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 $\forall_{i,j|i \neq j} A_i A_j = \varnothing$ 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 $\bigsqcup_{i=1}^{\infty} A_i \in \mathfrak{A}$, 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)**.**

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

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 сверху).*

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,* $A \subset \bigsqcup_{i=1}^{n} B_i$ *(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 $A_1, A_2, \ldots, A_n \in \mathfrak{A}$ **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. $\square$

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

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

$\square$

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*

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

Доказательство. Известно

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

Отсюда

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

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

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

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. $\square$

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 $P(B_n) \longrightarrow P(B)$. $\square$

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

Пример. Биномиальное распределение: $P(\xi = k) = \binom{n}{k} p^k q^{n-k}$. Обозначение: $\mathrm{Bin}(n; p)$, *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: $\mathrm{NB}(r; p)$. Также это обобщается на произвольное r при помощи гамма-функции.
В случае $r = 1$ распределение называется геометрическим. Геометрическое распределение — количество неудач до первого успеха.

Пример. Распределение Пуассона:

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

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

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 $P(\xi \in B) = \int\limits_B p(x)\ \mathrm{d}x$. 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* $\eta = c\xi + d = U[ac + d; bc + d]$

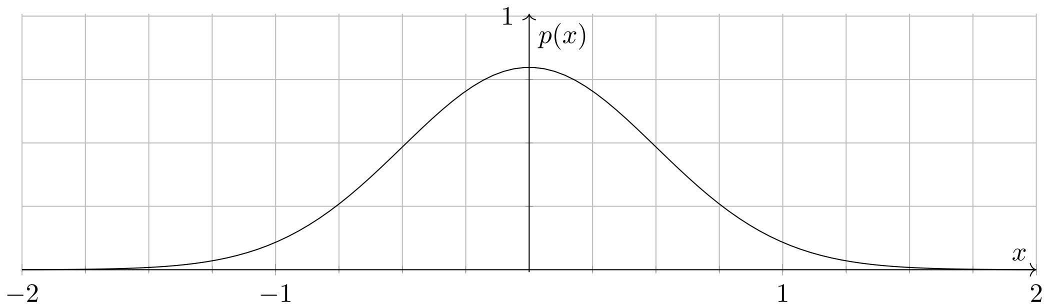
Доказательство.

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

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

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

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

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

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

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.
Она является функцией распределения сингулярной случайной величины.

Теорема 2 (Теорема Лебега)**.** *Пусть F — функция распределения. Тогда существуют единственные $a, b, c \geq 0$, $a + b + c = 1$, F_{disc}, F_{ac} и F_{sing} : $F = aF_{\mathrm{disc}} + bF_{\mathrm{ac}} + cF_{\mathrm{sing}}$.*
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 $p = (p_1;\dots;p_m)$, 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

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

Рассмотрим $P(S_{n,1} = k_1; S_{n,2} = k_2; \dots; S_{n,m} = k_m)$, 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 $P(\xi \in B) = \int\limits_B P\ \mathrm{d}\mu_n$ (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

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

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:

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

2.3 Независимые случайные величины

Пример. Пусть случайная величина задана таблицей

$\xi \backslash \eta$	y_1	y_2	$\dots$
x_1	p_{11}	p_{12}	$\dots$
x_2	p_{21}	p_{22}	$\dots$
$\vdots$	$\vdots$	$\vdots$	

$$p_{ij} = P(\xi = x_i, \eta = y_j) : \sum_{i,j} p_{ij} = 1$$

Теперь хотим найти вероятность:

$$P(\xi = x_i) = \sum_j P(\xi = x_i, \eta = y_i) = \sum_j = p_{ij}$$

Аналогично:

$$P(\eta = y_j) = \sum_i P(\xi = x_i, \eta = y_j) = \sum_i = p_{ij}$$

Теперь рассмотрим $x_i = y_i = i, i \in \mathbb{N}$
Хотим найти $P(\xi + \eta = k)$:
Во-первых $k \geq 2$ И теперь мы просто рассматриваем диагонали в табличке и получаем:

$$P(\xi + \eta = k) = \sum_{i=1}^{k-1} p_{i,k-i}$$

Пусть теперь $p(x, y)$ – совместная плотность на (X, Y)

$$p_X(x) = \int\limits_{D_x} p(x, y) dy$$

$$p_Y(y) = \int\limits_{D_y} p(x, y) dx$$

Тогда:

$$F_{X+Y}(t) = P(X + Y \leq t) = \iint\limits_{x+y \leq t} p(x, y) dx dy$$

Раньше мы определяли независимость вот так:

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

Определение 15. $X_1, \dots, X_n$ – независимы $\Leftrightarrow \forall B_1, \dots, B_n \in \mathfrak{B}$

$$P(X_1 \in B_1, \dots X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

$\{X_n\}_{n=1}^{\infty}$ – независимы $\forall m \in \mathbb{N} X_{1,\dots,m}$ – независимы

Теорема 4. *Критерий независимости*

$$X_1, \dots, X_n - \textit{независимы} \Leftrightarrow F_{x_1,\dots,x_n}(t_1, \dots, t_n) = \prod_{i=1}^{n} F_{x_i}(t_i), \forall t_1, \dots, t_n \in \mathbb{R}$$

Теорема 5. *Критерий независимости для дискретных с.в.*

$$X_1, \dots, X_n - \textit{дискретные нез с.в.} \Leftrightarrow P(X_1 = x_{i_1}, \dots, X_n = x_{i_n}) = \prod_{i=1}^{n} P(X_i = x_{i_1}), \forall x_{i_j} \in \textit{значения} X_i$$

Доказательство. ($\Rightarrow$) очевидна из определения
($\Leftarrow$)

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in \{x_{11}, \dots, x_{1,k_1}\}, X_n \in \{x_{n1}, \dots, x_{n,k_n}\} = \sum_{j_1 \dots j_n} P(X_1 = x_{1j_1}, \dots, X_n = x_{nj_n}) = \dots = P(X \in B_1) \cdot \dots \cdot P(X_n \in B_n)$$

□

Теорема 6. *Критерий независимости для абсолютно непрерывных с.в.*

$$X_1, \dots, X_n - \text{ нез с.в.} \Leftrightarrow p(x_1 \dots, x_n) = \prod_{i=1}^{n} p_i(x_i), p_i \text{плотность} X_i$$

Доказательство. ($\Rightarrow$) очевидна из определения
($\Leftarrow$)

$$= F(t_1 \dots t_n) = \int\limits_{-\inf}^{t_1} \cdots \int\limits_{-\inf}^{t_n} p(x_1 \dots x_n) dx_1 \dots dx_n = \prod_{i=1}^{n} F_i(t_i) = \prod_{i=1}^{n} \int\limits_{-\inf}^{t_i} p_i(x_i) dx_i$$

То есть мы получили, что совместная функция распределения равна произведению одномерных. В общем-то ровно то, что и хотели. □

Пример. $X_1, X_2 \sim Bern(p)$ независимы

$$X_1 + X_2 \sim Bin(2, p)$$

Доказательство.

$$\begin{aligned} P(X_1 + X_2 = k) &= P(X_1 + X_2 = k | X_2 = 0) + P(X_1 + X_2 = k | X_2 = 1) \\ &= P(X_1 = k | X_2 = 0) + P(X_1 = k - 1 | X_2 = 1) \\ &= \begin{cases} (1-p)^2 & k = 0 \\ 2p(1-p) & k = 1 \\ p^2 & k = 2 \end{cases} \end{aligned}$$

Тогда, если $X_1 \dots X_n$ - $Bern$ независимы и одинаковы распределены (i.i.d)

$$S_n = X_1 + \dots + X_n \sim Bin(n, p)$$

Доказывается по индукции: Пусть $X_1 \dots X_{n-1} \sim Bin(n-1, p)$

$$P(S_n = n) = P(S_{n-1} + X_n = n) = P(S_{n-1} = n | X_n = 0) + P(S_{n-1} = n - 1 | X_n = 1) = p^n$$

$$P(S_n = 0) = P(S_{n-1} + X_n = 0) = P(S_{n-1} = 0 | X_n = 0) = (1-p)^n$$

$$\begin{aligned} P(S_n = k) &= P(S_{n-1} + X_n = k) = P(S_{n-1} = k | X_n = 0) + P(S_{n-1} = k - 1 | X_n = 1) \\ &= \binom{n-1}{k-1} p^k (1-p)^{n-k} + \binom{n-1}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

□

Пример. $X_1 \sim Pois(\lambda_1), X_2 \sim Pois(\lambda_2)$ – независимы

$$\begin{aligned}
P(X_1 + X_2 = k) &= \sum_{i=0}^{k} P(X_1 = k-i) \cdot P(X_2 = i) \\
&= \sum_{i=0}^{k} e^{-\lambda_1} \cdot \frac{\lambda_1^{k-i}}{(k-i)!} \cdot e^{-\lambda_2} \frac{-\lambda_2^i}{i!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^{k-i} \lambda_2 i \\
&= \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1+\lambda_2)^k}{k!} \\
&\Rightarrow X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)
\end{aligned}$$

Пример. $X_1, X_2 \sim Geom(p)$ – нез. $k \in \mathbb{Z}_+$ (количество неудач до первого успеха).

$$P(X_1+X_2 = k) = \sum_{i=0}^{k} P(X+1 = i) P(X_2 = k-i) = \sum_{i=0}^{k} p(1-p)^i p(1-p)^{k-i} = (k+1)p^2(1-p)^k \sim NB(2,p)$$

Пример. Общая схема для непрерывных величин.
X, Y - нез. $p_{X,Y} = p_X(x) p_Y(y)$

$$P(X+Y \le t) = \iint\limits_{x+y \le t} p_X(x) p_Y(y) dx dy = \int\limits_{-\infty}^{+\infty} p_X(x) \int\limits_{-\infty}^{t-x} p_Y(y) dy dx = \int\limits_{-\infty}^{+\infty} p_X(x) p_Y(t-x) dx$$

Пример. $X_1, X_2 \sim Exp(\lambda)$ – независимы

$$p_x = \lambda e^{-\lambda x} \cdot \mathbb{1}(x \ge 0)$$

$$p_{X+Y}(t) = \int\limits_{0}^{+\infty} \lambda^2 e^{-\lambda x} e^{-\lambda(t-x)} \mathbb{1} dx (t-x \ge 0) = \lambda^2 \int\limits_{0}^{t} e^{-\lambda t} dx = \lambda^2 t e^{-\lambda t} \sim \Gamma(2, \lambda)$$

Аналогично:

$$X_1 \dots X_n \sim Exp(\lambda) i.i.d \Rightarrow \sum_{k=1}^{n} X_k \sim \Gamma(n, \lambda)$$

Определение 16. $(X_1 \dots X_n) \sim$ нез $N(0,1)$ – стандартный гауссовский вектор

$$\Leftrightarrow p(x_1 \dots x_n) = \frac{1}{\sqrt{2\pi}^n} \exp -\frac{1}{2} \sum_{i=1}^{n} x_i^2 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Пример. $X, Y \sim$ нез $N(0,1)$

$$\begin{aligned}
P_{X+Y}(t) &= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2} - \frac{(t-x)^2}{2}\right) dx = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \exp -\frac{1}{2}(2x^2 - 2tx + t^2) dx = \left[y = x - \frac{t}{2} \right] \\
&= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \exp -\frac{1}{2}(2y^2 + \frac{t^2}{2}) dy = \frac{1}{\sqrt{2\pi 2}} e^{-\frac{t^2}{4}} \underbrace{\int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{\frac{-y^2}{1}} dy}_{1} = \frac{1}{\sqrt{4\pi}} e^{-\frac{t^2}{4}} \sim N(0,2)
\end{aligned}$$

В общем случае: $X \sim N(\nu_1, \sigma_1^2), Y \sim N(\nu_2, \sigma_2^2)$ нез $\Rightarrow X + Y \sim N(\nu_1 + \nu_2, \sigma_1^2 + \sigma_2^2)$

2.4 Об интегралах

$(\Omega; \mathfrak{A}; P)$ – вероятностное пространство. X – случайный вектор, P_X распределение с.в

$$P_X(B) = P(X \in B) = P(\{\omega : X(w) \in B\})$$

$g : \mathbb{R}^n \to \mathbb{R}$, измерима

$$\int\limits_{\Omega} g(X(\omega))P(d\omega) = \int\limits_{\mathbb{R}^n} g(x)P_X(dx) = \begin{cases} \sum\limits_k g(x_k)P(X = x_k) & X\text{- дискретная} \\ \int g(x)p(x)dx & X\text{− непрерывная} \end{cases}$$

Теорема 7. *Теорема Фубини X, Y – независимые св, $g : R^2 \to R$, изм. Тогда*

$$\iint g(x,y)P_{x,y}(dx,dy) = \int \left[\int g(x,y)dF_X(x)\right] dF_Y(y) = \int \left[\int g(x,y)dF_Y(y)\right] dF_X(x)$$

Пример. X, Y – нез

$$P(X + Y \le t) = \iint\limits_{x+y\le t} P_X(dx)P_Y(dy) = \int\limits_{\mathbb{R}} \left[\int\limits_{y\le t-x} P_Y(dy)\right] dF_X(x) = \int\limits_{\mathbb{R}} F_Y(t-x)dF_X(x)\text{- свертка}$$

X, Y – непр: $p(t) = \int\limits_{\mathbb{R}} P_Y(t-x)P_X(x)dx$

3 Характеристики случайных величин

3.1 Матожидание

Определение 1. X - с.в.

$$EX = \int\limits_{\mathbb{R}} xdF(x) = \begin{cases} \sum x_k P(X = x_k) & X \text{ – дискретная} \\ \int\limits_{\mathbb{R}} xp(x)dx & X \text{ – непрерывная} \end{cases}$$

Свойство 1.1. *$g : R^n \to R, X_1 \ldots X_n$ - св*

$$Eg(X_1 \ldots X_n) = \int\limits_{\mathbb{R}} ydF_Y(y) = \int\limits_{R^n} g(x_1 \ldots x_n)P_{X_1 \ldots X_n}(dx_1 \ldots dx_n)$$

Свойство 1.2. $E(c_1X_1 + c_2X_2) = c_1EX_1 + c_2EX_2$

Свойство 1.3. X, Y**- *нез*** $\Rightarrow EXY = EX \cdot EY$

Свойство 1.4. $P(X \ge 0) = 1 \Rightarrow EX \ge 0$

Теорема 1. *Неравенство Маркова*

$$X \ge 0, \exists EX \Rightarrow P(X \ge c) \le \frac{EX}{c}, \forall c > 0$$

Доказательство.

$$P(X \le c) = \int\limits_{x\ge c} dF(x) \le \int\limits_{x\ge c} \frac{x}{c} \le \frac{1}{c}\int\limits_{\mathbb{R}} xdF(x) = \frac{1}{c}EX$$

□

Замечание.

$$E\mathbb{1}(X \in B) = P(X \in B)$$

Свойство 1.5. $X \geq 0, EX = 0 \Rightarrow P(x = 0) = 1$

Доказательство.

$$1 = P(X \geq 0) = P(X \geq c) + P(X < c) \leq 0 + P(X < c) \Rightarrow P(X < c) = 1, \forall c > 0$$

□

Свойство 1.6. *X - дискретная. $\mathbb{N}$ – носитель X, $\Rightarrow EX = \sum\limits_{k=1}^{\infty} P(X \geq k)$*

Доказательство. $\sqsupset p_i = P(X = i)$

$$EX = p_1 + 2p_2 + 3p_3 \dots = (p_1 + p_2 + \dots) + (p_2 + p_3 + \dots) + \dots = \sum_{k=1}^{\infty} P(X \geq k)$$

□

Свойство 1.7. *X - непр, $X \geq 0 \Rightarrow \int\limits_0^{+\infty} P(X \geq r)dr$*

Пример. $\sqsupset P(X = 1000) = 0.01, P(X = 10) = 0.99 \Rightarrow EX = 9.9{+}10 = 19.9$ Этот пример должен показывать, что иногда реальное распределение может быть не очень "хорошим и "среднее"aka матожидание не очень хорошо будет отображать реальное положение дел.

3.2 Дисперсия

Определение 2. X - с.в.

$$\operatorname{Var} X = DX = E(X - EX)^2 \text{– дисперсия}$$

Свойство 2.1. $\operatorname{Var} X \geq 0$

Свойство 2.2.

$$E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = EX^2 - 2(EX)^2 + EX^2 = EX^2 - (EX)^2$$

Свойство 2.3.

$$\operatorname{Var}(aX) = a^2 \operatorname{Var} X$$

Свойство 2.4.

$$\begin{aligned}\operatorname{Var}(X \pm Y)X = E(X \pm Y)^2 - (E(X \pm Y))^2 &= EX^2 \pm 2XY + EY^2 - (EX)^2 \mp 2EXEY - (EY)^2 \\ &= \operatorname{Var} X + \operatorname{Var} Y \pm 2(EXY - EX \cdot EY\end{aligned}$$

Свойство 2.5. *X, Y – нез* $\Rightarrow \operatorname{Var}(X \pm Y) = \operatorname{Var} X + \operatorname{Var} Y$

Свойство 2.6. $\operatorname{Var} c = 0$

Теорема 2. *Неравенство Чебышева X - с.в, $\exists EX, \operatorname{Var} X \Rightarrow P(|X - EX| \geq \varepsilon) \leq \frac{\operatorname{Var} X}{\varepsilon^2}$*

Доказательство.

$$P(|X - EX| \geq \varepsilon) = P((X - EX)^2 \geq \varepsilon) \leq_{\text{Маркова}} \frac{E(X - EX)^2}{\varepsilon^2}$$

□

Замечание.

$$P(|X - EX| \geq 3\sqrt{\operatorname{Var} X}) \leq \frac{1}{9}$$

Свойство 2.7. $\operatorname{Var} X = 0 \Rightarrow P(x = c) = 1$

Доказательство.

$$P(|X - EX| \geq 0) = 1$$

$$P(|X - EX| \geq \varepsilon) + P(|X - EX| \geq \varepsilon) \leq P(|X - EX| \leq \varepsilon) \leq 1$$

□

Свойство 2.8.

$$\operatorname{Var} X = \min_{a \in \mathbb{R}} E(X-a)^2$$

Доказательство.

$$f(a) = EX^2 - 2aEX + a^2 \Leftrightarrow a_* = \frac{2EX}{2} = EX - \min$$

□

Свойство 2.9. $\operatorname{Var}(X + c) = \operatorname{Var} X$

Пример. $I_C : P(X = c) = 1; EX = c, \operatorname{Var} X = 0$

Пример. $Bern(p)$

$$EX = 1p + 0(1-p) = p$$

$$\operatorname{Var} X = EX^2 - (EX)^2 = (1^2 p + 0^2(1-p)) - p^2 = p(1-p) = pq$$

Пример. $Bin(n, p), S_n \sim Bin(n, p), X_1 \dots X_n - -\text{i.i.d.} Bern(p)$

$$ES_n = \sum EX_i = np$$

$$\operatorname{Var} S_n = np(1-p)$$

Пример. $Pois(\lambda)$

$$EX = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

$$E(X(X-1)) = EX^2 - EX = EX^2 - (EX)^2 + (EX)^2 - EX = \operatorname{Var} X + (EX)^2 - EX$$
$$= \sum_{k=2}^{\infty} k(k-1) \frac{e^{-\lambda}\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2} e^{-\lambda}}{(k-2)!} = \lambda^2 \Rightarrow \operatorname{Var} X = \lambda$$

Пример. $X_1, X_2 \sim Geom(p)$ – нез. $k \in \mathbb{Z}_+$ (количество неудач до первого успеха).$EX = \sim_{k=1}^{\infty} P(x \geq k)$

$$P(X \geq k) = \sum_{i=k}^{\infty} (1-p)p = p \frac{(1-p)^k}{1-(1-p)} = (1-p)^k$$

$$E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1)(1-p)^(k-2)p = (1-p)^2 p \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = (1-p)^2 p \sum_{k=2}^{\infty} \frac{\delta^2 (1-p)^k}{\delta p^2}$$
$$= (1-p)^2 p \frac{\delta^2}{\delta p^2} (\sum_{k=2}^{\infty} (1-p)^k) = (1-p)^2 p \frac{\delta^2}{\delta p^2} (\frac{(1-p)^2}{p}) = 2\frac{(1-p)^2}{p^2} \Rightarrow \operatorname{Var} X = \frac{1-p}{p}$$

Если носитель $\mathbb{N} : Y = X + 1, EY = \frac{(1-p)}{p} + 1 = \frac{1}{p}$

Пример. $U[a,b] : X \sim U[a,b], Y \sim U[0,1] \Rightarrow X = (b-a)Y + a$

$$EY = \int\limits_0^1 x dx = \frac{1}{2}$$

$$EY^2 = \int\limits_0^1 x^2 dx = \frac{1}{3} \Rightarrow VarY = \frac{1}{3} - \frac{1}{4} = frac112$$

$$EX = (b-a)\frac{1}{2} + a = \frac{a+b}{2}; \operatorname{Var} X = \frac{(b-a)^2}{12}$$

Пример. $N(\mu, \sigma^2), Y \sim N(0,1), X \sim N(\mu, \sigma^2) \Rightarrow X = \sigma Y + \mu$

$$EY = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} x e^{-x^2/2} dx = 0$$

$$EY^2 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx = 1$$

$$EY = 0, \operatorname{Var} Y = 1 \Rightarrow EX = \mu, \operatorname{Var} X = \sigma^2$$

Оценим $3 - \sigma$ отклонение:

$$P(|X - EX| > 3\sqrt{\operatorname{Var} X}) = P(|X - \mu| > 3\sigma) = P(|\frac{X-\mu}{\sigma}| > 3) = 1 - P(|\frac{X-\mu}{\sigma}| < 3)$$
$$= 1 - \Phi(3) + \Phi(-3) = 2(1 - \Phi(3)) \approx 0.00135$$

Пример. $Exp(\lambda), p(x) = \lambda e^{-\lambda x} \mathbb{1}(x \geq 0)$

$$EX = \int\limits_0^{+\infty} x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^{+\infty} + \frac{1}{\lambda} \int\limits_0^{+\infty} \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$EX^2 = \int\limits_0^{+\infty} x^2 \lambda e^{-\lambda x} = -x^2 e^{-\lambda x}\Big|_0^{+\infty} + \frac{2}{\lambda} \int\limits_0^{+\infty} x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$\operatorname{Var} X = \frac{1}{\lambda^2}$$

Пример. $X_1 \dots X_n - -i.i.d. X_i \sim Exp(\lambda), S_n = X_1 + \dots + X_n \sim \Gamma(n, \lambda)$

$$ES_n = \frac{n}{\lambda}; \operatorname{Var} S_n = \frac{n}{\lambda^2}$$

Пример. $Cauchy(0,1)$; $\nexists EX$

$$EX = \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} \frac{x dx}{1 + x^2}$$

3.3 Другие характеристики случайных величин

3.3.1 Моменты

Определение 3. $k \in \mathbb{N}$.
Момент порядка $k : EX^k$.
Абсолютный момент порядка $k : E|X|^k$
Центральный момент порядка $k : E(X - EX)^k$
Абсолютный центральный момент порядка $k : E|X - EX|^k$

3.3.2 Коэффиценты ассиметрии/эксцесса

Определение 4. Коэффицент ассиметрии: $\frac{E(X-EX)^3}{\sigma^3}, \sigma = \sqrt{\operatorname{Var} X}$ Если он > 0, то правый хвост длиннее левого. Если < 0, то наоброт.

Определение 5. Коэффицент эксцесса: $\frac{E(X-EX)^4}{\sigma^4} - 3$.
Если он > 0, то пик более крутой. Если < 0, то более пологий.

3.3.3 Мода

Определение 6. Дискретный случай: $x_k^* : x_k^* = arg\max_{x_k} P(X = x_k)$ Непрерывный случай: $x^* : x_k^* = arg\max_{x\in\mathbb{R}} p(x)$

Замечание. Но с модами бывают путаницу, потому что бывают **унимодальные** и **многомодальные** распределения. И если у первых один и глобальный, и минимальный максимум, то у вторых может быть и несколько локальных (да и глобальных) максимумов.

3.3.4 Квантиль

Определение 7. $\alpha \in (0,1)$ q_α - квантиль порядка α распределения $P_X \Leftrightarrow P(X \geq q_\alpha) \geq 1-\alpha, P(X \leq q_\alpha) \geq \alpha$

Замечание. В дискретном случае квантиль определяется неоднозначно.

$$q_\alpha = \sup_{x\in\mathbb{R}}\{F(x) \leq \alpha\} = \inf_{x\in\mathbb{R}}\{F(x) \geq \alpha\}$$

Замечание. Важное замечание, что в непрерывных фукнциях, площадь под графиком плотности слева от q_α является α, а справа $1-\alpha$

Определение 8. Квантиль при $\alpha = \frac{1}{2}$ называется медианой.

Теорема 3. $m = med(X), \forall a \exists E|x-a| \Leftarrow m = arg\min_{a\in\mathbb{R}} E|x-a|$

Доказательство. Не уменьшая общности, $m = 0$, иначе введем $X' = X - m$.

$$E|X-c| \geq^? E|X|$$

$$> 0 : |X-c| - |X| = \begin{cases} -c & X > c \\ c-2X & 0 < X \leq c \\ c & X \leq 0 \end{cases}$$

$$\begin{aligned} \Rightarrow E(|X-c|-|X|) &= E(|X-c|-|X|)\mathbb{1}(X \leq 0) + E(|X-c|-|X|) = E(|X-c|-|X|)\mathbb{1}(X > 0) \\ &\geq cE\mathbb{1}(X \leq 0) - cE\mathbb{1}(X > 0) \\ &= c(2P(X \leq 0) - 1) \geq 0 \end{aligned}$$

$$c < 0 : X'' = -X, c'' = -c$$

□

Пример. $X \sim N(0,1), Y = X^2. X, Y$ – не независимы

$$P(X \leq 1, Y \leq 1) = P(X \leq 1, -1 \leq X \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1$$

$$P(X \leq 1) = \Phi(1), P(Y \leq 1) = 2\Phi(1) - 1$$

Но:

$$\Phi(1)(2\Phi(1)-1) \neq 2\Phi(1) - 1$$

$$EX = 0, EY = EX^2 = 1, EXY = EX^3 = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} x^3 e^{-x^2/2} dx = 0$$

3.4 Характеристики нескольких случайных величин

3.4.1 Ковариация

Определение 9. Ковариация

$$\operatorname{cov}(X,Y) = E(X-EX)(Y-EY)$$

Свойство 9.1. $\operatorname{cov}(X,Y) = E(XY-(EX)Y-X(EY)+EXEY) = E(XY)-EXEY$

$$\operatorname{Var}(X\pm Y) = \operatorname{Var}X + \operatorname{Var}Y \pm 2\operatorname{cov}(X,Y)$$

$$\operatorname{Var}(X_1+\dots+X_n) = \sum_{i=1}^{n}\operatorname{Var}X_n + 2\sum_{i<j}\operatorname{cov}(X_i,X_j)$$

Свойство 9.2. $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$

Свойство 9.3. $\operatorname{cov}(X,X) = \operatorname{Var}X$

Свойство 9.4. $\operatorname{cov}(c_1X_1+c_2X_2,Y) = c_1\operatorname{cov}(X_1,Y)+c_2\operatorname{cov}(X_2,Y)$

Свойство 9.5. $\operatorname{cov}(X,c) = 0$

Свойство 9.6. $\operatorname{cov}(X+c,Y) = \operatorname{cov}(X,Y)$

Свойство 9.7. X,Y – *нез* $\Rightarrow \operatorname{cov}(X,Y) = 0$

3.4.2 Корреляция

Определение 10. Коэффициент корреляции

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{Var}X\cdot\operatorname{Var}Y}}$$

Свойство 10.1. $|\rho(X,Y)|\le 1$

$$\rho(X,Y) = \frac{\operatorname{cov}(X-EX,Y-EY)}{\sqrt{\operatorname{Var}X\cdot\operatorname{Var}Y}} = \operatorname{cov}\frac{X-EX}{\sqrt{\operatorname{Var}X}},\frac{Y-EY}{\sqrt{\operatorname{Var}Y}} = \operatorname{cov}(\widetilde{X},\widetilde{Y})$$

$$E\widetilde{X}=0;\operatorname{Var}\widetilde{X}=1 \Rightarrow \operatorname{Var}(\widetilde{X}+\widetilde{Y}) = \operatorname{Var}\widetilde{X}+\operatorname{Var}\widetilde{Y}+2\operatorname{cov}(\widetilde{X},\widetilde{Y}) \Rightarrow \rho(X,Y)\ge -1$$
$$\operatorname{Var}(\widetilde{X}-\widetilde{Y}) = 2-2\rho(X,Y)\ge 0 \Rightarrow \rho(X,Y)\le 1$$

Свойство 10.2. $X, Y = aX+b$

$$\rho(X,Y) = \frac{\operatorname{cov}(aX+b,X)}{\sqrt{\operatorname{Var}(aX+b)\operatorname{Var}X}} = \frac{a\operatorname{cov}(X,X)}{|a|\operatorname{Var}X} = \operatorname{sign}a$$

Свойство 10.3. $\rho(X,Y)=1$

$$\operatorname{Var}(\widetilde{X}-\widetilde{Y}) = 2-2=0 \Rightarrow \widetilde{X}-\widetilde{Y} = C \Rightarrow \frac{Y-EY}{\sqrt{\operatorname{Var}Y}} = \frac{X-EX}{\sqrt{\operatorname{Var}X}} - C$$

Свойство 10.4. $\rho(X,Y)=-1$

$$\operatorname{Var}(\widetilde{X}+\widetilde{Y}) = 2-2=0 \Rightarrow \widetilde{X}+\widetilde{Y} = C \Rightarrow \frac{Y-EY}{\sqrt{\operatorname{Var}Y}} = -\frac{X-EX}{\sqrt{\operatorname{Var}X}} - C$$

Свойство 10.5. $|\rho(X,Y)|=1 \Leftrightarrow Y=aX+b, a\ne 0$

$$\begin{cases} a>0 \Leftrightarrow & p(X,Y)=1\\ a<0 \Leftrightarrow & p(X,Y)=-1\end{cases}$$

3.5 Характеристики случайного вектора

Определение 11. $X = (X_1 \dots X_n)$ – случайный вектор.
$EX = (EX_1, \dots EX_n)$ –вектор матожиданий
$\operatorname{Var} X = E(X - EX)(X - EX)^T = (\operatorname{cov}(X_i, X_j)_{i,j}$ – матрица ковариаций.

Свойство 11.1. $EAX = AEX, A \in \mathbb{M}_{m\times n}(\mathbb{R})$

Свойство 11.2. $\operatorname{Var} X = (\operatorname{Var} X)^T$

Свойство 11.3. $\operatorname{Var} X \geq 0$ – *неотрицательная определенность*

Доказательство.

$$t^T \operatorname{Var} X t = t^T (E(X - EX)(X - EX)^T) t = Et^T (X - EX)(X - EX)^T t = E(t^T (X - EX))^2 \geq 0$$

□

Свойство 11.4. $\operatorname{Var} AX = EA(X - EX)(X - EX)^T A^T = A \operatorname{Var} X A^T$

Свойство 11.5. $\operatorname{Var}(X + c) = \operatorname{Var} X$

Свойство 11.6. X, Y - *нез* $\Rightarrow \operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$

Доказательство. Как выглядит матрица ковариаций?

$$i = j \Rightarrow \operatorname{cov}(X_i + Y_i, X_i + Y_i) = \operatorname{Var} X_i + \operatorname{Var} Y_i$$
$$i \neq j \Rightarrow \operatorname{cov}(X_i + Y_i, X_j + Y_j) = \operatorname{cov}(X_i, X_j) + \underbrace{\operatorname{cov}(Y_i, X_j)}_{0} + \underbrace{\operatorname{cov}(X_i, Y_j)}_{0} + \operatorname{cov}(Y_i, Y_j)$$

□

3.5.1 Полиномиальное распределение

Определение 12. $Poly(n, p), p = (p_1 \dots p_m)$ – полиномиальное распределение

$$P(S_n^{(1)} = k_1, \dots, S_n^{(m)} = k_m = p_1^{k_1} \cdot \dots \cdot p_m^{k_m} = \frac{n!}{k_1! \cdot \because k_m!}$$

Пример. $S_n = \sum\limits_{i=1}^{n} X_i, X_i$– нез, $X_i \sim Poly(1, p)$

$$ES_n = nEX_1 = [EX_1^{(j)} = p_j] = np$$

$$\operatorname{Var}_n = n \operatorname{Var}(X_i), \operatorname{Var}(X_i) = (\operatorname{cov}(X_i^{(k)}, X_i^{(j)}))_{1 \leq k,j \leq m}$$

$$k = j \Rightarrow \operatorname{Var} X_i^{(k)} = E(X_i^{(k)})^2 - (EX_i^{(k)})^2 = p_j - p_k^2 = p_k(1 - p_k)$$
$$k \neq j \Rightarrow \operatorname{cov}(X_i^{(k)}, X_i^{(j)}) = \underbrace{EX_i^{(k)} X_i^{(j)}}_{0} - EX_i^{(k)} \cdot EX_i^{(j)} = -p_j p_j$$

3.5.2 Многомерное нормальное распределение

Пример.

$$X = (X_1\dot{X}_n), X_i\text{– нез}, X_i \sim N(0,1), p(x) = \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{||x_i||^2}{2}\right) \Rightarrow EX = 0, \operatorname{Var} X = I_n$$

$$Y = AX_\mu, Y \sim N(\mu, AA^T) \Rightarrow EY = \mu, \operatorname{Var} Y = \operatorname{Var} AX = AI_nA_T = AA_T$$

$$U \sim N(\mu, \Sigma), \Sigma = \Sigma^T, \Sigma \geq 0 \Rightarrow U = \sqrt{\Sigma}X + \mu$$

Утверждение. $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \Rightarrow \begin{cases} X \sim N(\mu_1, \Sigma_1 1) \\ Y \sim N(\mu_2, \Sigma_2 2) \end{cases}$

Доказательство.

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

Причем имеет место следующие соотношения (так как U ортогональна

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} U_{11}^T & U_{21}^T \\ U_{12}^T & U_{22}^T \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

$$\begin{pmatrix} U_{11}^T & U_{21}^T \\ U_{12}^T & U_{22}^T \end{pmatrix} \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

Тогда можем представить X, Y вот так:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} \sqrt{\Lambda_1} & 0 \\ 0 & \sqrt{\Lambda_2} \end{pmatrix} \begin{pmatrix} U_{11}^T & U_{21}^T \\ U_{12}^T & U_{22}^T \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

Если расписать это отдельно для X (для Y аналогично), то получится, что это линейная комбинация ξ_1, ξ_2:

$$X = \mu_1 + \begin{pmatrix} U_{11}\sqrt{\Lambda_1} & U_{12}\sqrt{\Lambda_2} \\ U_{21}\sqrt{\Lambda_1} & U_{22}\sqrt{\Lambda_2} \end{pmatrix} \begin{pmatrix} U_{11}^T\xi_1 & U_{21}^T\xi_2 \\ U_{12}^T\xi_1 & U_{22}^T\xi_2 \end{pmatrix} = \mu_1 + U_{11}\sqrt{\Lambda_1}U_{11}^T\xi_1 + \cdots + U_{22}\sqrt{\Lambda_2}U_{22}^T\xi_2$$

Дальше уже чисто техническая работа: сгруппировать слагаемые и показать, что матрица $\operatorname{Var} X = \Sigma_{11}$ и аналогично для Y. □

Определение 13. Некоррелированность X, Y некореллированны $\Leftrightarrow \rho(X,Y) = \operatorname{cov}(X,Y) = 0$

Утверждение. $X = (X_1 \dots X_n) \sim N(\mu, \Sigma)$ *тогда* X_i, X_j*– нез* $\Leftrightarrow \operatorname{cov}(X_i, X_j) = 0$

Доказательство. ($\Rightarrow$) уже знаем
($\Leftarrow$) рассмотрим $\begin{pmatrix} X_i \\ X_j \end{pmatrix}$

$$\begin{pmatrix} X_i \\ X_j \end{pmatrix} \sim N(\begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \sigma_{x_i}^2 & 0 \\ 0 & \sigma_{x_j}^2 \end{pmatrix})$$

$$\sigma_{x_i} = \sigma_{x_j} = 0 \Rightarrow \text{Оба вырожденны} \Rightarrow \text{они независимы}$$

$$\sigma_{x_i}, \sigma_{x_j} > 0 : p(x_i, x_j) = \frac{1}{2\pi\sqrt{\sigma_{x_i}^2\sigma_{x_j}^2}}\exp(-\frac{1}{2}\begin{pmatrix} x_i - \mu_i \\ x_j - \mu_j \end{pmatrix}^T \begin{pmatrix} \sigma_{x_i}^2 & 0 \\ 0 & \sigma_{x_j}^2 \end{pmatrix} \begin{pmatrix} x_i - \mu_i \\ x_j - \mu_j \end{pmatrix}) = p_{X_i}(x_i) \cdot p_{X_j}(x_j)$$

В случае, если только одно ровно нулю, то все тоже хорошо получится. □

3.6 Вероятностные неравенства

Неравенство Маркова: $X \geq 0, \exists EX \Rightarrow P(X \geq c) \leq \frac{EX}{c}, \forall c > 0$
Неравенство Чебышева: $\exists EX, \operatorname{Var} X \Rightarrow P(|X - EX| \geq \varepsilon) \leq \frac{\operatorname{Var} X}{\varepsilon^2}$

Утверждение. *Слабый закон больших чисел*

$$X_1 \ldots X_n - -i.i.d, \mu, \sigma^2, \widetilde{X} = \frac{\sum\limits_{i=1}^{n} x_i}{n} \Rightarrow P(|\widetilde{X} - \mu| \geq \varepsilon) \xrightarrow{n\to\infty} 0, \forall \varepsilon \geq 0$$

Доказательство. $P(|\widetilde{X} - \mu| \geq \epsilon) \leq \frac{\operatorname{Var} X}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$ □

Определение 14. Вероятностное КБШ $L_2 = \{X : EX^2 < \infty\}$
Рассмотрим функцию: $(\ldots, \ldots) : L2 \times L2 \to \mathbb{R}.(X, Y) = EXY$
Это очень похоже на скалярное произведение, а для него у нас есть КБШ: $|EXY| \leq EX^2 \cdot EY^2$. Что неправда, а значит не скалярное произведение. Но вот если рассмотрим функцию $(X, Y) = E(X - EX)(Y - EY)$, она линейна, аргументы перестановочны, ну и $(X, X) \geq 0$, а значит это скалярное произведение. Тогда получаем свойство $\operatorname{cov}^2(X, Y) \leq \operatorname{Var} X \cdot \operatorname{Var} Y$

Определение 15. Вероятное неравенство Гёльдера $E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}, pq = 1$

Определение 16. Вероятностное неравенство Минковского $(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$

Определение 17. Вероятностное неравенство Йенсенна $g-$ выпукла вниз $\Rightarrow g(EX) \leq Eg(X)$

Определение 18. Неравенство Ляпунова $(E|X|^p)^(1/p) \leq (E|X|^q)^{1/q}, p < q$

Замечание. Из неравенства Ляпунова следует, что если существует k-ый момент, то существуют и все предыдущие.

3.7 Условные распределения. Условные математические ожидания и дисперсии

Определение 19. Условное распределение
$X_1 \ldots X_n-$ дискретные :

$$P(X_1 = x_1, \ldots X_n = x_n | X_{k+1} = x_{k+1} \ldots X_n = x_n) = \frac{P(X_1 = x_1, \ldots X_n = x_n)}{P(X_{k+1} = x_{k+1} \ldots X_n = x_n)}$$

$X_1 \ldots X_n-$ непрерывные :

$$p(x_1, \ldots x_n | x_{k+1}, \ldots x_n) = \frac{p_{X_1 \ldots X_n}(x_1, \ldots x_n)}{p_{X_{k+1} \ldots X_n}(x_{k+1}, \ldots x_n)}$$

Пример. $(X, Y)-$ диск : $P(X = k) = \sum\limits_j P(X = k | Y = y_j) \cdot P(Y = y_j)$
$(X, Y)-$ непр : $P_X(x) = \int p(x|y) \cdot p(y) dy$

3.7.1 Условное матожидание

Определение 20. Условное матожидание
Дискретные: $E(X|Y) = \sum x_k p(X = x_k | Y)$
Непрерывные: $E(X|Y) = \int xp(x|y) dx$

Замечание. Обычное матожидание минимизирует: $EX = arg \min_a E(X - a)^2$
Условное: $E(X|Y) = arg \min_{f(Y) \in L_2} E(X - f(Y))^2$
То есть идет минимизация по функции.

Замечание. Давайте рассмотрим $M = \{f(Y) : Ef^2(Y) < \infty\}$
Геометрической интерпретацией матожидания $E(X|Y)$ будет ортогональная проекция X на M.

Свойство 20.1. $E(X|X) = X$

Свойство 20.2. X, Y *– нез* $\Rightarrow E(X|Y) = EX$

Свойство 20.3. *Формула полной вероятности для матожидания:* $EX = E_Y[E(X|Y)]$

3.7.2 Условная дисперсия

Определение 21. Условная дисперсия

$$\operatorname{Var}(X|Y) = E(X - E(X|Y))^2 = E(X^2|Y) - (E(X|Y))^2$$

Свойство 21.1. $\operatorname{Var} Y = E \operatorname{Var}(Y|X) + \operatorname{Var} E(X|Y)$

Доказательство. $\operatorname{Var} Y = EY^2 - (EY)^2 = E\operatorname{Var}(Y|X) + \underbrace{EE^2(Y|X) - E^2E(Y|X)}_{\operatorname{Var} E(X|Y)}$

$$EY^2 = E(E(Y^2|X)) = E(\operatorname{Var}(Y|X) + E^2(Y|X))$$

$$EY = EE(Y|X)$$

□

Пример.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \operatorname{Var} X & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{Var} Y \end{pmatrix}) \Rightarrow E(Y|X) = \mu_2 + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}(X - \mu_1)$$

Доказательство.

$$U = Y - \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}(X - \mu_1)$$

$$\operatorname{cov}(U, X) = \operatorname{cov}(X,Y) - \operatorname{cov}(X,Y) = 0 \Rightarrow U, Y - \text{нез}$$

$$\begin{aligned} E(Y|X) &= E(Y - \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}X + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}X|X) \\ &= E(U|X) + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}E(X|X) = EU + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}X \\ &= \underbrace{\mu_2}_{EY} + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}(X - \underbrace{\mu_1}_{EX}) \end{aligned}$$

□

3.8 Введение в методы Монте-Карло, ЦПТ

Пусть мы хотим вычислить какой-то интеграл:

$$\int\limits_A f(x_1 \dots x_n) dx_1 \dots dx_n = \frac{\mu A}{\mu A} \int\limits_A f(x_1 \dots x_n) dx_1 \dots dx_n = \mu A \cdot E_{U(A)} f(X_1 \dots X_n)$$

$$\Rightarrow I_n = \frac{1}{n} \sum_{i=1}^{n} f(X_{i1}, \dots X_{in}), (X_{i1}, \dots X_{in}) - -i.i.d \sim U(A)$$

Тогда по закону больших чисел $I_n \approx Ef(X_1 \dots X_n)$ Ну и в целом методы, когда мы генерируем много каких-то случайных величин, аггрегируя результат, называются **методами Монте-Карло**.

Замечание. У нас уже была теорема Муавра-Лапласа, которая утверждала, что

$$S_n = \sum_{i=1}^{n} X_i, X_i - -i.i.d, X_i \sim Bern(p)$$

$$\sup_{x\in\mathbb{R}} |P(\frac{S_n - np}{\sqrt{npq}} \le x) - \Phi(x)| \underset{n\to\infty}{\longrightarrow} 0$$

Теорема 4. *Центральная предельная теорема. ЦПТ*

$$S_n = \sum_{i=1}^{n} X_i, X_i - -i.i.d, \exists EX, \operatorname{Var} X$$

$$\sup_{x\in\mathbb{R}} |P(\frac{S_n - nEX_1}{\sqrt{n \operatorname{Var} X}} \le x) - \Phi(x)| \underset{n\to\infty}{\longrightarrow} 0$$

3.9 Сходимости

Определение 22. почти наверное
$X_1 \dots X_n$ – с.в, которые заданы на $(\Omega, \mathfrak{A}, P)$

$$X_n \xrightarrow{a.s.} X \Leftrightarrow \begin{cases} P(\{w : \lim\limits_{n\to\infty} X_n(w) = X(w)\}) = 1 \\ P(\{w : \lim\limits_{n\to\infty} X_n(w) \neq X(w)\}) = 0 \end{cases}$$

Определение 23. По вероятности. P

$$X_n \xrightarrow[n\to\infty]{P} X \Leftrightarrow \begin{cases} P(|X_n - X| \ge \varepsilon) \underset{n\to\infty}{\longrightarrow} 0, \forall\varepsilon > 0 \\ P(|X_n - X| < \varepsilon) \underset{n\to\infty}{\longrightarrow} 1 \end{cases}$$

Определение 24. В среднем порядка p, L_p

$$X_n \xrightarrow{L_p} X \Leftrightarrow E|X_n - X|^p \underset{n\to\infty}{\longrightarrow} 0$$

Определение 25. По распределению, d $X_n : (\Omega_n, \mathfrak{A}_n, P_n), X : (\Omega, \mathfrak{A}, P)$

$$X_n \xrightarrow{id} X \Leftrightarrow Ef(X_n) \to Ef(X), \forall f - \text{ непр и ограниченной}$$

Теорема 5.

1. *a.e $\Rightarrow$ P из сходимости почти наверное следует сходимость по вероятности*
2. *$L_p \Rightarrow P$ из сходимости в среднем следует сходимость по вероятности*
3. *$P \Rightarrow d$ из сходимости по вероятности следует сходимость по распределению*

Доказательство. $(P \Rightarrow d)$

$$|E(f(X_n) - f(X)| \le |E(f(X_n) - f(X)|\mathbb{1}(|X_n - X| \ge \underbrace{\delta_\varepsilon}_{\text{из непр f}}) + |E\underbrace{|f(X_n) - f(X)|}_{<\varepsilon}\mathbb{1}(|X_n - X| < \underbrace{\delta_\varepsilon}_{\text{из непр f}})$$

$$\le 2MP(|X_n - X| \ge \delta_\varepsilon) + \varepsilon$$

$(L_p \Rightarrow P)$

$$P(|X_n - X| \ge \varepsilon) \le P(|X_n - X|^p \ge \varepsilon^p) \le \frac{E|X_n - X|^p}{\epsilon^p} \to 0$$

$(a.s \Rightarrow P), O = \{\omega : X_n(\omega) \not\to X(\omega)\}, P(O) = 0$

$$\varepsilon > 0, A_n = \bigcup_{m\ge n} \{\omega : |X_m(\omega) - X(\omega)| \ge \varepsilon\}$$

$$A_{n+1} \subset A_n, A = \bigcap A_n \Rightarrow P(A_n) \to P(A)$$
$$\omega \in \overline{O} \Rightarrow \omega \notin A_n \Rightarrow \omega \in \overline{A} \Rightarrow \overline{O} \subset \overline{A} \Rightarrow A \subset O$$
$$P(|X_n - X| \geq \varepsilon) \leq P(A_n)$$

□

Замечание. Никакие другие импликации не верны.

Пример. ($d \not\Rightarrow P$)
Возьмем $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, $X(\omega_i) = (-1)^i$, $X_n = (-1)^{i+n}$

$$E(X) = \frac{f(X(\omega_1)) + f(X(\omega_2))}{2} = \frac{f(1) + f(-1)}{2} = Ef(X_n)$$
$$P(|X_n(w_i) - X(w_i)| > \varepsilon) \not\to 0$$

Пример. ($a.s \not\Rightarrow L_p$), ($P \not\Rightarrow L_p$)

Возьмем пространство: $([0,1], \mathfrak{B}, U[0,1]), X(w) = 0, X_n(w) = \begin{cases} e^n & w \in [0, \frac{1}{n}] \\ 0 & \text{otherwise} \end{cases}$

$$X_n \xrightarrow{a.s.} X; E|X_n - X|^2 = \frac{e^{np}}{n} \not\to_{p>0} 0$$

Пример. ($P \not\Rightarrow a.s.$)
$X_{2^n}(w) = \begin{cases} \frac{1}{2^n} & w \in w \in [0, \frac{1}{2^n}] \\ 0 & \text{otherwise} \end{cases}$

$m, \exists n, 2^n < m < 2^{n+1}, m = 2^n + k \Rightarrow X_m = \begin{cases} \frac{1}{2^n} & w \in w \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \\ 0 & \text{otherwise} \end{cases}$

$$P(|X_m(w) - X(w)| > \varepsilon) \to 0$$

Замечание. Если в предыдущий пример подставить вместо 2^n что-то побольше, например, $n!$, то докажется ещё и ($L_p \not\Rightarrow a.s.$)

Определение 26. F, F_n– функции распределения

$$F_n(X) \rightharpoonup F(x) \Leftrightarrow F_n(x) \to F(x) \forall x \in C(F)$$

Теорема 6.

$$X_n \xrightarrow{d} X \Leftrightarrow F_n \rightharpoonup F$$

Доказательство. ($\Leftarrow$)
$x_0 \in C(F) : f_\varepsilon(t) = \begin{cases} 1, t \leq x_0 \\ \text{линейно} \\ 0, t \geq x_0 + \varepsilon \end{cases}$

$$F_n(x_0) = \int\limits_{-\infty}^{x_0} dF_n(x) = \int\limits_{-\infty}^{x_0} f_\varepsilon(t) dF_n(x) \leq \underbrace{\int\limits_{-\infty}^{+\infty} f_\varepsilon(t) dF_n(x)}_{Ef_\varepsilon(X_n)} \to \underbrace{\int\limits_{-\infty}^{x_0+\varepsilon} f_\varepsilon(t) dF_n(x)}_{Ef_\varepsilon(X)} \leq F(x_0 + \varepsilon)$$

($\Rightarrow$)$f^*_\varepsilon(t) = f_\varepsilon(t + \varepsilon$ □

Замечание. $F_n(x) - F_n(y) \rightharpoonup F(x) - F(y), \forall x, y\ inC(F)$

Замечание. $F \in C(\mathbb{R}) : \sum\limits_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$

Замечание. F_n, F – диск, $x_0, x_1 \dots x_k$ – носитель, тогда:

$$\underbrace{p_{n,k}}_{P(X_n=x_k)} \to \underbrace{p_k}_{P(X=x_k)}$$

Теорема 7. *Свойства сходимости*

1. $x_n \xrightarrow{a.s.} x : |x_n| \leq^{a.e.} Y, EY < \infty \Rightarrow EX_n \to EX$
2. $X_n \xrightarrow{P} x, f \in C(\mathbb{R}) \Rightarrow f(X_n) \xrightarrow{P} f(X)$
3. $X_n - Y_n \xrightarrow{P} 0, \begin{cases} X_n \xrightarrow{P} X & \Rightarrow Y_n \xrightarrow{P} Y \\ Y_n \xrightarrow{P} Y & \Rightarrow X_n \xrightarrow{P} X \end{cases}$
4. $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} 0 \Rightarrow X_n Y_n \xrightarrow{P} 0$
5. $X_n \xrightarrow{d} C \Leftrightarrow X_n \xrightarrow{P} C$
6. $X_n \xrightarrow{d} C, Y_n \xrightarrow{P} c \Leftrightarrow X_n + Y_n \xrightarrow{d} X + c, X_n + Y_n \xrightarrow{d} cX, \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} (c \neq 0)$
7. *Арифметические операции сохраняют сходимость по вероятности.*

Доказательство. Только идеи.
(1) по сути теорема Лебега о мажорированной сходимости
(2) доказывается с помощью определний сходимости по вероятности и непрерывности
... □

3.9.1 Сходимости по множеству функций распределения

Определение 27. $\mathbf{F}$ – множество функций распределений.
$\mathbf{G}$ – множество функций, монотонно возрастающих, непр справа и $G(+\infty) \leq 1, G(-\infty) \geq 0$

Теорема 8. *Хейли*

$$\{G_n\} \in \mathbf{G} \Rightarrow \exists G_{n_k} : G_{n_k} \rightharpoonup G \in \mathbf{G}$$

Следствие 0.1. $\forall G_{n_k} \rightharpoonup G \Rightarrow G_n \rightharpoonup G$

Пример.

$$F_n(T) = \begin{cases} 0 & t \leq -n \\ \frac{1}{2} & -n \leq t < n \\ 1 & t \geq n \end{cases}$$

$$F_n(t) \underset{n\to\infty}{\longrightarrow} \frac{1}{2}$$

Определение 28. плотное распределение

$$\forall \varepsilon > 0, \exists N : \inf_n P(-N \leq X_n \leq N) > 1 - \varepsilon$$

Определение 29. Класс определяющий распределение
$\mathbf{L}$ – какое-то подмножество непрерыных и ограниченных функций
$\mathbf{L}$ определяет распределение $\Leftrightarrow F \in \mathbf{F}, G \in \mathbf{G}, \forall f \in \mathbf{L} : \int f dG = \int f dF \Rightarrow G = F$

Теорема 9. $\{F_n\}_{n=1}^{\infty}, F_n \in \mathbf{F}, \mathbf{L}$ *– опр распределение. Тогда:*

$$\exists \lim F_n = F \in \mathbf{F} \Leftrightarrow \begin{cases} \{F_n\}_{n=1}^{\infty} \text{ плотно} \\ \exists \lim \int f dF_n \forall f \in \mathbf{L} \end{cases}$$

Доказательство. ($\Leftarrow$) очевидно
($\Rightarrow$)$\exists F_{n_k} \rightharpoonup F_1 \in \mathbf{G}$

$$\varepsilon > 0, \exists N : F_n(N) > \underbrace{F_n(N)}_{1 \geq F_n(N) \geq 1-\varepsilon} - \underbrace{F_n(-N)}_{F_n(-N) < \varepsilon} > 1 - \varepsilon,$$

x_0 – точка непрерывности. $F_1, x_0 \geq N : 1 \geq F_n(N) \geq 1 - \varepsilon$.
$x_1 \leq -N : F_n(x_1) < \varepsilon \Rightarrow (F_1(+\infty) = 1), (F_1(-\infty) = 0), 1 \geq F_1(N) \geq 1 - \varepsilon(F_1(x_1) < \varepsilon \Rightarrow F_1 \in \mathbf{F}$
$\lim \int f dF_{n_k} = \int f dF_1 = \int f dF_2 \Rightarrow F_1 = F_2$ – любая подпоследовательность, сходится к одному распределению $\Rightarrow \exists \lim F_n = F$ □

Следствие 0.1. $\mathbf{L}$ *– опр распределение,* $\int f dF_n \to \int f dF$, $\forall f \in \mathbf{L}, F \in \mathbf{G}$
Пусть верно одно из следующих:

1. $\{F_n\}_{n=1}^{\infty}$ *плотно*
2. $F \in \mathbf{F}$
3. $f(x) \equiv 1 \in \mathbf{L}$

Тогда: $F \in \mathbf{F}, F_n \rightharpoonup F$

1. $C^{(k)}(\mathbb{R})$
2. $C_0^{(k)} = \{f : \mathbb{R} \to \mathbb{R}, f \in C^{(k)}(\mathbb{R}), \operatorname{supp} f -$ *компакт*$\}$
   $f \in C^{(k)}(\mathbb{R}) : \exists f_n \in C_0^{(k)} : f_n(x) \to f(x)$
   $\int f_n df \to \int f dF \Rightarrow F = G$
   $\int f_n'' dG \to \int f'' dG$

Доказательство. $F_n(x) \rightharpoonup F(x), x \in C(F)$
$\int f dF_n \underbrace{=}_{\text{из инт по частям}} - \int F_n f' dx \underbrace{\longrightarrow}_{\text{тк А непр пв, то сход пв}} - \int F f' dx = \int f dF$
3)$\{e^{itx}\}_{t \in \mathbb{R}}$ □

4 Характеристические функции

4.1 Определение

Определение 1. X - с.в. $f(t) = Ee^{itX} = Ecos(tX) + iEsin(tX)$ – **характеристическая функция случайной величины**, $f_X(t) = \int e^{itx} dF(x)$

4.2 Свойства

Свойство 1.1. $f(0) = 1$

Свойство 1.2. $|f(t)| \leq E|e^{itx}| = 1$

Свойство 1.3. $Y = aX + b \Rightarrow f_Y(t) = Ee^{itaX+b} = e^{itb} \cdot f_x(at)$

Свойство 1.4. $U = -X \Rightarrow f_U(t) = Ee^{-itX} = \overline{f_X(t)}$ *– комплексное сопряжение*

Свойство 1.5. X, Y *– нез* $\Rightarrow f_{X+Y}(t) = Ee^{itX} \cdot Ee^{itY} = f_X(t) \cdot f_Y(t)$

Свойство 1.6. f *– равномерно непрерывны на* $\mathbb{R}$:

$$|f(t+h) - f(t)| = |Ee^{itX+ihX} - Ee^{it}| \leq E|e^{ihX} - 1| \xrightarrow{h \to 0} 0$$

$\Rightarrow \sup\limits_{t \in \mathbb{R}} |f(t+h) - f(t)|$ *определен равномерно-непрерывно*
$\Rightarrow e^{ihx} \to 1, |e^{ihx} - 1| \leq 2$

Свойство 1.7. $\exists EX^k \Rightarrow \exists f^{(k)}(c)$

Свойство 1.8. $f(t) = Ee^{itX}, f'(t) = EiXe^{itX}$

Свойство 1.9. $f'0 = iEX$

Свойство 1.10. $\forall t_1 \dots t_m \in \mathbb{R}, \lambda_1 \dots \lambda_m \in \mathbb{C} : \sum\limits_{k,j} f(t_k - t_j)\lambda_k \overline{\lambda_j} \geq 0$

Теорема 1. *Бохнера-Хинчина*

$$P_X - \textit{симметрична} \Leftrightarrow f(t) \in \mathbb{R}, \forall t \in \mathbb{R}$$

Доказательство. $(\Rightarrow) f(t) = E\cos(tX) + i\underbrace{E\sin(tX)}_{=0}$

$(\Leftarrow) f_X(t) = \overline{f_X(t)} = f_{-X}(t) \Rightarrow Y = -X, \int e^{itX} dF_X = \int e^{itY} dF_y \Rightarrow F_X = F_Y$ □

Определение 2. Решетчатые св
X – решетчатая $\Leftrightarrow P(X = a + kh) = 1, k \in \mathbb{Z}$. макс h - шаг решетки

Утверждение. X – *реш* $\Leftrightarrow |f(\frac{2\pi h}{n}| = 1, k \in \mathbb{Z}, h$- *шаг решетки*

4.3 Примеры

Пример. $I_c : f(t) = e^{itc}$

Пример. $Bern(p) : f(t) = p + qe^{it}$

Пример. $Bin(n,p) : f(t) = (p + qe^{it})^n$

Пример. $Pois(\lambda)$:

$$f(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}$$

Пример. $U[0,1] : f(t) = \int\limits_0^1 e^{itx} dx = \frac{e^{it}-1}{it}$

Пример. $U[-1,1] : f(t) = \frac{1}{2} \cdot \frac{e^{it} - e^{-it}}{it} = \frac{\sin t}{t}$

Пример. $N(0,1)$:

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

$$f'(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} ixe^{itx - \frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} e^{itx - \frac{x^2}{2}} \Big|_{-\infty}^{+\infty} - t \underbrace{\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx}_{f_t} = -f(t) \cdot t$$

$$\Rightarrow f(t) = e^{\frac{-t^2}{2}}$$

Пример. $N(\mu, \sigma^2) : Y = \sigma X + \mu \Rightarrow f_Y(t) = e^{it\mu} \cdot e^{-\frac{\sigma^2 t^2}{2}}$

Пример. $Exp(\lambda) : f(t) = \int\limits_0^{+\infty} \lambda e^{itx - \lambda x} dx = \frac{\lambda}{it - \lambda} e^{itx - \lambda x} \Big|_0^{+\infty} = \frac{-\lambda}{it - \lambda}$

Пример. $(n, \lambda) : X_1 \dots X_n, X_i \sim Exp(\lambda), S_n = \sum X_i, f_{S_n}(t) = (\frac{-\lambda}{it - \lambda})^n$

4.3.1 Многомерные

Определение 3. X- с.в. $f_X(t) : \mathbb{R}^n \to \mathbb{C}, f_X(t) = Ee^{i\langle t,X\rangle}$ - хар фун для с.в.

Пример. $N(\mu, \Sigma) : f_X(t) = \exp(i\langle t, \mu\rangle - \frac{1}{2}t^T\Sigma t)$

Доказательство. (С большой вероятностью это дичь, тк отсебятина). Каждая из компонент стандартного гауссовского вектора независима и имеет х.ф. $f(t_i) = \exp\left(-\frac{1}{2}t_i^2\right)$, а х.ф. по сути является матожиданием, но матожидание произведения независимых с.в. является произведением их матожиданий, тогда характеристическая функция стандартного гауссовского вектора размерности n имеет вид:

$$f(\boldsymbol{t}) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}t_i^2\right) = \exp\left(-\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}\right)$$

где $\boldsymbol{t} = (t_1, t_2, ..., t_n)^T$ - вектор переменных.

Ну а дальше просто воспользуемся свойствами х.ф. (*откуда у нас свойства многомерной х.ф... :(*):

$$N(\mu, \Sigma^2) : Y = \Sigma X + \mu \Rightarrow f_Y(t) = e^{i\langle t,\mu\rangle} \cdot e^{-\frac{t^T\Sigma t}{2}} = \exp(it^T\mu - \frac{1}{2}t^T\Sigma t)$$

□

Замечание. X – многомерное норм распределение $\Leftrightarrow \sum c_k X_k$ распределено нормально или вырождено

Доказательство. Для доказательства этого свойства воспользуемся характеристической функцией. Пусть $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ - многомерное нормально распределенное случайное вектор со средним $\boldsymbol{\mu}$ и ковариационной матрицей $\boldsymbol{\Sigma}$. Рассмотрим линейную комбинацию компонент вектора $\boldsymbol{X}$:

$$Y = \sum_{k=1}^{n} c_k X_k$$

Характеристическая функция случайной величины Y выражается следующим образом:

$$\varphi_Y(t) = \mathbb{E}[e^{itY}] = \mathbb{E}\left[e^{it\sum_{k=1}^{n} c_k X_k}\right] = \mathbb{E}\left[\prod_{k=1}^{n} e^{itc_k X_k}\right] = e^{it\boldsymbol{c}^T\boldsymbol{\mu} - \frac{1}{2}t^2\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}}$$

где $\boldsymbol{c} = (c_1, c_2, ..., c_n)$ - вектор коэффициентов линейной комбинации.

Таким образом, характеристическая функция случайной величины Y имеет вид нормальной плотности вероятности с параметрами $\boldsymbol{\mu}_Y = \boldsymbol{c}^T\boldsymbol{\mu}$ и $\boldsymbol{\Sigma}_Y = \boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}$. Если $\boldsymbol{\Sigma}_Y$ вырождена, то распределение Y вырождено. В противном случае, Y имеет многомерное нормальное распределение. □

4.4 Формула обращения

Теорема 2. *Формула обращения f- х.ф. $X \Rightarrow$*

$$\forall y > x \in C(F) : F(y) - F(x) = \frac{1}{2\pi} \lim_{\sigma\to 0} \int\limits_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-ity}}{it} \cdot f(t) \cdot e^{-\frac{\sigma^2 t^2}{2}} dt$$

Более того, если $\frac{f(t)}{t}$ – интегрируема, можно внести предел под интеграл.

Замечание. По другому: $F(y) - F(x) = \frac{1}{2\pi} \lim_{A\to\infty} \int\limits_{-A}^{+A} \frac{e^{-itx} - e^{-ity}}{it} \cdot f(t)dt$

Доказательство. $f(t) = \int\limits_{\mathbb{R}} e^{itx} p(x)dx$,
$p(z) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} e^{-itz} f(t)dt$
$\Rightarrow F(y) - F(x) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} \frac{e^{-itz}|_x^y}{-it} f(t)dt$
Так через пару действий мы придем к тому, что нужно для X - непр.
$U_\delta = X + Y_\delta, X, Y_\delta$ – нез $Y_\delta \sim N(\mu, \delta^2) \Rightarrow U_\delta$ – непр
$P(U_\delta \le t) = \iint_{X+Y\le t} dP(dx, dy) = \int\limits_{\mathbb{R}} dF_{Y_\delta}(y) \int\limits_{-\infty}^{t-y} dF_X(x) = \int\limits_{\mathbb{R}} \underbrace{F_X(t-y)p_\delta(y)}_{p_{U_\delta}} dy$
Тогда $F_{U_\delta}(y) - F_{U_\delta}(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \frac{e^{it} - e^{-it}}{it} \cdot f(t) \cdot e^{-\frac{\sigma^2 t^2}{2}} dt$– тк.$X \xrightarrow{d} X \Rightarrow U_\delta \xrightarrow{d} X, Y_\delta \xrightarrow{P} 0$
То есть мы сделали предельный переход по δ (в целом произвели "сглаживание"X при помощи Y_δ). ☐

Теорема 3. *Леви*
X_n, X – с.в.
f_n, f – х.ф.

$$X_n \xrightarrow{d} X \Leftrightarrow f_n(t) \to f(t) \forall t$$

Доказательство. Напрямую следует из того, что обсуждалось в начале прошлой лекции. ☐

Лемма 1. $P(|x| > \frac{2}{U}) \le \frac{1}{U} \int\limits_{-U}^{U} (1 - f(t))dt$*, где* X *– св,* f *– хф* X*.*

Доказательство.

$$\frac{1}{U}\int\limits_{-U}^{U}(1-f(t))dt = \frac{1}{U}\int\limits_{-U}^{U}\int\limits_{-\infty}^{+\infty}(1-e^{itx})dF(x)dt = \frac{1}{U}\int\limits_{-\infty}^{+\infty}\int\limits_{-U}^{U}(1-e^{itx})dtdF(x)$$
$$= \frac{1}{U}\int\limits_{-\infty}^{+\infty}\left(2U - \frac{e^{iUx} - e^{-iUx}}{ix}\right)dF(x) = 2\int\limits_{-\infty}^{+\infty}(1 - \frac{\sin Ux}{Ux})dF(x)$$
$$\ge 2 \cdot \int\limits_{|x| > \frac{2}{U}} \left(1 - \frac{|\sin Ux|}{|Ux|}\right) dF(x) \ge 2 \cdot \int\limits_{|x|>\frac{2}{U}} \frac{1}{2} dF(x) = P\left(|x| > \frac{2}{U}\right)$$

☐

Теорема 4. $(X_n)_{n=1}^{\infty}$ *– с.в,* $(f_n)_{n=1}^{\infty}$ *и* $f_n(t) \to f(t), \forall t \in \mathbb{R}$ *(неизвестно, является ли* f *х.ф.!)* $\Rightarrow$ *(условия равносильны)*

1. *f – х.ф X*
2. *$f(0) = 1$ и f непр в 0*
3. *(F_n) – плотная*

Доказательство. $(1 \Rightarrow 2)$ очевидно
$1 \Leftrightarrow 3$ вытекает из доказательств раньше
$2 \Rightarrow 3$
Воспользуемся леммой, и теоремой о среднем, тогда для произвольного U:

$$\lim_{n\to\infty} \int\limits_{|x|>\frac{2}{U}} dF_n(x) \le \lim_{n\to\infty} \frac{1}{U} \int\limits_{-U}^{I} (1 - f_n(t_{Un}))dt = \frac{1}{U}\int\limits_{-U}^{I}(1 - f_n(t_U))dt = 2(1 - f(t_U)) < \varepsilon$$

☐

4.5 Слабый ЗБЧ, ЦПТ

Утверждение. *Слабый закон больших чисел*

$$\{X_i\}_{i=1}^{\infty} - -i.i.d, EX_1 = \mu, S_n = \sum_{i=1}^{n} X_i \Rightarrow \frac{S_n}{n} \xrightarrow{P} \mu$$

Доказательство.

$$\frac{S_n}{n} \xrightarrow{P} \mu \xleftrightarrow{\text{для вырожденных}} \frac{S_n}{n} \xrightarrow{d} \mu \Leftrightarrow f_{\frac{S_n}{n}}(t) \to e^{it\mu} \Leftrightarrow f_{\frac{S_n}{n}-\mu}(t) \to 1 = e^{it0}$$

$$f_{x_0} = 1 + i\mu t + o(t), t \to 0 \text{ (Тейлор)}$$

$$f_{S_n} = (1 + i\mu t + o(t))^n, t \to 0$$

$$f_{\frac{S_n}{n}}(t) = f_{S_n}(\frac{t}{n}) = (1 + i\mu\frac{t}{n} + o(\frac{t}{n}))^n \xrightarrow{n\to\infty} e^{it\mu}$$

□

Теорема 5. *ЦПТ для i.i.d с.в.*
$\{X_i\}_{i=1}^{\infty} - -i.i.d, EX_i = \mu, \operatorname{Var} X_i = \sigma^2 > 0 \Rightarrow$

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} U \sim N(0,1) \Leftrightarrow \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(X)| \xrightarrow{n\to\infty} 0, (F_n(x) = P(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x))$$

Доказательство. $U_i = \frac{X_i - \mu}{\sigma}, EU_i = 0, \operatorname{Var} U_i = 1, f'(0) = 0$

$$\frac{\sum U_i}{\sqrt{n}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

$$f_{U_i}(t) = 1 - \frac{t^2}{2} + o(t^2) \Rightarrow f_{\frac{\sum U_i}{\sqrt{n}}}(t) = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-\frac{t^2}{2}}$$

□

4.6 Ещё свойства

Утверждение. X *– целочисленная* $\Rightarrow f$ *–* 2π*-периодическая функция*

Утверждение. *Рассмотрим промежуток* $[-\pi, \pi]$: $\{e^{ith}\}_{h\in\mathbb{Z}}$ *– полное и ортогональное семейство функций.*

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g}(x)dx$$

$$\sum_k p_k e^{itk} = \sum_k c_k e^{itk} \Rightarrow \langle f, e^{ith}\rangle = c_j \langle e^{itj}, e^{itj}\rangle = c_j \cdot 2\pi$$

$$\Rightarrow c_j = \frac{1}{2\pi}\langle f, e^{ith}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{ith}dt = P(X = j)$$

4.6.1 Оценка погрещности в теорему Пуассона

Лемма 2. $\mathbb{R}\mathrm{e} U < 0 \Rightarrow \begin{cases} |e^U - 1| \leq |U| \\ |e^U - U - 1| \leq \frac{|U|^2}{2} \end{cases}$

Доказательство.

$$|e^U - 1| = |\int\limits_0^U e^t dt| \overset{t=Uv}{=} |\int\limits_0^1 U e^{Uv} dv| \leq |U| \int\limits_0^1 |e^{Uv}| dv \leq |U|$$

$$|e^U - U - 1| = |\int\limits_0^U (e^t - 1) dt| = \dots \leq \frac{|U|^2}{2}$$

□

Лемма 3. $a_k, b_k, |a_k|, |b_k| \leq 1 \Rightarrow |\overbrace{\prod_{k=1}^n a_k}^{A_n} - \overbrace{\prod_{k=1}^n b_k}^{B_n}| \leq \sum_{k=1}^n |a_k - b_k|$

Доказательство.

$$|A_{n-1}a_n - B_{n-1}b_n| = |A_{n-1}a_n - A_{n-1}b_n + A_{n-1}b_n - B_{n-1}b_n| \leq$$

$$\leq |A_{n-1}| \cdot |a_n - b_n| + |b_n| \cdot |A_{n-1} - B_{n-1}| \leq |a_n - b_n| + |A_{n-1} - B_{n-1}|$$

□

Теорема 6. *Об оценке погрешности в теореме Пуассона*
$\{X_i\}_{i=1}^\infty$ X_i*– целочисленны, независимы.* $p_i = P(X_i = 1), 1 - q_i - p_i = P(X_i = 0) \Rightarrow q_i = P(X_i \notin \{0, 1\})$

$$S_n = \sum_{i=1}^n X_i, \lambda = \sum_{i=1}^n p_i (n\text{- фикс}) \Rightarrow |P(S_n = k) - e^\lambda \frac{\lambda^k}{k!}| \leq \sum_{i=1}^n p_i^2 + 2\sum_{i=1}^n q_i$$

Доказательство.

$$f_{X_i} = (1 - p_i - q_i) + p_i e^{it} + q_i \gamma_i(t), \gamma_i(t) - \text{некоторая х.ф.} = 1 + p_i(e^{it} - 1) + q_i(\gamma_i(t) - 1)$$

$$f_{S_n} = \prod_{i=1}^n f_{X_i}$$

$$\psi_i(t) = e^{p_i(e^{it} - 1)}$$

$$\varphi_i(t) = \prod_{i=1}^n \psi_i(t) = e^{\sum p_i(e^{it} - 1)} \sim Pois(\lambda)$$

$$|f_i(t) - \psi_i(t)| = (1 + p_i(e^{it} - 1)) - e^{p_i(e^{it} - 1)} + q_i(\gamma_i(t) - 1) \underset{\text{по лемме 2}}{\leq} p_i^2 |e^{it} - 1|^2 + 2q_i$$

$$|e^{it} - 1|^2 = (e^it - 1)(e^{-it} - 1) = (1 - e^{it} - e^{-it} + 1) = 1 - cos(t)$$

$$\frac{1}{2\pi} \int\limits_{-\pi}^{\pi} (\frac{p_i^2}{2} |e^{it} - 1|^2 + 2q_i) dt = 2q_i + \frac{p_i^2}{4\pi} \int\limits_{-\pi}^{\pi} |e^{it} - 1|^2 dt = 2q_i + \frac{p_i^2}{2\pi} \int\limits_{-\pi}^{\pi} (1 - cos(t)) dt = 2q_i + p_i^2$$

Тогда:

$$|P_(S_n = k) - e^{-\lambda}\frac{\lambda^k}{k!}| = \frac{1}{2\pi}|\int\limits_{-\pi}^{\pi}(f_{S_n}(t)e^{-itk} - \varphi_n(t)e^{-itk})dt| \le \frac{1}{2\pi}|\int\limits_{-\pi}^{\pi}(\prod_{i=1}^{n} f_i(t) - \prod_{i=1}^{n}\psi_i(t))dt|$$

$$\le \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\sum_{i=1}^{n}|f_i(t) - \psi_i(t)|dt \le \frac{1}{2\pi}\sum_{i=1}^{n}\int\limits_{-\pi}^{\pi}|f_i - \psi_i|dt \le \sum_{i=1}^{n}p_i^2 + 2\sum_{i=1}^{n}q_i$$

□

Следствие 3.1. *Оценка Пуассона*

$$X = np, p_{i,n} = \frac{\lambda}{n}, q_i = 0 \Rightarrow |P(S_n = k) - e^{-np}\frac{(np)^2}{k!}| \le \sum_{k=1}^{n}\frac{\lambda^2}{n^2} = \frac{\lambda^2}{n}$$

4.7 Неравенства

Теорема 7. *Неравенство Эссеена*
F, G – функции распределения, f, g – х.ф

$$\sup_{x\in\mathbb{R}}|G'(x)| \le M \Rightarrow \forall T > 0 : \sup_{x\in\mathbb{R}}|F(x) - G(x)| \le \frac{2}{\pi}\int\limits_{0}^{T}\frac{|f(t)| - g(t)|}{|t|}dt + \frac{24}{\pi T}\sup|G'(x)|$$

Доказательство. Без доказательства. □

Теорема 8. *Неравенство Берри-Эссеена*
$(X_i)_{i=1}^{\infty}$ – i.i.d. $S_n = \frac{\sum_{i=1}^{n} X_i - nX_1}{\sqrt{n}\operatorname{Var}X_1}$, $F_n(t) = P(S_n \le t)$, $E|X_1 - EX_1|^3 = \beta_3$, $\sigma = \sqrt{\operatorname{Var}X_1}$
Тогда:

$$\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(X)| \le \frac{c\beta_3}{\sigma^3\sqrt{n}}$$

Доказательство. Н.У.О $(\frac{X-EX}{\sigma}) : EX_1 = 0, \operatorname{Var}X_1 = 1 \Rightarrow \beta_3 = E|X_1|^3$
И нам нужно проверить $\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(X)| \le \frac{c\beta_3}{\sigma^3\sqrt{n}}$
$f_{S_n}(t) = f_{X_1}^n(\frac{t}{\sqrt{n}})$
Возьмем $T = \frac{\sqrt{n}}{5\beta_3}$?
Посчитаем х.ф: $f_{X_1}(t) = 1 - \frac{t^2}{2} + \frac{(it)^3}{6}EX_1^3(\cos(tX\theta_1 + isin(tX\theta_2))), |\theta_1|, |\theta_2| \le 1$
$f_{X_1}(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2}{2n} + \frac{(it)^3}{n^{\frac{3}{2}}6}EX_1^3(\cos(\frac{t}{\sqrt{n}}X\theta_1 + isin(\frac{t}{\sqrt{n}}X\theta_2)))$
ценим при $t < |T|$:

$$1-|f_{X_1}(\frac{t}{\sqrt{n}}| \le |1-f_{X_1}(\frac{t}{\sqrt{n}})| = |\frac{t^2}{2n}+\frac{(it)^3}{n^{3/2}6}EX_1^3(\cos(\frac{t}{\sqrt{n}}X\theta+isin(\frac{t}{\sqrt{n}}X\theta)))| \overset{|\cos t|,|\sin t|\le 1}{\le} |\frac{t^2}{2n}+\frac{t^3\beta_3}{3n^{3/2}}| \le \frac{1}{25}$$

Таким образом у нас получилось, что х.ф. отделима от нуля $|f_{X_1}(\frac{t}{\sqrt{n}})| \ge \frac{24}{25}$
$f_{S_n}(t) = f_{X_1}^n(\frac{t}{\sqrt{n}}) = \exp(n\ln f_{X_1}(\frac{t}{\sqrt{n}}))$
Замечание.

$$f(t) = 1 = iEX \cdot t - \frac{t^2}{2}EX^2 - \frac{it^3}{6}EX^3 + \dots$$

$$lnf(t) = S_1 it + \frac{S_2(it)^2}{2} + \frac{S_3(it)^3}{6}$$

S_k называются семинвариантами, причем $S_1 = EX, S_2 = \sigma^2$

$$lnf_{X_1}(\frac{t}{\sqrt{n}}) = -\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}}(lnf)'''(\theta\frac{t}{\sqrt{n}})$$

$$ln'''f(s) = \frac{f'''f^2 - 3f''f' + 2(f')^3}{f^3} = \frac{E(iX_i)^3e^{iX_is}f^2 - 3E(iX_i)^2e^{iX_is} - 3EX_ie^{iX_is} + 2(EiX_ie^{iX_is})^3}{f^3(s)}$$

$$\beta_1 \leq \beta_2^{1/2} \leq \beta_3^{1/3}(\text{по н-ву Ляпунова})$$

$$|f_{X_1}(\frac{t}{\sqrt{n}})| \geq \frac{24}{25}, |t| < \frac{sqrtn}{5\beta_3}, |f(s)| \leq 1 \Leftarrow |ln'''(\theta\frac{t}{\sqrt{n}})| \leq \frac{\beta_3 + 3\beta_1\beta_2 + 2\beta_1^3}{(\frac{24}{25})^3} \leq 7\beta_3$$

Замечание. Имеет место неравенство: $|e^z - 1| \leq |z|e^{|z|}$

Оценим:

$$\begin{aligned}
\left|f_{X_1}^n(\frac{t}{\sqrt{n}})) - e^{-t^2/2}\right| &= \left|e^{nlnf_{X_1}(\frac{t}{\sqrt{n}})} - e^{-t^2/2}\right| \\
&= \left|e^{-t^2/2} - e^{-\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}}(lnf)'''(\theta\frac{t}{\sqrt{n}})}\right| \\
&\leq e^{-t^2/2}\left|\frac{(it)^3}{6\sqrt{n}}(lnf''')(\theta\frac{t}{\sqrt{n}})\right| \exp\left(\left|\frac{t^3}{6\sqrt{n}}lnf'''(\theta\frac{t}{\sqrt{n}})\right|\right) \\
&\leq e^{\frac{-t^2}{2}}\frac{t^3}{6\sqrt{n}}7\beta_3 \exp\left(\frac{t^3}{6\sqrt{n}}\right)(lnf''') \\
&\leq \frac{t}{6}\frac{\beta_3|t|^3}{\sqrt{n}}e^{\frac{t^2}{4}}
\end{aligned}$$

$$\left|\int_0^T \frac{|f_{S_n}(t) - \Phi(t)}{|t|}dt\right| \leq \frac{7}{6}\frac{\beta_3}{\sqrt{n}} \underbrace{\int_0^T |t|^2e^{-\frac{t^2}{4}}dt}_{\text{заведомо сходится}}$$

□

Следствие 3.1. *Оценка интегральной теоремы Муавра-Лапласа:*

$$\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(x| \leq c \cdot \frac{p^2 + q^2}{\sqrt{pq}\sqrt{n}}$$

Доказательство.

$$E\left|X - EX\right|^3 = E\left|X - p\right|^3 = (1-p)^3p + p^3(1 - p_= pq(p^2 + q^2) \Rightarrow C\frac{(p^2+q^2)pq}{(pq)^{3/2}\sqrt{n}} = C \cdot \frac{p^2+q^2}{\sqrt{pqn}}$$

□

5 Моделирование случайных величин

На самом деле, в реальности сгенерировать реально "случайное"число практически невозможно, поэтому в современных языках программирования генерируют "псевдослучайных"чисел. Тут мы рассмотрим, как это будет происходить.

Самым простым примером будет **линейный генератор**, где мы просто берем число, как-то его преобразуем (складываем, умножаем), а потом берем остаток по довольно большому модулю.

Теперь подумаем как работает "‘random.random()"‘ из Python, которая симулирует равномерное распределение на $[0, 1]$. И как правило, когда мы говорим "возьмем случайное подразумеваем именно равномерное распределение.

В C++ всё сложнее и многие распределения уже реализованы как-то.

А что в Java? Есть класс Random. Для генерации мы используем фукнцию nextDouble().

Мораль в чем? Обычно в языках программирования стараются симулировать в первую очередь равномерное распределение (либо дискретное, либо непрерывное).

То есть, обычно $rand() \sim U[0,1]$

А теперь мы рассмотрим как генерировать другие распределения.

Как получить $N(\mu, \sigma)$: ЦПТ, 4 задачка из лабы.

$F^{-1}(rand()) \sim F$

То есть как можно поступать? Либо пользоваться предельными соотношениями, либо через обратную функцию (по формуле выше), либо приближенными методами (rejection assembling и тд).

Что с дискретными распределениями? Можно симулировать их через $U[0,1]$. Например, $Bern(p)$ можно так сделать ‘rand() < p ? 1 : 0‘. Значит автоматически научились генерировать и $Bin(n,p)$ (просто сгенерим n бернуллевских) или можно даже быстрее как-нибудь.

Хотим ещё Пуассона $Pois(\lambda)$ симулировать: $\max\{j : \sum_{k=1}^{j} \ln U_k \leq e^{\lambda}\}$, тогда получится, что $P(X \leq k) = \int_k^{\infty} p_{\Gamma k,n}(x)dx$ и если как-нибудь посчитать, то это ровно то, что нужно.

6 Что ещё происходит в теории вероятностей?

Курс почти закончился, но его, понятное дело, не хватает, чтобы обсудить всю эту науку. Что же в ней есть ещё?

Мы сформулировали слабый ЗБЧ, а есть ещё несколько усиленных ЗБЧ, где утверждается, что сумма сходится к матожиданию почти наверное, а не по вероятности, как в слабом.

Одно из достаточных условий: $\lim\limits_{n\to\infty} \frac{\sum \operatorname{Var} X_i}{n} < \infty$ (условие Колмогорова)

Ещё можно рассматривать предельные теорему не только когда все св одинаковые, но и другие. Там много всего интересного.

Например, есть вероятности больших уклонений, например, что-то про $\lim\limits_{\sqrt{n}x_n\to\infty} P(S_n > x_n\sqrt{n})$

Также можно упомянуть про случайные процессы, которые является ответвлением от классической теории вероятностей. Мы рассматривали случайные величины, а там, грубо говоря, индексированные случайные величины $X_t, t \in T \subset \mathbb{R}$. Там есть важные классы процессов: Гауссовские и Пуассоновские.

Не менее важным разделом является математическая статистика (с ней мы подробнее познакомимся в следующем семестре). Там мы, решаем в каком-то смысле обратную задачу теории вероятностей – пытаемся по реальным данным понять, что распределение у нас они представляют.

Что такое слабо-зависимые величины? Может утверждаться $\operatorname{cov}(X_i, X_j) = 0$ или что-то такое. В некоторых других источниках рассказывается именно в таком виде, то есть, с чуть более слабыми ограничениями.