + P(A_n\overline{[M_1; M_2]})$$

Левую часть суммы мы разобрали, а правая мала т.к. M_2, что $P((-\infty; M_1) \cup (M_2; +\infty)) < \frac{\varepsilon}{2}$.
Осталось предъявить случайную величину $\xi(\omega) = \omega$. □

Типы распределений.

Дискретные случайные величины и распределения.

Определение 5. Случайная величина ξ называется **дискретной**, если существует такое не более чем счётное множество E, что $P_\xi(E) = 1$.

Пример. Вырожденное: $P(\xi = c) = 1$. Обозначают $I(c)$ или I_c.

Пример. Распределение Бернулли: $P(\xi = 0) = p$, $P(\xi = 1) = q = 1 - p$. Обозначение: $\operatorname{Bern}(p)$.

Пример. Биномиальное распределение: $P(\xi = k) = \binom{n}{k} p^k q^{n-k}$. Обозначение: $\operatorname{Bin}(n; p)$,

Пример. Отрицательное биномиальное распределение: $\xi = (\min n : S_n = r) - r$, где $r \in \mathbb{n}$. То есть

$$P(\xi = k) = \binom{k + r - 1}{r - 1} p^r q^k$$

Обозначение: $\operatorname{NB}(r; p)$. Также это обобщается на произвольное r при помощи гамма-функции.
В случае $r = 1$ распределение называется геометрическим. Геометрическое распределение — количество неудач до первого успеха. Обозначается $\operatorname{Geom}(p)$.

Пример. Распределение Пуассона:

$$P(\xi = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad k \in \mathbb{Z}_+$$

Обозначение: $\operatorname{Pois}(\lambda)$.

Определение 6. Носителем случайной величины ξ называется минимально по включению замкнутое множество E, удовлетворяющее условию $P(\xi \in E) = 1$.

Абсолютно непрерывные случайные величины и распределения.

Определение 7. Величина ξ (или её случайно распределение) называется **абсолютно непрерывной**, если существует $p \in L(\mathbb{R} \to \mathbb{R})$ с неотрицательными значениями такая что $P(\xi \in B) = \int_B P(x)\ \mathrm{d}x$. В таком случае p называется **плотностью** ξ.

Свойство 7.1. *В таком случае*

$$F(t) = \int\limits_{-\infty}^{t} p(x)\ \mathrm{d}x$$

То есть плотность почти всюду равна производной функции распределения.

Свойство 7.2. $P(\xi = c) = 0$.

Свойство 7.3. *Функция распределения абсолютно непрерывной случайной величины непрерывна на* $\mathbb{R}$.

Свойство 7.4.

$$P(x_0 \leqslant \xi \leqslant x_0 + h) = P(x_0 < \xi < x_0 + h) = F(x_0 + h) - F(x_0) = p(x_0)h + o(h)$$

Свойство 7.5. *Пусть* $E = \operatorname{supp} p$. *Тогда* E *является носителем по нашему прошлому определению.*

Пример.

$$p(x) = \frac{1}{b-a}\chi_{[a;b]}$$

Тогда

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a;b) \\ 1 & x \geqslant b \end{cases}$$

Обозначение: $U[a;b]$.

Утверждение. *Пусть* $\xi = U[a;b]$, $c > 0$. *Тогда* $\eta = c\xi + d = U[ac + d; bc + d]$

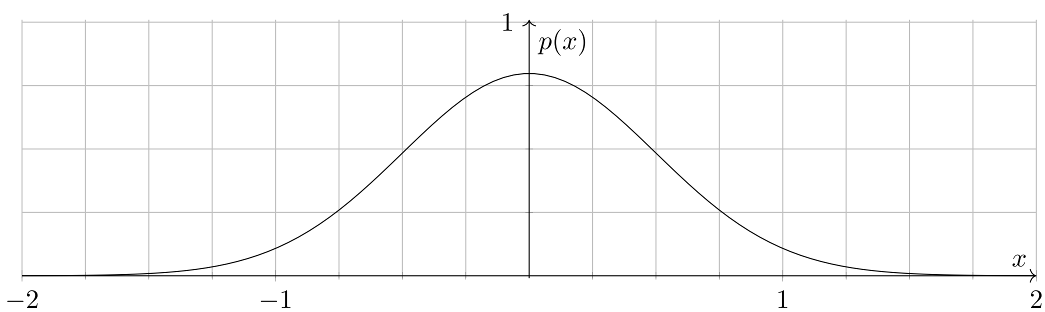
Доказательство.

$$P(\eta \leqslant t) = P(c\xi + d \leqslant t) = P\left(\xi \leqslant \frac{t-d}{c}\right)$$

Несложно проверить, что это именно $U[ac + d; bc + d]$. □

Пример. Нормальное (гауссовское) распределение: $\mathrm{N}(\mu; \sigma^2)$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



$\mathrm{N}(0;1)$ — стандартное нормальное распределение. Ещё оно обозначается $\Phi(x)$.

Утверждение. *Пусть* $\xi = \mathrm{N}(\mu;\sigma^2)$, $\eta = a\xi + b$. *Тогда* $\eta = \mathrm{N}(a\mu_b; a^2\sigma^2)$.

Доказательство. Пусть $a > 0$. Обозначим $y = ax + b$. Тогда

$$P(\eta \leqslant t) = P(a\xi + b \leqslant t) = P(\xi \leqslant \frac{t-b}{a}) = F_\xi(\frac{t-b}{a}) = \frac{1}{2\pi\sigma^2}\int\limits_{-\infty}^{\frac{t-b}{a}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,\mathrm{d}x =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2a^2}}\int\limits_{-\infty}^{t} \exp\left(-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2a^2}}\int\limits_{-\infty}^{t} \exp\left(-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right)$$

Что и требовалось доказать. При $a < 0$ аналогично. ☐

Следствие 1.1. *Если* $\xi = \mathrm{N}(0;1)$, *то* $\sigma\xi + \mu = \mathrm{N}(\mu;\sigma^2)$.

Пример. Распределение Коши: $\mathrm{Cauchy}(x_0;\gamma)$.

$$p(x) = \frac{1}{\pi\gamma}\cdot\frac{1}{1+\left(\frac{x-x_0}{\gamma}\right)^2}$$

Тогда

$$F(t) = \frac{1}{\pi\gamma}\int\limits_{-\infty}^{t}\frac{\mathrm{d}x}{1+\left(\frac{x-x_0}{\gamma}\right)^2} = \frac{1}{\pi}\tan^{\text{-1}}\left(\frac{t-x_0}{\gamma}\right) + \frac{1}{2}$$

Пример. Экспоненциальное распределение: $\mathrm{Exp}(\lambda)$.

$$p(x) = \lambda e^{-\lambda x}\chi_{\mathbb{R}_+}$$

Тогда

$$F(t) = 1 - e^{-\lambda t}\chi_{\mathbb{R}_+}$$

Пример. Г-распределение: $\Gamma(k,\lambda)$. Сначала для $k \in \mathbb{N}$, $\lambda > 0$.

$$p(x) = \frac{\lambda^k x^{k-1}}{(k-1)!}e^{-\lambda x} \qquad x \geqslant 0$$

Для $k = \alpha > 0$ заменим $(k-1)!$ на $\Gamma(\alpha)$.

Утверждение. *Пусть* ξ *— абсолютно непрерывная случайная величина с плотностью* p_ξ. *Пусть* $g \in C^{(1)}(\mathbb{R}\to\mathbb{R})$ *— строго монотонна. Пусть* $\eta = g(\xi)$. *Тогда*

$$p_\eta(y) = p_\xi(g^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right| = p_\xi(g^{-1}(y))\left|\frac{1}{g'(g^{-1}(y))}\right|$$

Доказательство. Во-первых, у условиях теоремы g^{-1} существует. Также второе равенство следует из первого в силу теоремы об обратном отображении.
Не умаляя общности, g строго возрастает. Тогда

$$F_\eta(y) = P(g(\xi) \leqslant y) \overset{g\uparrow}{=} P(\xi \leqslant g^{-1}(y)) = \int\limits_{-\infty}^{g^{-1}(y)} p_\xi(x)\,\mathrm{d}x$$

Продифференцировав это равенство по y получим искомое равенство. ☐

Сингулярные случайные величины и распределения.

Определение 8. x называется **точкой роста** монотонно возрастающей функции f, если $\forall\varepsilon > 0\ f(x+\varepsilon) - f(x-\varepsilon) > 0$.

Определение 9. Случайная величина ξ (и её распределение) называются **сингулярными**, если $F_\xi \in C(\mathbb{R})$ и мера множества точек F_ξ роста равна нулю.

Пример. Функция Кантора — функция, которая выглядит так: в нуле она равна нулю, в единице — единице, а во всех остальных точках строится так: отрезок $[0;1]$ делится на три части и в средней части равна среднему значению краёв (т.е. $\frac{1}{2}$). И так далее.
Оня является функцией распределения сингулярной случайной величины.

Теорема 2 (Теорема Лебега)**.** *Пусть F — функция распределения. Тогда существуют единственные F_{disc}, F_{ac} и F_{sing}, которые в сумме дают F.*
Без доказательства.

2.2 Многомерные случайные величины.

Распределение многомерных случайных величин.

Определение 10. Вектор ξ называется **случайным вектором** или **многомерной случайной величиной**, если ξ_i — случайная величина.

Определение 11. Распределением случайного вектора называется функция P_ξ, определённая на $\mathfrak{B}^n$, заданная так:

$$P_\xi(B_1;\dots;B_n) = P(\{(\omega_1;\dots;\omega_n) \mid \xi_1(\omega_1) \in B_1 \wedge \dots \wedge \xi_n(\omega_n) \in B_n\})$$

Последнее обычно обозначается так: $P(\xi_1 \in B_1, \xi_2 \in B_2;\dots;\xi_n \in B_n)$.

Определение 12. Функцией распределения случайного вектора ξ называется функция

$$F_\xi(t_1;\dots;t_n) = P_\xi(\forall i \in [1:n]\ \xi_i \leqslant t_i)$$

Утверждение.

$$\begin{aligned} P(\forall i \in [1:n]\ a_i < \xi_i \leqslant b_i) = \\ & F(b_1;b_2;\dots;b_n) \\ & - F(a_1;b_2;\dots;b_n) - F(b_1;a_2;\dots;b_n) - \dots - F(b_1;b_2;\dots;a_n) \\ & + \vdots \\ & \pm F(a_1;a_2;\dots;a_n) \end{aligned}$$

То есть в этой сумме участвует F от a_i и b_i в произвольном сочетании, при этом минус стоит там, где нечётное количество a_i.

Доказательство.

$$P(\forall i \in [1:n]\ a_i < \xi_i \leqslant b_i) = P(\xi_1 \leqslant b_1, \forall i \in [2:n]\ a_i < \xi_i \leqslant b_i) - P(a_1 \geqslant \xi_1, \forall i \in [2:n]\ a_i < \xi_i \leqslant b_i)$$

Проведя такую операцию несколько раз, получим искомое. □

Замечание. Если ввести обозначение $\Delta_{a_i;b_i}F = F(\cdot;\dots;b_i;\cdot;\dots;\cdots) - F(\cdot;\dots;a_i;\cdot;\dots;\cdots)$, то арифметическая сумма из утверждения выше записывается как

$$\Delta_{a_1;b_1}\Delta_{a_2;b_2}\cdots\Delta_{a_n;b_n}F$$

Свойство 12.1.

$$F(+\infty; \cdots; +\infty) = 1$$

$$F(-\infty; \cdots; -\infty) = 0$$

F непрерывна справа.

Теорема 3. *Если функция распределения удовлетворяет трём свойствам выше, то она является функцией распределения некоторого случайного вектора.*

Доказательство. Аналогично одномерному случаю. □

Определение 13. Случайный вектор (и его распределение) называется **дискретным**, если существует не более чем счётное множество E такое что $P(\xi \in E) = 1$.

Пример. Полиномиальное распределение. Пусть $p = (p_1; \ldots; p_m)$, где $\sum p_i = 1$. Обозначается $\mathrm{Poly}(n; p)$. Физическая интерпретация такая: мы бросаем кубик с m гранями n раз. Пусть $S_{n,j}$ — количество исходов типа j в n независимых испытаниях. Тогда искомая случайная величина обладает распределением

$$P(S_{n,1} \in B_1; S_{n,2} \in B_2; \ldots; S_{n,m} \in B_m)$$

Рассмотрим $P(S_{n,1} = k_1; S_{n,2} = k_2; \ldots; S_{n,m} = k_m)$, где $\sum k_m = n$. Чему равна такая вероятность? Ну,

$$\frac{n!}{k_1!k_2! \cdots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

Замечание. Несложно заметить, что штука справа — слагаемые в сумме $(p_1 + p_2 + \cdots + p_m)^n$.

Определение 14. Случайный вектор ξ (и его распределение) называется **абсолютно непрерывным**, если существует $p \in L(\mathbb{R}^n \to \mathbb{R})$ с неотрицательными значениями такая что $P(\xi \in B) = \int_B P \, \mathrm{d}\mu_n$ (тут интеграл n-кратный). В таком случае p называется **плотностью** ξ.

Пример. Случайный вектор ξ имеет стандартное многомерное нормальное распределение $\mathrm{N}(\mathbb{0}_n; E_n)$, если его плотность равна

$$p(x_1; \ldots; x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{\sum\limits_{i=1}^{n} x_i^2}{2}\right)$$

Свойство 14.1. *Несложно заметить, что это произведение плотностей одномерных стандартных нормальных распределений.*

Пример. Случайный вектор η имеет многомерное нормальное распределение $\mathrm{N}(\mu; \Sigma)$ (где $\mu \in \mathbb{R}^n$, Σ — симметричная матрица $n \times n$ с неотрицательными собственными числами), если он равен $\sqrt{\Sigma}\xi + \mu$, где ξ — стандартный многомерный гауссовский вектор.

Замечание. В случае $\Sigma > 0$ можно написать плотность:

$$p(y) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu)\right)$$

Независимые случайные величины.

Определение 15. Случайные величины $\xi_1; \ldots; \xi_n$ называются **независимыми**, если для любых борелевских множеств $B_1; \ldots; B_n$

$$P(\xi_1 \in B_1 \land \cdots \land \xi_n \in B_n) = P(\xi_1 \in B_1) \cdot \cdots \cdot P(\xi_n \in B_n)$$

Определение 16. Случайные величины $\{\xi_i\}_{i=1}^{\infty}$ **независимы**, если

$$\forall m \ \xi_1; \ldots; \xi_m \text{ независимы}$$

Теорема 4. *Случайные величины независимы тогда и только тогда, когда их совместная функция распределения равна произведению одномерных функций распределения.*

Доказательство. В одну сторону очевидно, в другую — доказательства не будет. □

Теорема 5 (Критерий независимости дискретных случайных величин)**.** *Пусть $\xi_1;\ldots;\xi_n$ — дискретны. Пусть ξ_k имеет своим множеством значений $\{x_{i,k}\}_i$. Тогда эти величины независимы тогда и только тогда, когда*

$$\forall x_{\textit{индексы}}\ P(\xi_1 = x_{i_1,1} \wedge \cdots \wedge \xi_n = x_{i_n,n}) = P(\xi_1 = x_{i_1,1}) \cdot \cdots \cdot P(\xi_n = x_{i_n,n})$$

Доказательство. Необходимость очевидна, теперь достаточность, которая тоже довольно проста. Условие $\xi_k \in B_k$ можно представить как $\xi_k = x_{j_1,k} \vee \cdots \vee \xi_k = x_{j_m,k}$. Отсюда используем сумму вероятностей и предпосылку импликации. □

Теорема 6 (Критерий независимости абсолютно непрерывных случайных величин)**.** *Абсолютно непрерывные случайные величины независимы тогда и только тогда, когда их совместная плотность равна произведению одномерных плотностей.*

Доказательство.

$$\int\limits_{-\infty}^{t_1} \cdots \int\limits_{-\infty}^{t_n} p(x_1;\ldots;x_n)\ \mathrm{d}x_1 \cdots \mathrm{d}x_n = F(t_1;\ldots;t_n) = F(t_1) \cdot \cdots \cdot F(t_n) = \prod \int\limits_{-\infty}^{t_k} p(x_k)\ \mathrm{d}x_k$$

Это если нам дана независимость, а узнать мы хотим указанное в теореме (дифференцируем указанное равенство). Следствие обратно сами докажете, используя аналогичные формулы. □

Пример. Пусть ξ_1 и ξ_2 — независимые Бернуллиевские случайные величины с параметром p. Что будет, если их сложить? Очевидно, получится биномиальное распределение $\mathrm{Bin}(2;p)$.
Тривиально по индукции докажете, что это для любого количества одинаковых независимых $\mathrm{Bern}(p)$ работает.

Пример. Пусть $\xi_1 = \mathrm{Pois}(\lambda_1)$, $\xi_2 = \mathrm{Pois}(\lambda_2)$ независимы. Чему равна сумма этих случайных величин пуассона? Получится $\mathrm{Pois}(\lambda_1 + \lambda_2)$, докажете, опять же, сами, рассмотрев $P(\xi_1 + \xi_2 = k)$.

Пример. Пусть $\xi_1, \ldots, \xi_n \sim \mathrm{Geom}(p)$ и независимы. Тогда сумма имеет распределение $\mathrm{NB}(n;p)$

Пример. Пусть $\xi_1, \xi_2 \sim \mathrm{Exp}(\lambda)$ независимы. И мы опять считаем сумму $\xi_1+\xi_2$. Как несложно заметить, нам нужно посчитать

$$p_{\xi_1,\xi_2}(t) = \int\limits_{-\infty}^{\infty} p_1(x)p_2(t-x)\ \mathrm{d}x$$

Если $t \leqslant 0$, ответ 0. Иначе искомая плотность равна $\lambda^2 t e^{-\lambda t}$. Это же Gamma-распределение $\Gamma(2;\lambda)$.

Пример. Сложение нормальных распределений. Пусть $\xi_1 \sim \mathrm{N}(\mu_1;\sigma_1^2)$, $\xi_2 \sim \mathrm{N}(\mu_2;\sigma_2^2)$ независимы, тогда $\xi_1 + \xi_2 \sim \mathrm{N}(\mu_1+\mu_2;\sigma_1^2+\sigma_2^2)$. Докажите это сами.

3 Об интегралах.

Замечание. Пусть $(\Omega;\mathfrak{A};P)$ — вероятностное пространство, ξ — случайный вектор с распределением P_ξ. Пусть $g\colon \mathbb{R}^n \to \mathbb{R}$ измерима. Тогда

$$\int_\Omega g(\xi(\omega))\ P(\mathrm{d}\omega) = \int_{\mathbb{R}^n} g(x)\ P_\xi(\mathbb{d}x)$$

Для дискретного случая это равно $\sum g(x_k)P(\xi = x_k)$, а для абсолютно непрерывного — $\int g(x)p(x)\ \mathrm{d}x$.

Теорема 1 (Теорема Фубини)**.** *Пусть ξ, η —независимые случайные величины, $g\colon \mathbb{R}^2 \to \mathbb{R}$ измерима. Тогда*

$$\iint g(x,y)\ P_{\xi,\eta}(\mathrm{d}x,\mathrm{d}y) = \iint g(x;y)\ \mathrm{d}F_\xi(x)\ \mathrm{d}F_\eta(y) = \iint g(x;y)\ \mathrm{d}F_\eta(y)\ \mathrm{d}F_\xi(x)$$

4 Числовые характеристики случайных величин.

4.1 Математическое ожидание.

Определение 1. Пусть ξ — случайная величина, тогда её **математическим ожиданием** называется

$$\mathsf{E}\,\xi = \int\limits_{\mathbb{R}} x \,\mathrm{d}F_\xi(x)$$

Свойство 1.1. *Пусть $\xi_1; \ldots; \xi_n$ — случайные величины, $g\colon \mathbb{R}^n \to \mathbb{R}$. Тогда*

$$Eg(\xi_1, \ldots, \xi_n) = \int\limits_{\mathbb{R}^n} g(\xi_1, \ldots, \xi_n)\ P_{\xi_1;\ldots;\xi_n}(\mathrm{d}x_1; \ldots; \mathrm{d}x_n)$$

Свойство 1.2. *Математическое ожидание линейно.*

Свойство 1.3. *Математическое ожидание произведения независимых величин равно произведению их математических ожиданий (обратное неверно).*

Теорема 1 (Неравенство Маркова)**.** *Если $\xi \geqslant 0$ и существует $\mathsf{E}\,\xi$, то*

$$\forall c > 0\ P(X \geqslant c) \leqslant \frac{\mathsf{E}\,\xi}{c}$$

Доказательство.

$$P(X \geqslant c) = \int_{x \geqslant c} \mathrm{d}F(x) \leqslant \int_{x \geqslant c} \frac{x}{c}\mathrm{d}F(x) \leqslant \frac{1}{c}\int\limits_{\mathbb{R}} x\mathrm{d}F(x) = \frac{1}{c}E\xi$$

□

Следствие 0.1. *Ожидание индикатора $\xi \in B$ равно $P(\xi \in B)$.*

Свойство 1.4. *Если $P(\xi \geqslant 0) = 1$, то $\mathsf{E}\,\xi \geqslant 0$.*

Свойство 1.5. *Если $\xi \geqslant 0$ и $\mathsf{E}\,\xi = 0$, то $P(\xi = 0) = 1$.*

Доказательство.

$$1 = P(\xi \geqslant 0) = P(\xi \geqslant 0) + P(\xi < c) \leqslant P(\xi < c) \leqslant 1$$

Отсюда тут везде равенства, значит $P(\xi < c) = 1$. □

Свойство 1.6. *Если ξ дискретно с носителем $\mathbb{N}$, то $\mathsf{E}\,\xi = \sum\limits_{k=1}^{\infty} P(\xi \geqslant k)$.*

Очевидно.

Свойство 1.7. *Если ξ непрерывно $\xi \geqslant 0$, то $\mathsf{E}\,\xi = \int\limits_0^{+\infty} P(\xi \geqslant r)\ \mathrm{d}r$*

4.2 Дисперсия.

Определение 2. **Дисперсией случайной величины** ξ называется $\operatorname{Var}\xi = \mathsf{E}(\xi - \mathsf{E}\,\xi)^2$. Также обозначается $\mathsf{D}\,\xi$.

Свойство 2.1. $\mathsf{D}\,\xi \geqslant 0$.

Свойство 2.2.

$$\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2$$

Свойство 2.3.

$$\mathsf{D}(\xi + c) = \mathsf{D}\,\xi$$

Свойство 2.4.

$$\mathsf{D}\, a\xi = a^2\,\mathsf{D}\,\xi$$

Свойство 2.5. $\mathsf{D}(\xi \pm \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta \pm 2(\mathsf{E}\,\xi\eta - \mathsf{E}\,\xi \cdot \mathsf{E}\,\eta)$

Следствие 0.2. *Если ξ, η независимы, то* $\mathsf{D}(\xi \pm \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta$.

Теорема 2 (Неравенство Чебышёва)**.** *Пусть ξ — случайная величина, существуют $\mathsf{E}\,\xi$ и $\mathsf{D}\,\xi$. Тогда*

$$\forall \varepsilon > 0 \;\; P(|\xi - \mathsf{E}\,\xi| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}\,\xi}{\eta^2}$$

Доказательство. Примените неравенство Маркова для $P(|\xi - \mathsf{E}\,\xi|^2 \geqslant \varepsilon^2)$. ☐

Свойство 2.6.

$$P(\xi - \mathsf{E}\,\xi \geqslant 0) = 1$$

.

Доказательство. См. свойство математического ожидания. ☐

Свойство 2.7.

$$\mathsf{D}\,\xi = \min_{a \in \mathbb{R}} \mathsf{E}(\xi - a)^2$$

Доказательство. Искомая формула под минимумом — парабола, вершину которой посчитаете сами и найдёте, что она именно в нужной точке. ☐

4.3 Примеры.

Пример.

$$\mathrm{Bern}(p)$$

$$\mathsf{E}\,\xi = p$$

$$\mathsf{D}\,\xi = pq$$

Пример.

$$\mathrm{Bin}(n, p)$$

Чтобы доказать свойства, проще всего вспомнить, что это сумма независимых Вернуллиевских величин.

$$\mathsf{E}\,\xi = np$$

$$\mathsf{D}\,\xi = npq$$

Пример.

$$\mathrm{Pois}(\lambda)$$

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

Для подсчёта дисперсии заметим, что $\mathsf{E}(\xi(\xi - 1)) = \mathsf{D}\,\xi + (\mathsf{E}\,\xi)^2 - \mathsf{E}\,\xi$. Тогда

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

Отсюда

$$\mathsf{D}\,\xi = \lambda$$

Пример.

$$\mathrm{Geom}(p)$$

Будем считать носителем $\mathbb{Z}_+$.

$$\mathsf{E}\,\xi = \sum_{k=1}^{\infty} P(x \geqslant k) = \sum_{k=1}^{\infty}\sum_{i=k}^{\infty}(1-p)^i p = \sum_{k=1}^{\infty} p\frac{(1-p)^k}{1-(1-p)} = \frac{1-p}{p}$$

$$\mathsf{E}(\xi(\xi-1)) = \sum_{k=2}^{\infty} k(k-1)(1-p)^k p = (1-p)^2 p \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = (1-p)^2 p \sum_{k=2}^{\infty} \frac{\partial^2 (1-p)^k}{\partial p^2} = (1-p)^2 p \frac{\partial^2}{\partial p^2}\sum_{k=2}^{\infty}(1-p)^k \stackrel{\text{дальше сам}}{=} 2$$

$$\mathsf{D}\,\xi = \frac{1-p}{p^2}$$

Пример. Равномерное распределение на $[a;b]$.
Заметим тот факт, что ξ — равномерное распределение на $[a;b]$ можно выразить через η — равномерное распределение на $[0;1]$:

$$\xi = (b-a)\eta + a$$

Отсюда

$$\mathsf{E}\,\xi = \frac{a+b}{2}$$

$$\mathsf{D}\,\xi = \frac{(a+b)^2}{12}$$

Пример.

$$\mathrm{N}(\mu;\sigma^2)$$

$$\mathsf{E}\,\xi = \mu$$

$$\mathsf{D}\,\xi = \sigma^2$$

Сами как-нибудь проинтегрируете, мне лень.

Пример.

$$\mathrm{Exp}(\lambda)$$

$$\mathsf{E}\,\xi = \int_0^{+\infty} x\lambda e^{-\lambda x}\,\mathrm{d}x = -xe^{-\lambda x}\Big|_0^{+\infty} + \frac{1}{\lambda}\int_0^{+\infty}\lambda e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$$

$$\mathsf{E}\,\xi^2 = \int_0^{+\infty} x^2\lambda e^{-\lambda x}\,\mathrm{d}x = -x^2e^{-\lambda x}\Big|_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty}\lambda x e^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda^2}$$

$$\mathsf{D}\,\xi = \frac{1}{\lambda^2}$$

Пример.

$$(n,\lambda) = \mathrm{Exp}(\lambda) + \cdots + \mathrm{Exp}(\lambda)$$

$$\mathsf{E}\,\xi = \frac{n}{\lambda}$$

$$\mathsf{D}\,\xi = \frac{n}{\lambda^2}$$

Доказывать для нецелого n не будем.

Пример.

$$\mathrm{Cauchy}(0;1)$$

$$\mathsf{E}\,\xi = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\,\mathrm{d}x}{1+x^2}$$

Мат. ожидание не существует.

4.4 Моменты и связанное с ними.

Определение 3. **Моментом случайной величины** ξ порядка k называется $\mathsf{E}\,\xi^k$.

Определение 4. **Абсолютным моментом случайной величины** ξ порядка k называется $\mathsf{E}\,|\xi|^k$.

Определение 5. **Центральным моментом случайной величины** ξ порядка k называется $\mathsf{E}(\xi - \mathsf{E}\,\xi)^k$.

Определение 6. **Абсолютным центральным моментом случайной величины** ξ порядка k называется $\mathsf{E}\,|\xi - \mathsf{E}\,\xi|^k$.

Определение 7. **Коэффициентом ассиметрии случайной величины** ξ называется $\frac{\mathsf{E}(\xi-\mathsf{E}\,\xi)^3}{\sigma^3}$.

Определение 8. **Коэффициентом эксцесса случайной величины** ξ называется $\frac{\mathsf{E}(\xi-\mathsf{E}\,\xi)^4}{\sigma^4} - 3$.

4.5 Мода.

Определение 9. **Модой дискретной случайной величины** ξ называется число $x_k^* = \operatorname{argmax}_{x_k} P(\xi = x_k)$.

Определение 10. **Модой непрерывной случайной величины** ξ называется число $x^* = \operatorname{argmax}_{x\in\mathbb{R}} p(x)$.

Определение 11. Случайная величина называется n**-модальной** (**унимодальной** для $n = 1$, **бимодальной** для $n = 2...$), если у неё n *локальных* максимумов.

4.6 Квантиль, медиана.

Определение 12. **Квантилью случайной величины** ξ порядка $\alpha \in (0;1)$ называется такое число q_α, что $P(\xi \geqslant q_\alpha) \geqslant 1 - \alpha$ и $P(\xi \leqslant q_\alpha) \leqslant \alpha$.
Квантиль порядка $\frac{1}{2}$ называется **медианой**.

Замечание. В дискретном случае определяется неоднозначно. Есть несколько способов это исправить, обращайте внимание на то, какой используется в Вашей задаче.

Теорема 3. *Пусть* $m = \operatorname{med}\xi$*. Тогда* $m = \operatorname{argmin}_{a\in\mathbb{R}} \mathsf{E}\,|\xi - a|$*, если для любого a искомое ожидание существует.*

Доказательство. Н.У.О. $m = 0$, иначе сдвинем ξ. Сравним $\mathsf{E}\,|\xi - c|$ и $\mathsf{E}\,|\xi|$. Пусть $c > 0$. Тогда

$$|\xi - c| - |\xi| = \begin{cases} -c & \xi > c \\ c - 2\xi & 0 < \xi \leqslant c, c - 2\xi \geqslant -c \\ c & x \leqslant 0 \end{cases}$$

отсюда

$$\mathsf{E}(|\xi-c|-|\xi|) = \mathsf{E}(|\xi-c|-|\xi|)\mathbb{1}(\xi \leqslant 0) + \mathsf{E}(|\xi-c|-|\xi|)\mathbb{1}(\xi > 0) \geqslant c\,\mathsf{E}\,\mathbb{1}(\xi \leqslant 1) - c\,\mathsf{E}\,\mathbb{1}(\xi > c) = c(2P(\xi \leqslant 0) - 1) \geqslant 0$$

То есть $\mathsf{E}\,|\xi - c|$ больше либо равно $\mathsf{E}\,|\xi|$. □

4.7 Ковариация. Коэффициентом корреляции.

Определение 13. **Ковариацией случайных величин** ξ и η называется $\operatorname{cov}(\xi;\eta) = \mathsf{E}(\xi - \mathsf{E}\,\xi)\,\mathsf{E}(\eta - \mathsf{E}\,\eta)$

Свойство 13.1.

$$\operatorname{cov}(\xi;\eta) = \mathsf{E}(\xi\eta) - \mathsf{E}\,\xi\,\mathsf{E}\,\eta$$

Свойство 13.2.

$$\mathsf{D}(\xi_1 + \cdots + \xi_n) = \sum \mathsf{D}\,\xi_i + \sum \operatorname{cov}(\xi_i;\xi_j)$$

Свойство 13.3.

$$\operatorname{cov}(\xi;\eta) = \operatorname{cov}(\eta;\xi)$$

Свойство 13.4.

$$\operatorname{cov}(\xi;\xi) = \mathsf{D}(\xi)$$

Свойство 13.5. *Ковариация линейна по обоим аргументам.*

Свойство 13.6.

$$\operatorname{cov}(\xi;c) = 0$$

Свойство 13.7. *Ковариация независимых величин равна нулю. Обратное в общем случае неверно.*

Пример. Пусть $\xi \sim \mathrm{N}(0;1)$, $\eta = \xi^2$. Они не независимы, при этом

$$\mathsf{E}\,\xi = 0$$

$$\mathsf{E}\,\xi\eta = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} x^3 e^{-x^2/2}\,\mathrm{d}x = 0$$

Значит корреляция равна нулю.

Определение 14. **Коэффициентом корреляции случайных величин** ξ и η называется $\rho(\xi;\eta) = \frac{\operatorname{cov}(\xi;\eta)}{\sqrt{\mathsf{D}\,\xi\,\mathsf{D}\,\eta}}$

Свойство 14.1.

$$|\rho(\xi;\eta)| \leqslant 1$$

Доказательство. Чтобы это доказать, посмотрим на следующее свойство и рассмотрим $\mathsf{D}(\tilde{\xi} + \tilde{\eta})$ и $\mathsf{D}(\tilde{\xi} - \tilde{\eta})$. □

Свойство 14.2.

$$\rho(\xi;\eta) = \operatorname{cov}\left(\underbrace{\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}}_{\tilde{\xi}};\underbrace{\frac{\eta - \mathsf{E}\,\eta}{\sqrt{\mathsf{D}\,\eta}}}_{\tilde{\eta}}\right)$$

Свойство 14.3. *Если* $\eta = a\xi + b, a \neq 0$*, то*

$$\rho(\xi;\eta) = \operatorname{sign} a$$

Свойство 14.4. *Если* $\rho(\xi;\eta) = 1$*, то* $\tilde{\xi} - \tilde{\eta} = c$*. Если* $\rho(\xi;\eta) = -1$*, то* $\tilde{\xi} + \tilde{\eta} = c$*.*

Определение 15. Две величины **некоррелированы**, если их ковариация (следовательно, коэффициент корреляции) равна нулю.

Утверждение. *Независимость влечёт некоррелированность, обратное в общем случае неверно.*

4.8 Характеристики векторных случайных величин.

Определение 16. **Математическим ожиданием случайного вектора** называется вектор математических ожиданий его компонент.

Теорема 4. $\mathsf{E}\,A\xi = A\,\mathsf{E}\,\xi$

Определение 17. **Дисперсией случайного вектора** ξ называется

$$\mathsf{E}(\xi - \mathsf{E}\,\xi)(\xi - \mathsf{E}\,\xi)^T$$

Несложно заметить, что это матрица ковариаций компонент.

Свойство 17.1.

$$\mathsf{D}\,\xi = (\mathsf{D}\,\xi)^T$$

Свойство 17.2.

$$\mathsf{D}\,\xi \geqslant 0$$

(Речь идёт про неотрицательную определённость.)

Доказательство. Рассмотрим $T^T\,\mathsf{D}\,\xi T$

$$T^T\,\mathsf{D}\,\xi T = T^T\,\mathsf{E}(\xi - \mathsf{E}\,\xi)(\xi - \mathsf{E}\,\xi)^T T = \mathsf{E}\,\underbrace{T^T(\xi - \mathsf{E}\,\xi)}\,\underbrace{(\xi - \mathsf{E}\,\xi)^T T} = \mathsf{E}(T^T(\xi - \mathsf{E}\,\xi))^2 \geqslant 0$$

□

Свойство 17.3.

$$\mathsf{D}(\xi + c) = \mathsf{D}\,\xi$$

Свойство 17.4. *Если ξ и η независимы, то*

$$\mathsf{D}(\xi + \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta$$

Доказательство. Если $i = j$, то просто имеем дисперсии на диагоналях, и для одномерных дисперсий мы знаем такой результат.
Иначе раскроем коварицаии по линейности получим искомое. □

Пример.

$$\text{Poly}(n, p)$$

Вспомним, что наша случайная величина является суммой независимых случайных величин с распределением $\text{Poly}(1, p)$. Математические ожидание же $\xi_1 \sim \text{Poly}(1, p)$ равно $\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$.

Дисперсия также считается как n умножить на $\mathsf{D}\,\xi_1$. Осталось только это посчитать.

$$\mathsf{D}\,\xi_1 = \begin{pmatrix} \text{cov}(\xi_1^{(1)}, \xi_1^{(1)}) & \text{cov}(\xi_1^{(1)}, \xi_1^{(2)}) & \cdots & \text{cov}(\xi_1^{(1)}, \xi_1^{(m)}) \\ \text{cov}(\xi_1^{(2)}, \xi_1^{(1)}) & \text{cov}(\xi_1^{(2)}, \xi_1^{(2)}) & \cdots & \text{cov}(\xi_1^{(2)}, \xi_1^{(m)}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\xi_1^{(m)}, \xi_1^{(1)}) & \text{cov}(\xi_1^{(m)}, \xi_1^{(2)}) & \cdots & \text{cov}(\xi_1^{(m)}, \xi_1^{(m)}) \end{pmatrix}$$

Если $i = j$, то $\text{cov}(\xi_1^{(i)} l \xi_1^{(i)}) = \mathsf{D}\,\xi_1^{(i)} = \mathsf{E}(\xi_1^{(i)})^2 - (\mathsf{E}\,\xi_1^{(i)})^2 = p_i(1 - p_i)$.
Если $i \neq j$, то $\text{cov}(\xi_1^{(i)} l \xi_1^{(j)}) = \mathsf{E}(\xi_1^{(i)}\xi_1^{(j)}) - \mathsf{E}\,\xi_1^{(i)}\,\mathsf{E}\,\xi_1^{(j)} = -p_i p_j$. Посмотрим на $\mathsf{E}(\xi_1^{(i)}\xi_1^{(j)})$. Тут просто выражение под мат.ожиданием ноль, ведь в ξ_1 только одна компонента равна единице.

Пример.

$$\text{N}(\mu; \Sigma)$$

Дл этого рассмотрим стандартный Гауссовский вектор, а потом рассмотрим нестандартный. Ну, у него компоненты — независимые $\text{N}(0; 1)$. У каждого мат. ожидание ноль, а дисперсия один. И в силу независимости ковариации равны нулю, а значит $\mathsf{E}\,\xi = \mathbb{0}$, $\mathsf{D}\,\xi$ — единичная матрица.
Теперь $\text{N}(\mu; \Sigma)$. Мы знаем, что если $\eta = A\xi + \mu$ и $\xi \sim \text{N}(\mathbb{0}; E)$, то $\eta \sim \text{N}(\mu; AA^T)$, а отсюда $\mathsf{E}\,\eta = \mu$, $\mathsf{D}\,\eta = AA^T$.
При этом на линейной алгебре доказывали, что для $\Sigma \geqslant 0$ и $\Sigma = \Sigma^T$ у Σ существует корень. Отсюда μ; подходит под предыдущий случай.

Утверждение.

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \sim \text{N}(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \Rightarrow \xi_1 \sim \text{N}(\mu_1; \Sigma_{11}), \xi_2 \sim \text{N}(\mu_2; \Sigma_{22})$$

Доказательство. Рассмотрим спектральное представление Σ:

$$\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} = \begin{pmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{pmatrix}\begin{pmatrix}\Lambda_1 & \mathbb{0}\\ \mathbb{0} & \Lambda_2\end{pmatrix}\begin{pmatrix}U_{11}^T & U_{21}^T\\ U_{12}^T & U_{22}^T\end{pmatrix}$$

Корень из этой байды равен тому же самому, только с корнями $\sqrt{\Lambda_i}$. В итоге

$$\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} = \begin{pmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{pmatrix}\begin{pmatrix}\sqrt{\Lambda_1} & \mathbb{0}\\ \mathbb{0} & \sqrt{\Lambda_2}\end{pmatrix}\begin{pmatrix}U_{11}^T & U_{21}^T\\ U_{12}^T & U_{22}^T\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$$

Где ξ_1 и ξ_2 — нормальные Гауссовские векторы.
Для того, чтобы проверить то, что мы хотим, надо всего лишь посмотреть на эту формулу и узнать, что будет её первой компонентой, и что — второй. Получим именно то, что нужно. □

Утверждение. *Для нормальных случайных векторов некоррелированность влечёт независимость.*

Доказательство. Пусть мы знаем некоррелированность. Тогда

$$\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \sim \mathrm{N}(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}; \begin{pmatrix}\sigma_{\xi_1}^2 & \mathbb{0}\\ \mathbb{0} & \sigma_{\xi_2}^2\end{pmatrix})$$

Если $\sigma_{\xi_1} = 0$ или $\sigma_{\xi_2} = 0$, то всё очевидно. Иначе

$$p(x_1; x_2) = \frac{1}{2\pi\sqrt{\sigma_{\xi_1}^2\sigma_{\xi_2}^2}}\underbrace{\exp\left(-\frac{1}{2}\begin{pmatrix}x_1-\mu_1\\ x_2\mu_2\end{pmatrix}^T\begin{pmatrix}\sigma_{\xi_1}^{-2} & \mathbb{0}\\ \mathbb{0} & \sigma_{\xi_2}^{-2}\end{pmatrix}\begin{pmatrix}x_1-\mu_1\\ x_2-\mu_2\end{pmatrix}\right)}_{\exp\left(-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_{\xi_1}^2}+\frac{(x_2-\mu_2)^2}{\sigma_{\xi_2}^2}\right)\right)}$$

□

5 Вероятностные неравенства.

Замечание. См. также 1 2

Теорема 1 (Слабый закон больших чисел)**.** *Пусть $\xi_1; \ldots; \xi_n$ — независимые одинаково распределённые величины. И пусть у них ожидание μ и дисперсия σ^2. Тогда*

$$\forall\varepsilon > 0\ P(|\overline{\xi} - \mu| \geqslant \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$$

Где

$$\overline{\xi} = \frac{\sum\limits_{i=1}^{n}\xi_i}{n}$$

Доказательство.

$$P(|\overline{\xi} - \mu| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}\,\overline{\xi}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

□

Определение 1.

$$L_2 = \{\xi \mid \mathsf{E}\,\xi^2 < \infty\}$$

Утверждение. *Пусть $\xi, \eta \in L_2$, Тогда*

$$|\,\mathsf{E}\,\xi\eta| \leqslant \mathsf{E}\,\xi^2\,\mathsf{E}\,\eta^2$$

Доказательство. Это просто неравенство Коши — Буняковского — Шварца для скалярного произведения $(\xi; \eta) \mapsto \mathsf{E}\,\xi\eta$. □

Утверждение. *Пусть $\xi, \eta \in L_2$, Тогда*

$$(\operatorname{cov}(\xi;\eta))^2 \leqslant \mathsf{D}\,\xi\,\mathsf{D}\,\eta$$

Доказательство. Это просто неравенство Коши — Буняковского — Шварца для скалярного произведения $(\xi;\eta) \mapsto \mathsf{E}(\xi - \mathsf{E}\,\xi)(\eta - \mathsf{E}\,\eta)$. □

Утверждение. *Если g выпукла вниз, то $g(\mathsf{E}\,\xi) \leqslant \mathsf{E}\,g(\xi)$.*

Доказательство. неравенство Йенсена. □

Утверждение. *Если p, q — сопряжённые показатели, то*

$$\mathsf{E}\,|\xi\eta| \leqslant (\mathsf{E}\,|\xi|^p)^{1/p}(\mathsf{E}\,|\eta|^q)^{1/q}$$

Доказательство. Неравенство Гёльдера. □

Утверждение. *Если $p \geqslant 1$, то*

$$(\mathsf{E}\,|\xi+\eta|^p)^{1/p} \leqslant (\mathsf{E}\,|\xi|^p)^{1/p} + (\mathsf{E}\,|\eta|^p)^{1/p}$$

Доказательство. Неравенство Минковского. □

Утверждение (неравенства Ляпунова)**.** *Пусть $p < q$. Тогда*

$$(\mathsf{E}\,|\xi|^p)^{1/p} \leqslant (\mathsf{E}\,|\xi|^q)^{1/q}$$

6 Условные распределения, мат. ожидания и дисперсии.

Утверждение. *Пусть $\xi_1;\ldots;\xi_n$ — дискретный вектор. Тогда*

$$P(\xi_1 = x_1, \ldots, \xi_k = x_k \mid \xi_{k+1} = x_{k+1}, \ldots, \xi_n = x_n) = \frac{P(\xi_1 = x_1, \ldots, \xi_n = x_n)}{P(\xi_{k+1} = x_{k+1}, \ldots, \xi_n = x_n)}$$

Пусть $\xi_1;\ldots;\xi_n$ — абсолютно непрерывный вектор. Тогда

$$p(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n) = \frac{p_{\xi_1;\ldots;\xi_n}(x_1, \ldots, x_n)}{p_{\xi_{k+1};\ldots;\xi_n}(x_{k+1}, \ldots, x_n)}$$

Следствие 0.1. *Если $(\xi;\eta)$ дискретны, то*

$$P(\xi = k) = \sum_j P(\xi = k \mid \eta = y_j) P_\eta(\eta = y_i)$$

Если $(\xi;\eta)$ непрерывны, то

$$p(x) = \int p(x \mid y) p_\eta(y)\ \mathrm{d}y$$

Определение 1. Условное математическое ожидание $E(\xi \mid \eta)$ в дискретном случае равно

$$\sum_k x_k P(\xi = x_k \mid \eta)$$

в непрерывном случае равно

$$\int x p(x \mid y)\ \mathrm{d}x$$

Свойство 1.1.

$$\mathsf{E}(\xi \mid \eta) = \operatorname*{argmin}_{f(\eta) \in L_2} \mathsf{E}(\xi - f(\eta))^2$$

Свойство 1.2.

$$\mathsf{E}(\xi \mid \xi) = \xi$$

Свойство 1.3. *Если ξ и η независимы, то*

$$\mathsf{E}(\xi \mid \eta) = \mathsf{E}\,\xi$$

Утверждение (Формула полной вероятности для математического ожидания)**.**

Свойство 1.4.

$$\mathsf{E}(\xi) = \mathsf{E}\,\mathsf{E}(\xi \mid \eta)$$

Определение 2. Условная дисперсия $\mathsf{D}(\xi \mid \eta)$ равна

$$\mathsf{E}((\xi - \mathsf{E}(\xi \mid \eta))^2 \mid \eta) = \mathsf{E}(\xi^2 \mid \eta) - (\mathsf{E}(\xi \mid \eta))^2$$

Свойство 2.1.

$$\mathsf{D}\,\eta = \mathsf{E}\,\mathsf{D}(\eta \mid \xi) + \mathsf{D}\,\mathsf{E}(\xi \mid \eta)$$

Доказательство.

$$\mathsf{E}\,\eta^2 = \mathsf{E}\,\mathsf{E}(\eta^2 \mid \xi) = \mathsf{E}\left(\mathsf{D}(\eta \mid \xi) + (\mathsf{E}(\eta \mid \xi))^2\right)$$
$$\mathsf{E}\,\eta = \mathsf{E}\,\mathsf{E}(\eta \mid \xi)$$

Тогда

$$\mathsf{D}\,\eta = \mathsf{E}\,\eta^2 - (\mathsf{E}\,\eta)^2 = \mathsf{E}\,\mathsf{D}(\eta \mid \xi) + \underbrace{\mathsf{E}(\mathsf{E}(\eta \mid \xi))^2 - (\mathsf{E}\,\mathsf{E}(\eta \mid \xi))^2}_{\mathsf{D}\,\mathsf{E}(\xi \mid \eta)}$$

□

Пример. Пусть

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathrm{N}(\mu; \Sigma)$$

Тогда

$$\mathsf{E}(\xi_2 \mid \xi_1) = \mu_2 + \frac{\operatorname{cov}(\xi_1; \xi_2)}{\mathsf{D}\,\xi_1}(\xi_1 - \mu_1)$$

Для доказательства введём

$$\eta = \xi_2 - \frac{\operatorname{cov}(\xi_1; \xi_2)}{\mathsf{D}\,\xi_1}\xi_1$$

Несложно заметить, что $\operatorname{cov}(\eta; \xi_1) = 0$, а значит они независимы. Тогда

$$\mathsf{E}(\xi_2 \mid \xi_1) = \mathsf{E}(\eta \mid \xi_1) + \frac{\operatorname{cov}(\xi_1; \xi_2)}{\mathsf{D}\,\xi_1}\,\mathsf{E}(\xi_1 \mid \xi_1)$$

7 Сходимости.

Определение 1. Пусть ξ_n, ξ — случайные величины на одном пространстве. Говорят, что ξ_n **сходится к ξ почти наверное**, если

$$P(\omega \mid \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)) = 1$$

Определение 2. Пусть ξ_n, ξ — случайные величины на одном пространстве. Говорят, что ξ_n **сходится к ξ по вероятности**, если

$$\forall \varepsilon > 0 \ P(|\xi_n - \xi| \geqslant \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Определение 3. Пусть ξ_n, ξ — случайные величины на одном пространстве. Говорят, что ξ_n **сходится к ξ в среднем порядка** p, если

$$\mathsf{E}\,|\xi_n - \xi|^p \xrightarrow[n\to\infty]{} 0$$

Определение 4. Пусть ξ_n, ξ — случайные величины. Говорят, что ξ_n **сходится к ξ по распределению**, если

$$\forall f \text{ непрерывна и ограничена } \mathsf{E}\, f(\xi_n) \to \mathsf{E}\, f(\xi)$$

Теорема 1. *Сходимость почти наверное влечёт сходимость по вероятности.*
Сходимость в среднем порядка p влечёт сходимость по вероятности.
Сходимость по вероятности влечёт сходимость по распределению.

Доказательство.$p \to d$

$$|\mathsf{E}(f(\xi_n)-f(\xi))| \leqslant \mathsf{E}\,|f(\xi_n)-f(\xi)|\mathbb{1}(|\xi_n-\xi| \geqslant \delta_\varepsilon)+\mathsf{E}\,|f(\xi_n)-f(\xi)|\mathbb{1}(|\xi_n-\xi| < \delta_\varepsilon) \leqslant 2MP(|\xi_n-\xi| \geqslant \delta_\varepsilon)+\varepsilon$$

Здесь δ_ε — из определения непрерывности f, M — то, чем ограничена f.

$L_p \to p$

$$P(|\xi_n - \xi| \geqslant \varepsilon) \leqslant P(|\xi_n - \xi|^p \geqslant \varepsilon^p) \leqslant \frac{\mathsf{E}\,|\xi_n - \xi|^p}{\varepsilon^p} \longrightarrow 0$$

$a.s. \to p$ Пусть $O = \{\omega \mid \xi_n(\omega) \nrightarrow \xi(\omega)\}$. Известно $P(O) = 0$.
Пусть $\varepsilon > 0$, $A_1 = \bigcup\limits_{m \geqslant n} \{\omega \mid |\xi_m(\omega) - \xi(\omega)| > \varepsilon\}$.
Очевидно, $A_{n+1} \subset A_n$, следовательно $P(A_n) \to P(A)$, где $A = \bigcap\limits_n A_n$.
Пусть $\omega \in O^{\complement}$. Тогда

$$\exists N(\varepsilon, \omega)\ \forall n > N\ |\xi_n(\omega) - \xi(\omega)| < \varepsilon$$

Это значит $\omega \notin A_N$, следовательно $\omega \in A^{\complement}$. Отсюда $P(A) = 0$.
Осталось заметить, что

$$P(|\xi_n - \xi| \geqslant \varepsilon) \leqslant P(A_n)$$

□

Замечание. Ни одна из обратных импликаций в общем случае неверна. Ни одна из импликаций между $a.s.$ и L_p не влечёт другую.

Пример.

$$d \nrightarrow p$$

Пусть $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. Пусть

$$\xi(\omega_i) = (-1)^i \qquad \xi_n(\omega_i) = (-1)^{i+n}$$

Тогда

$$\mathsf{E}\, f(\xi) = \frac{f(\xi(\omega_1)) + f(\xi(\omega_2))}{2} = \frac{f(1) + f(-1)}{2} = \mathsf{E}\, f(\xi_n)$$

Однако

$$P(\xi_n(\omega_i) - \xi(\omega_i)) = \pm 2 \nrightarrow 0$$

Пример.

$$a.s. \nrightarrow L_p$$

Возьмём равномерное распределение на $[0; 1]$. Пусть $\xi \equiv 0$,

$$\xi_n(\omega) = \begin{cases} e^n & \omega \in [0; 1/n] \\ 0 & \text{otherwise} \end{cases}$$

Очевидно, сходимость почти наверное есть. Но

$$\mathsf{E}\,|\xi_n - \xi|^p = \frac{e^{np}}{n} \nrightarrow 0$$

Пример.

$$p \nrightarrow a.s.$$

Пусть

$$\xi_{2^n}(\omega) = \begin{cases} \frac{1}{2^n} & \omega \in [0; \frac{1}{2^n}] \\ 0 & \text{otherwise} \end{cases}$$

$$\forall m \ \exists n : m \in (2^n; 2^{n+1}) \ \xi_m(\omega) = \begin{cases} \frac{1}{2^n} & \omega \in [\frac{m-2^n}{2^n}; \frac{m-2^n+1}{2^n}] \\ 0 & \text{otherwise} \end{cases}$$

Тогда

$$P(|\xi_m(\omega) - \xi(\omega)| > \varepsilon) \to 0$$

Пример.

$$L_p \nrightarrow a.s.$$

Возьмём предыдущий пример и заменим 2^n на $n!$.

Теорема 2. *Пусть F_n, F — функции распределения. Тогда сходимость по распределению равносильна*

$$\forall x \in C(F) \ F_n(x) \to F(x)$$

Доказательство. $\Rightarrow$ Пусть $x_0 \in C(F)$. Пусть

$$f_\varepsilon(t) = \begin{cases} 1 & t \leqslant x_0 \\ 0 & t \geqslant x_0 + \varepsilon \\ \text{линейно} & \text{otherwise} \end{cases}$$

$$F_n(x_0) = \int\limits_{-\infty}^{x_0} \mathrm{d}F_n(x) = \int\limits_{-\infty}^{x_0} f_\varepsilon(t) \ \mathrm{d}F_n(x) \leqslant= \int\limits_{-\infty}^{+\infty} f_\varepsilon(t) \ \mathrm{d}F_n(x) \to \int\limits_{-\infty}^{x_0+\varepsilon} f_\varepsilon(t) \ \mathrm{d}F(x) \leqslant F(x+\varepsilon)$$

$\Leftarrow$ Пока не будем доказывать.

$\square$

Замечание. Условие в теореме равносильно.

$$\forall x, y \ F_n(x) - F_n(y) \to F(x) - F(y)$$

Замечание. Если $f \in C(\mathbb{R})$, то

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$$

Замечание. Если F_n, F дискретны и имеют одинаковый носитель, то для сходимости по распределению достаточно доказать, что вероятность $\xi_n = x_k$ сходится к $\xi = x_k$.

Теорема 3 (Свойства сходимостей)**.**

1. *Пусть $\xi_n \xrightarrow{a.s.} \xi$, $|\xi_n| \overset{a.s.}{\leqslant} \eta$, $\mathsf{E}\eta < +\infty$. Тогда $\mathsf{E}\xi_n \to \mathsf{E}\xi$.*
2. *Пусть $\xi_n \xrightarrow{p} \xi$, $f \in C(\mathbb{R})$. Тогда $f(\xi_n) \xrightarrow{p} f(\xi)$.*
3. *Пусть $\xi_n - \eta_n \xrightarrow{p} 0$, $\xi_n \xrightarrow{p} \xi$, тогда $\eta_n \xrightarrow{p} \xi$.*
4. *Пусть $\xi_n - \eta_n \xrightarrow{p} 0$, $\eta_n \xrightarrow{d} \eta$, тогда $\xi_n \xrightarrow{p} \eta$.*
5. *Если $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{p} 0$, то $\xi_n\eta_n \xrightarrow{p} \xi$.*
6. *$\xi_n \xrightarrow{d} C$ тогда и только тогда, когда $\xi_n \xrightarrow{p} C$.*
7. *Если $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{p} C$, то $\xi_n + \eta_n \xrightarrow{d} \xi + C$, $\xi_n\eta_n \xrightarrow{d} \xi C$, если $C \neq 0$, то ещё $\frac{\xi_n}{\eta_n} \xrightarrow{d} \frac{\xi}{C}$.*

8. *Арифметические операции сохраняют сходимость по вероятности.*

Теорема 4. *Пусть $\mathcal{F}$ — множество функций распределений. Пусть $\mathcal{G}$ — множество неубывающих функций, непрерывных справа, $g(+\infty) \leqslant 1$, $g(-\infty) \geqslant 0$. Тогда для любой последовательности $\{g_n\} \subset \mathcal{G}$. Тогда существует подпоследовательность g_{n_k}, сходящаяся к $g \in \mathcal{G}$.*

Замечание. Для $\mathcal{F}$ это в общем случае неверно:

$$f_n(t) = \begin{cases} 0 & t < -n \\ \frac{1}{2} & t \in [-n; n) \\ 1 & t \geqslant n \end{cases}$$

Тогда эта штука сходится к $\frac{1}{2} \notin \mathcal{F}$.

Определение 5. Последовательность случайных величин называется **плотной**, если

$$\forall \varepsilon > 0 \ \exists N \ \inf_n P(-N \leqslant \xi_n \leqslant N) > 1 - \varepsilon$$

Определение 6. Класс $\mathcal{L}$ определяет распределение, если $\mathcal{L}$ — подмножество непрерывных ограниченных функций и

$$\forall f \in L \ \int f \ \mathrm{d}G = \int f \ \mathrm{d}F \Rightarrow F = G$$

Теорема 5. *Пусть $\{F_n\} \subset \mathcal{F}$, $\mathcal{L}$ определяет распределение. В таком случае существование $\lim F_n = F \in \mathcal{F}$ равносильно конъюнкции условий $\{F_n\}$ плотно и $\forall f \in \mathcal{L} \ \exists \lim \int f \ \mathrm{d}F_n$*

Доказательство. Следствие направо тривиально. Докажем налево. Известно, что у F_n есть сходящаяся в $\mathcal{G}$ подпоследовательность F_{n_k}. Пусть она сходится к $F_1 \in \mathcal{G}$.
Пусть $\varepsilon > 0$. Для него существует такое N, что

$$1 \geqslant F_n(N) > F_n(N) - F_n(-N) > 1 - \varepsilon$$

Пусть x_0 — точка непрерывности F_1 и $x_0 \geqslant N$. Тогда $1 \geqslant F_n(x_0) > 1 - \varepsilon$. Если мы рассмотрим $x_1 < -N$, то получим $F_n(x_1) < \varepsilon$. Отсюда F_1 на бесконечности равно 1, а на минус бесконечности — нулю. То есть $F_1 \in \mathcal{F}$.
$\lim \int f dF_{n_k} = \int f dF_1 = \int f dF_2 \Rightarrow F_1 = F_2$ — любая подпоследовательность сходится к одному распределению, следовательно $\exists \lim F_n F$ □

Следствие 0.1. *Пусть $\mathcal{L}$ определяет распределение, $\forall f \in \mathcal{L} \ \int f \ \mathrm{d}F_n \to \int f \ \mathrm{d}F$, $F \in \mathcal{G}$. Тогда из хотя бы одного из условий*

- *$\{F_n\}$ плотная.*
- *$F \in \mathcal{F}$.*
- *$f \equiv 1$.*

Следует $F_n \to F \in \mathcal{F}$.