Курс «Теория вероятностей»

Александр Вадимович Булинский

Весенний семестр 2021

Аннотация

Конспект подготовлен дилетантами, содержит кучу опечаток (2 и 8 лекции лучше не видеть, но главную идею понять несложно) и не является материалом, по которому стоит готовиться на 5. Пишущий эти строки человек рекомендует пользоваться книгой Альберта Николаевича Ширяева «Вероятность I», а еще лучше посещать лекции Александра Вадимовича. Но если вам теория вероятностей нужна только для того, чтобы ее сдать, то можно готовиться по этим конспектам. В целях более логичного изложения на бумаге редактор отступает от разбиения по лекциям и некоторые теоремы или утверждения относит туда, где они должны быть. Разбиение по подсекциям примерно соответствует программе экзамена 2021 года.

Литература, рекомендованная А. В. Булинским:

- А. Н. Ширяев – Вероятность-1
- В. Феллер – Введение в теорию вероятностей и ее приложения
- А. А. Боровков – Теория вероятностей
- Б. В. Гнеденко – Курс теории вероятностей
- Л. Б. Коралов, Я. Г. Синяй – Теория вероятностей и случайные процессы
- J. Jacod, P. Protter – Probability Essentials
- R. Durett – Probability: Theory and Examples

Все ссылки на июль 2021 года рабочие.

Содержание

1 Основные понятия теории вероятностей

Определение 1.1. Система $\mathcal{A}$ подмножеств Ω называется *алгеброй*, если

1. $\varnothing \in \mathcal{A}$
2. $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$
3. $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$

Определение 1.2. Система $\mathcal{F}$ подмножеств Ω называется *σ-алгеброй*, если

1. $\varnothing \in \mathcal{F}$
2. $A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{A}$
3. $A_1, A_2, ... \in \mathcal{F} \implies \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{F}$

Утверждение 1.1. *σ-алгебра является алгеброй подмножеств.*

Доказательство. Свойства (1) и (2) совпадают. Проверим (3):

$$A, B \in \mathcal{F} \implies A \cup B \cup \varnothing \cup \varnothing \cup ... \in \mathcal{F} \implies A \cup B \in \mathcal{F}$$

□

Утверждение 1.2. *Пересечение любой совокупности σ-алгебр является σ-алгеброй.*

Определение 1.3. $\sigma(M)$ — наименьшая (по включению) σ-алгебра, содержащая M.

Утверждение 1.3. $\sigma(M) = \bigcap\limits_{\alpha} S_\alpha$, $\{S_\alpha\}$ *— все σ-алгебры, содержащие M.*

Определение 1.4. Борелевской σ-алгеброй $\mathcal{B}(S)$ топологического пространства S называется наименьшая σ-алгебра, содержащая топологию этого пространства.

Определение 1.5. Система $\mathcal{M}$ подмножеств Ω называется π-системой, если $A, B \in \mathcal{M} \implies A \cap B \in \mathcal{M}$.

Определение 1.6. Система $\mathcal{D}$ подмножеств Ω называется λ-системой (системой Дынкина), если

1. $\Omega \in \mathcal{D}$
2. $A, B \in \mathcal{D},\ A \subseteq B \implies B \setminus A \in \mathcal{D}$
3. $A_1, A_2, ... \in \mathcal{D},\ A_k \uparrow A \implies A \in \mathcal{D}$ (замкнутость), т.е. $A_k \in \mathcal{D},\ A_n \subseteq A_{n+1} \implies \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{D}$

Теорема 1.1. *Совокупность $\mathcal{F}$ подмножеств Ω является σ-алгеброй тогда и только тогда, когда $\mathcal{F}$ является π- и λ-системой.*

Доказательство. $\implies$ Заметим, что $B \setminus A = B \cap (\Omega \setminus A)$. В эту сторону очевидно.
$\impliedby$ Пусть $\mathcal{F}$ одновременно π- и λ-система. $A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}$ (в силу свойств (1) и (2) λ-системы).

$$A_n \in \mathcal{F} \implies \bigcup_{k=1}^{N} A_k = \left(\left(\bigcup_{k=1}^{N} A_k\right)^C\right)^C = \left(\bigcap_{k=1}^{N} {A_k}^C\right)^C \in \mathcal{F}$$

$\bigcup\limits_{k=1}^{N} A_k \uparrow \bigcup\limits_{k=1}^{\infty} A_k$ при $N \implies \infty$, т.е. $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{F}$ по свойству (3) λ-системы. □

Обозначим за $\lambda(M)$ наименьшую λ-систему, содержащую M (заметим, что пересечение любой совокупности λ-систем — λ-система, наименьшая — пересечение всех, содержащих M).

Теорема 1.2. *Пусть π-система $\mathcal{M}$ вложена в λ-систему $\mathcal{D}$. Тогда $\sigma(\mathcal{M}) = \lambda(\mathcal{M}) \subseteq \mathcal{D}$.*

Доказательство. Без ограничения общности $\lambda(\mathcal{M}) = \mathcal{D}$. Пусть $A \in \mathcal{M}$. Обозначим $\mathcal{D}_A := \{B \subseteq \Omega : A \cap B \in \mathcal{D}\}$ — нетрудно показать, что это λ-система. Тогда $\mathcal{D} = \lambda(\mathcal{M}) \subseteq \mathcal{D}_A \implies \forall A \in \mathcal{M},\ \forall B \in \mathcal{D}\ A \cap B \in \mathcal{D}$.
$B \in \mathcal{D},\ \mathcal{D}(B) := \{A \subseteq \Omega : A \cap B \in \mathcal{D}\}$ — λ-система. Т.к. из сказанного выше $\mathcal{M} \subseteq \mathcal{D}(B) \implies \lambda(\mathcal{M}) = \mathcal{D} \subseteq \mathcal{D}(B)$
$\forall A, B \in \mathcal{D}\ A \cap B \in \mathcal{D} \implies \mathcal{D}$ — π-система. По теореме 1.1 $\mathcal{D}$ является σ-алгеброй. Значит, $\sigma(\mathcal{M}) \subseteq \mathcal{D}$. С другой стороны, $\mathcal{D} = \lambda(\mathcal{M}) \subseteq \sigma(\mathcal{M}) \implies \sigma(\mathcal{M}) = \lambda(\mathcal{M})$. □

Определение 1.7. *Измеримым пространством* $(S, \mathcal{B})$ называется множество S с выделенной σ-алгеброй подмножеств $\mathcal{B}$. Множества из $\mathcal{B}$ называются *измеримыми*.

Определение 1.8. *Мерой* на $(S, \mathcal{B})$ называется функция $\mu : \mathcal{B} \implies [0, +\infty]$: $\forall B_1, B_2, ... : B_i \cap B_j = \varnothing \implies \mu(\bigcup\limits_i B_i) = \sum\limits_i \mu(B_i)$.

Утверждение 1.4. $\mu(\varnothing) = 0 \iff \exists B : \mu(B) < \infty$.

Следствие 1.1. *Из счетной аддитивности меры следует конечная аддитивность меры.*

Определение 1.9. *Вероятностной мерой* на $(\Omega, \mathcal{F})$ называется такая счетно-аддитивная мера P, что $P(\Omega) = 1$.

Определение 1.10. *Вероятностным пространством* называется тройка $(\Omega, \mathcal{F}, P)$, $(\Omega, \mathcal{F})$ — измеримое пространство, P — вероятностная мера.

Определение 1.11. Говорят, что вероятностное пространство $(\Omega, \mathcal{F}, P)$ *дискретно*, если $|\Omega| \leq |\mathbb{N}|$.

Пусть даны $p_1, p_2, \ldots > 0$, $|\{p_k\}| = |\Omega|$. Тогда $P(A) = \sum\limits_{k : w_k \in A} p_k$.

Лемма 1.1. *Пусть $I_1, I_2, \ldots$ — последовательность попарно непересекающихся подмножеств $\mathbb{N}$. Тогда $\sum\limits_{k \in \bigcup\limits_n I_n} p_k = \sum\limits_n \sum\limits_{k \in I_n} p_k$.*

Теорема 1.3. *Если есть вероятностное пространство* $(\Omega, \mathcal{F}, P)$, $\mathcal{F} \neq 2^\Omega$, *то* $\exists \widetilde{P}$ *на* $(\Omega, 2^\Omega) : \widetilde{P}|_\mathcal{F} \equiv P$.

Определение 1.12. *(Классическое определение вероятности)* Пусть $|\Omega| = N$, $A \subseteq \Omega$, $|A| = n$. Тогда $P(A) = \frac{n}{N}$.

2 Свойства вероятности. Некоторые распределения.

2.1 Некоторые дискретные распределения

$\mathcal{F} = 2^\Omega,\ p_1, p_2, \dots \geq 0,\ \sum\limits_i^\infty p_i = 1,\ P(A) := \sum\limits_{i: w_i \in A} p_i.$
Если $|\Omega| = n,\ p_1 = p_2 = \dots = p_n = \frac{1}{n}$ — получаем классическое определение вероятности.

1. *Биномиальное распределение* $B(n, p)$ — распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких, что вероятность «успеха» в каждом из них постоянна и равна p.
   $\Omega = \{w = (i_1, \dots, i_n)\}$, $0 < p < 1$ — фиксированное число.
   $P(w) = p^k(1-p)^{n-k}$
   $P(\text{ровно k успехов}) := C_n^k p^k (1-p)^{n-k}$
   $P(\Omega) = \sum\limits_w P(w) = \sum\limits_{k=1}^n C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$

2. *Геометрическое распределение* — распределение первого «успеха» в серии испытаний Бернулли ($B(p)$).
   $\Omega = \{(1), (0, 1), (0, 0, 1), \dots\}$, $0 < p < 1$ — фиксированное число.

$$P(\text{успех наступит на } n \text{ испытании}) = (1-p)^{n-1} p$$

3. *Гипергеометрическое распределение* — в ящике лежат N белых и M черных шаров, вытягивают n шаров.

$$P(\text{вынуто } k \text{ черных шаров}) = \frac{C_M^k C_N^{n-k}}{C_{M+N}^n}$$

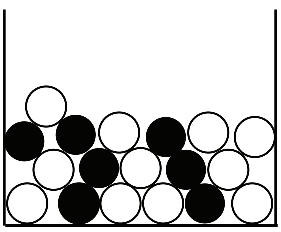


Рис. 1: Гипергеометрическое распределение

4. *Пуассоновское распределение* $Pois(\lambda)$ — число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью (λ) и независимо друг от друга. $\Omega = \{0, 1, 2, \dots\} \cong \mathbb{N} \cup \{0\}$.

$$P(\{n\}) = \frac{\lambda^n e^{-\lambda}}{n!}$$

2.2 Геометрические вероятности

$P(A) := \frac{\mu(A)}{\mu(\Omega)}$.

1. *Парадокс Бертрана.* Рассмотрим равносторонний треугольник, вписанный в окружность. Наудачу выбирается хорда окружности. Какова вероятность того, что выбранная хорда длиннее стороны треугольника?

   (a) *Метод «случайного центра».* Бросаем точку и проводим через нее хорду, перпендикулярную радиусу. Хорда длиннее стороны равностороннего треугольника, если выбранная точка находится внутри круга, вписанного в треугольник. $P = \frac{\pi\left(\frac{R}{2}\right)^2}{\pi R^2} = \frac{1}{4}$.

   (b) *Метод «случайных концов».* Фиксируем точку на окружности, берем произвольный угол α. Чтобы посчитать искомую вероятность, представим, что треугольник повёрнут так, что одна из его вершин совпадает с концом хорды. Заметим, что если другой конец хорды лежит на дуге между двумя другими вершинами треугольника, то длина хорды больше стороны треугольника. Длина рассмотренной дуги равна трети длины окружности, значит, искомая вероятность равна $P = \frac{\frac{1}{3}(2\pi R)}{2\pi R} = \frac{1}{3}$.

   (c) *Метод «случайного радиуса».* Случайно выбираем радиус, бросаем на радиус точку и строим через него хорду, перпендикулярную этому радиусу. Хорда длиннее стороны треугольника, если её центр ближе к центру, чем точка пересечения треугольника с зафиксированным радиусом, а сторона треугольника делит радиус пополам. $P = \frac{R}{2R} = \frac{1}{2}$.

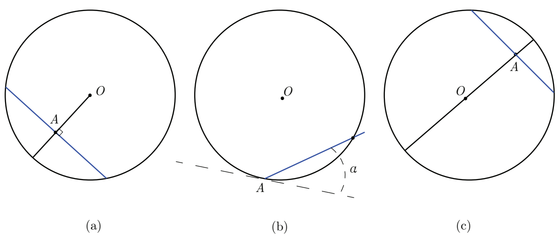


Рис. 2: Парадокс Бертрана. Синим отмечена полученная хорда

Суть парадокса: проблема точной формализации эксперимента. Тогда и только тогда, когда метод случайного выбора задан, проблема имеет чётко определённое решение.

2. *Игла Бюффона.* Рассмотрим лист бумаги, расчерченный параллельными прямыми на n полос длины L и ширины d ($|\Omega| = nd \cdot L$). Случайно бросаем на него иглу длиной $l < d$ (случайно выбираем центр (x, y), затем случайно выбираем угол поворота α). Какова вероятность того, что игла пересечет одну из линий решетки?
Пусть центр (x, y) оказался в полосе между $(k-1)$-ой и k-ой линией, причем $y > (k-1)d + \frac{d}{2}$ (случай, где центр находится ближе к $(k-1)$-ой линии, симметричен). Тогда условие того, что игла, повернутая на угол $0 < \alpha < \pi$ пересекла k-ую линию:

$$\begin{cases} kd - \frac{d}{2} < y \le kd \\ kd - y < \frac{l}{2}\sin\alpha \end{cases} \implies kd - \frac{l}{2}\sin\alpha < y \le kd$$

Получаем $\int\limits_0^\pi (kd - (kd - \frac{l}{2}\sin\alpha))d\alpha = l$ — для половины полосы.
Искомая вероятность $P = \frac{2l \cdot nL}{\pi d \cdot nL} = \frac{2l}{\pi d}$.

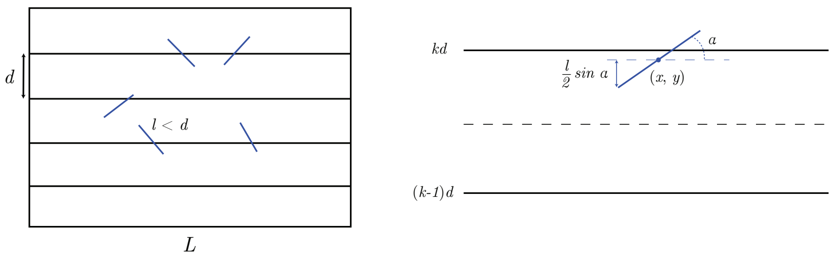


Рис. 3: Игла Бюффона

2.3 Свойства вероятностной меры

Теорема 2.1. *Пусть $A, B \in \mathcal{F}$, $A_n \in \mathcal{F}$. Справедливы следующие утверждения:*

1. $A \subseteq B \implies P(B \setminus A) = P(B) - P(A)$, *в частности,* $P(B) \ge P(A)$
2. $P(A \cup B) = P(A) + P(B) - P(AB)$
3. (субаддитивность) $P(\bigcup\limits_n A_n) \le \sum\limits_n P(A_n)$

Доказательство. 1. $B = A \sqcup (B \setminus A) = A \sqcup (B\overline{A})$. Т.к. $A(B \setminus A) = \varnothing$, то $P(B) = P(A) + P(B \setminus A)$.

2.
$$\begin{cases} P(A \cup B) = P(A) + P(B \setminus A) \\ P(B) = P(AB) + P(B \setminus A) \end{cases} \implies P(A \cup B) = P(A) + P(B) - P(AB)$$

3. $B_1 = A_1,\ B_2 = A_2 \setminus A_1,\ B_3 = A_3 \setminus (A_1 \cup A_2), \ldots$
$\bigsqcup\limits_i B_i = \bigcup\limits_i A_i \implies P(\bigcup\limits_i A_i) = P(\bigsqcup\limits_i B_i) = \sum\limits_i P(B_i) \le \sum\limits_i P(A_i)$ (мажорируем ряд).

$\square$

Теорема 2.2. *Пусть $A_1, \dots, A_n \in \mathcal{F}$. Тогда* $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i} P(A_i) - \sum\limits_{i,j} P(A_iA_j) + \sum\limits_{i,j,k} P(A_iA_jA_k) - \dots + (-1)^{n-1}P(A_1 \dots A_n)$.

Доказательство. По индукции. □

Определение 2.1. Неотрицательная функция $\mu : \mathcal{A} \implies \mathbb{R}$, $\mathcal{A} \subseteq 2^S$ — алгебра, называется *счетно-аддитивной*, если $\forall A_1, A_2, \dots \mathcal{A} : A_i \cap A_j = \varnothing \iff i \neq j$ и $\bigcup\limits_i A_i \in \mathcal{A}$ верно равенство $\mu(\bigcup\limits_i A_i) = \sum\limits_i \mu(A_i)$.

Определение 2.2. Неотрицательная функция $\mu : \mathcal{A} \implies \mathbb{R}$, $\mathcal{A} \subseteq 2^S$ — алгебра, называется *непрерывной в нуле*, если $\forall A_n \downarrow \varnothing$ при $n \implies \infty \implies \mu(A_n) \implies 0$ при $n \implies \infty$.

Лемма 2.1. *Пусть μ — неотрицательная конечно-аддитивная функция на $\mathcal{A}$ и $\forall A \in \mathcal{A}$ $\mu(A) < \infty$. Тогда $\mu \in C(\varnothing) \implies \mu$ непрерывна на монотонных последовательностях: $A_n \downarrow (\uparrow) A$ при $n \implies \infty \implies \mu(A_n) \implies \mu(A)$ при $n \implies \infty$.*

Доказательство. $\forall A \in \mathcal{A}\, \mu(A) < \infty \iff \mu(S) < \infty$. Пусть $A_n \downarrow A \in \mathcal{A}$. Тогда $\mu(A_n \setminus A) \implies 0$ (μ непрерывна в нуле), а $\mu(A_n \setminus A) = \mu(A_n) - \mu(A) \implies \mu(A_n) \implies \mu(A)$.
Для монотонно возрастающих последовательностей: положим $B_n = \overline{A_n}, B = \overline{A}$. Тогда $B_n \downarrow B \implies \mu(B_n) \implies \mu(B) \implies \mu(S) - \mu(A_n) \implies \mu(S) - \mu(A) \implies \mu(A_n) \implies \mu(A)$ при $n \implies \infty$. □

Теорема 2.3. *Пусть μ — конечная неотрицательная функция, заданная на алгебре $\mathcal{A}$ подмножеств S. Тогда μ счетно-аддитивна тогда и только тогда, когда μ конечно-аддитивна и непрерывна в нуле.*

Доказательство. $\implies$. Пусть μ счетно-аддитивна, тогда она, очевидно, является конечно-аддитивной. Покажем ее непрерывность в нуле: $A_n \in \mathcal{A} \downarrow \varnothing$. Положим $B_n = A_n \setminus A_{n+1}$. Очевидно, что $B_iB_j = \varnothing$ при $i \neq j$. Заметим, что $A_n = \bigcup\limits_{k=n}^{\infty} B_k$. $A_1 = \bigcup\limits_{k=1}^{\infty} B_k \implies \mu(A_1) = \sum\limits_{k=1}^{\infty} \mu(B_k)$ в силу счетной аддитивности μ. Поскольку ряд сходится, $\mu(A_n) \implies 0$ при $n \implies \infty$.
$\impliedby$. Возьмем $C_k \in \mathcal{A} : C_iC_j = \varnothing$ при $i \neq j$, $\bigcup\limits_k C_k \in \mathcal{A}$. Положим $A_n = \bigcup\limits_{k=n}^{\infty} C_k$. Очевидно, что $A_1 = C_1 \cup C_2 \cup \dots \cup C_{n-1} \cup A_n$, $n \geq 2 \implies A_n = A_1 \setminus (C_1 \cup \dots \cup C_{n-1}) \in \mathcal{A}$, причем $C_1, \dots C_{n-1}, A_n$ попарно не пересекаются. Очевидно, что $A_n \downarrow \varnothing$ и из конечной аддитивности $\mu(A_1) = \mu(C_1) + \dots + \mu(C_{n-1} + \mu(A_n))$. Из непрерывности в нуле: $\mu(A_n) \implies 0 \implies \sum\limits_{k=1}^{n} \mu(C_k) \implies \mu(A_1) = \mu(\bigcup\limits_{k=1}^{\infty} C_k) \implies \sum\limits_{k=1}^{n} \mu(C_k) = \mu(\bigcup\limits_{k=1}^{\infty} C_k)$. □

Теорема 2.4. *Пусть P, Q — вероятностные меры, заданные на измеримом пространстве $(S, \mathcal{B})$, совпадающие на π-системе $\mathcal{M}$. Тогда они совпадают на $\sigma(\mathcal{M})$.*

Доказательство. Положим $\mathcal{D} = \{A \in \mathcal{B} : P(A) = Q(A)\}$. Тогда по условию теоремы $\mathcal{M} \subseteq \mathcal{D}$. Заметим, что

1. $S \in \mathcal{D}$, т.к. $P(S) = Q(S) = 1$.
2. $A, B \in \mathcal{D},\ A \subseteq B \implies B \setminus A \in \mathcal{D}$, т.к. $P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$.
3. $A_n \in \mathcal{D},\ A_n \uparrow A \implies A \in \mathcal{D}$, т.к. $0 = \lim\limits_{n \implies \infty} 0 = \lim\limits_{n \implies \infty} P(A_n) - Q(A_n) = P(A) - Q(A) \implies P(A) = Q(A) \implies A \in \mathcal{D}$ (из непрерывности вероятности).

Значит, $\mathcal{D}$ является λ-системой. По теореме 1.2 $\mathcal{D} = \sigma(\mathcal{M})$. □

Теорема 2.5. (Каратеодори). *Любая вероятностная мера на алгебре $\mathcal{A} \subseteq 2^S$ однозначно продлевается на $\sigma(\mathcal{A})$.*

Доказательство. (схема)

1. Определим внешнюю меру следующим образом: $P^*(A) = \inf\{\sum\limits_n P(A_n) : A_n \in \mathcal{A}, A \subseteq \bigcup\limits_n A_n\}$.
2. Из теоремы 2.1 следует, что можно обойтись покрытиями, состоящими из попарно непересекающихся множеств $A_n \in \mathcal{A}$.
3. *Оказывается*, что $P^*|_{\mathcal{A}} \equiv P$.
4. $\forall A, B \in 2^S\ \rho(A, B) = P^*(A \triangle B)$.
5. Положим $\mathcal{A}^* = \{A \in 2^S : \exists A_n \in \mathcal{A} \rho(A_n, A) \implies 0$ при $n \implies \infty\}$. *Очевидно*, что $\mathcal{A} \subseteq \mathcal{A}^*$.
6. *Проверяется*, что $\mathcal{A}^*$ — σ-алгебра.
7. *Доказывается*, что $(S, \mathcal{A}^*, P^*)$ — вероятностное пространство.

Единственность продления следует из теоремы 2.4, т.к. алгебра есть π-система. □

3 Условные вероятности

3.1 Вероятностные меры на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Примеры распределений

Определение 3.1. Пусть P – вероятностная мера на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Её функцией распределения назовём функцию $F(x) := P((-\infty, x]), x \in \mathbb{R}$.

Теорема 3.1. *Пусть F – функция распределения. Тогда F обладает следующими свойствами:*

1. *F не убывает на $\mathbb{R}$*
2. *F непрерывна справа в каждой точке*
3. $\lim\limits_{x \to -\infty} F(x) = 0$
4. $\lim\limits_{x \to +\infty} F(x) = 1$

Доказательство. 1. Очевидно: если $x \le y$, то $(-\infty, x] \subset (-\infty, y]$. Осталось вспомнить теорему 2.1.

2. Если $x_n \downarrow x$, то $(-\infty, x_n] \downarrow (-\infty, x]$, и по непрерывности вероятностной меры получаем $F(x_n) \to F(x)$.

3. Рассмотрим $B_n = (-\infty, x_n]$. Если $x_n \downarrow -\infty$, то $B_n \downarrow \varnothing$, и $\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(B_n) = P(\varnothing) = 0$.

4. Аналогично пункту 3, только теперь берём $x_n \uparrow +\infty$.

□

Теорема 3.2. *Пусть F обладает свойствами из 3.1. Тогда на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ существует единственная вероятностая мера P, для которой F является функцией распределения.*

Доказательство. Допустим, что нашлась такая P, для которой F является функцией распределения. Тогда $P((a,b]) = P((-\infty, b] \setminus (-\infty, a]) = F(b) - F(a)$.

Пусть $A = \bigcup\limits_{k=1}^{m} (a_k, b_k]$, $(a_i, b_i] \cap (a_j, b_j] = \varnothing$ при $i \neq j$ (допускаем бесконечные промежутки). Тогда $P(A) = \sum\limits_{k=1}^{m} P((a_k, b_k]) = \sum\limits_{k=1}^{m} (F(b_k) - F(a_k))$. Легко видеть, что множества, на которых мы только что определили меру P, образуют алгебру подмножеств на прямой $\mathcal{A}$, и $P(A)$ не зависит от вида представления A в виде объединения.

Напоминание(2.3): P на $\mathcal{A}$ сигма-аддитивна $\iff$ конечно-аддитивна и непрерывна в $\varnothing$. Конечная аддитивность очевидна. Проверим, что она непрерывна в $\varnothing$. Тогда P будет вероятностной на $\mathcal{A}$.

Пусть $A_n = \bigcup\limits_{k=1}^{m_n} (a_k^{(n)}, b_k^{(n)}]$, $A_n \downarrow \varnothing$. Докажем, что $\lim_{n\to\infty} P(A_n) = 0$.

По свойствам 3 и 4 функции распределения (3.1) имеем $\forall \varepsilon > 0$ $\exists a = a(\varepsilon), b = b(\varepsilon) : F(a) < \frac{\varepsilon}{2}, F(b) > 1 - \frac{\varepsilon}{2}$. Тогда $P(\mathbb{R} \setminus (a,b]) = F(a) + 1 - F(b) < \varepsilon$.

$P(A_n) = P(A_n \cap (\mathbb{R} \setminus (a,b])) + P(A_n \cap (a,b])$. В предыдущей строчке написано, что первое слагаемое меньше ε.

Без ограничения общности считаем, что $A_n \downarrow \varnothing$, а также $\forall n\ A_n \subset (a,b]$. Будем рассматривать только $A_n \neq \varnothing$.

Итак, $A_n = \bigcup\limits_{k=1}^{m_n} (a_k^{(n)}, b_k^{(n)}]$. Рассмотрим $B_n = \bigcup\limits_{k=1}^{m_n} (x_k^{(n)}, b_k^{(n)}]$, где $x_k^{(n)}$ выбраны так, что $a_k^{(n)} < x_k^{(n)} < b_k^{(n)}$ и $F(x_k^{(n)}) - F(a_k^{(n)}) < \frac{2^{-n}\varepsilon}{m_n}$ (это возможно в силу того, что F непрерывна справа). Ещё рассмотрим $C_n = \bigcup\limits_{k=1}^{m_n} [x_k^{(n)}, b_k^{(n)}]$. Очевидно, C_n замкнуто, $C_n \subset A_n$ и $P(A_n) - P(C_n) \le \sum\limits_{k=1}^{m_n} \frac{2^{-n}\varepsilon}{m_n} = 2^{-n}\varepsilon$. $\bar{C}_n := \mathbb{R} \setminus C_n$ открыто. $\bigcap\limits_{n=1}^{\infty} C_n \subset \bigcap\limits_{n=1}^{\infty} A_n = \varnothing$, значит, $\bigcup\limits_{n=1}^{\infty} \bar{C}_n \supset \bar{\varnothing} = \mathbb{R}$. $\bigcup\limits_{n=1}^{\infty} \bar{C}_n \supset [a,b]$, значит, $\exists n_1 < \ldots < n_r : [a,b] \subset \bar{C}_{n_1} \cap \ldots \cap \bar{C}_{n_r}$, а тогда $C_{n_1} \cap \ldots \cap C_{n_r} \subset \overline{[a,b]}$. Но $\forall i\ C_{n_i} \subset [a,b]$, значит, $C_{n_1} \cap \ldots \cap C_{n_r} = \varnothing$.

$$P(A_{n_r}) = P\left(A_{n_r} \setminus \bigcup_{j=1}^{r} C_{n_j}\right) = P\left(A_{n_r} \cap \overline{\bigcup_{j=1}^{r} C_{n_j}}\right) = P\left(A_{n_r} \cup \bigcup_{j=1}^{r} \bar{C}_{n_j}\right) \leq$$
$$\sum_{j=1}^{r} P(A_{n_r} \cap \bar{C}_{n_j}) \leq \sum_{j=1}^{r} P(A_{n_j} \cap \bar{C}_{n_j}) \leq \varepsilon(\sum_{j=1}^{r} 2^{-n_j}) \leq \varepsilon.$$

Итак, всё. Доказали, что $\lim_{n\to\infty} P(A_n) = 0$, и получили, что мера является сигма-аддитивной. Осталось заметить, что по теореме Каратеодори меру можно однозначно продолжить с $\mathcal{A}$ на $\sigma\{\mathcal{A}\} = \mathcal{B}(\mathbb{R})$. $\square$

Определение 3.2. Функция $p(x)$ называется плотностью, если $p(x) \geq 0$ $\forall x \in \mathbb{R}$ и $\int\limits_{-\infty}^{+\infty} p(x)dx = 1$.

Утверждение 3.1. *Возьмём* $F(x) = \int\limits_{(-\infty,x]} p(u)du$. *Тогда* F *обладает свойствами 1-4 из 3.1, и следовательно является функцией распределения некоторой вероятностной меры.*

Пример 3.1. Равномерное распределение на отрезке $[a, b]$

$$p(x) = \frac{1}{b-a} \cdot \mathbb{1}([a, b])$$

Пример 3.2. Экспоненциальное распределение с параметром $\lambda > 0$

$$p(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}([0, +\infty))$$

Пример 3.3. Распределение Коши

$$p(x) = \frac{1}{\pi(1 + x^2)}$$

Пример 3.4. Распределение Парето с параметром $a > 0$

$$p(x) = \frac{1}{ax^{a+1}} \cdot \mathbb{1}([1, +\infty))$$

Пример 3.5. Гауссовское (нормальное) распределение $\mathcal{N}(a, \sigma)$ с параметрами $a \in \mathbb{R}$ и $\sigma > 0$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Определение 3.3. $\mathcal{N}(0, 1)$ – стандартное нормальное распределение.

3.2 Пополнение вероятностного пространства. Мера Лебега

Определение 3.4. Пусть $(\Omega, \mathcal{F}, P)$ – вероятностное пространство. Пусть $\mathcal{N} = \{B \subset \Omega : \exists A : A \supset B, P(A) = 0\}$. Тогда пополнением $\mathcal{F}$ назовём $\bar{\mathcal{F}} = \{A \cup B : A \in \mathcal{F}, B \in \mathcal{N}\}$. Это то же самое, что и $\sigma\{\mathcal{F}, \mathcal{M}\}$. Пополнением вероятностного пространства назовём $(\Omega, \bar{\mathcal{F}}, \bar{P})$, где $\bar{P}$ определена на $\bar{\mathcal{F}}$ следующим образом: $\bar{P}(C) = P(A)$, если $C = A \cup B$, где $A \in \mathcal{F}, B \in \mathcal{N}$.

Упражнение. Проверить корректность определения, т.е. что если
$A, A' \in \mathcal{F}, B, B' \in \mathcal{N}$ и $A \cup B = A' \cup B'$, то $\bar{P}(A \cup B) = \bar{P}(A' \cup B')$.

Определение 3.5. Мера μ на $(S, \mathcal{B})$ называется σ-конечной, если существует счётный набор множеств $\{S_n\}$ такой, что $\mu(S_n) \leq \infty$ и $S = \bigcup\limits_{n=1}^{\infty} S_n$.

Определение 3.6. Пусть $B \in \mathcal{B}(\mathbb{R})$. Тогда $B = \bigcup\limits_{n \in \mathbb{Z}} (B \cap (n, n+1])$. Введём μ_n - равномерное распределение на $(n, n+1]$. Тогда определим $\mu(B) = \sum\limits_{n \in \mathbb{Z}} \mu_n(B \cap (n, n+1])$. Это мера Лебега.

3.3 Условная вероятность

Определение 3.7. Пусть $(\Omega, \mathcal{F}, P)$ – вероятностное пространство, $P(B) > 0$. Условной вероятностью события A при условии B назовём

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

Теорема 3.3. *(Формула полной вероятности) Пусть* $\Omega = B_1 \sqcup \ldots \sqcup B_n \sqcup \ldots$ *Тогда* $P(A) = \sum_j P(AB_j) = \sum_j P(A|B_j)P(B_j)$.

4 Независимость событий

Пример 4.1. Студент из N билетов выучил n. Перед ним k человек. Какова вероятность, что студент вытащил "хороший" билет (то есть тот, который выучил)?

$A = \{$студент вытащил хороший билет$\}$

Ответ: $P(A) = \frac{n}{N}$.

Теорема 4.1. *Формула Байеса*

$\Omega = \bigsqcup\limits_i B_i$, $P(B_i) > 0\ \forall i \implies \forall k\ P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum\limits_i P(A|B_i)P(B_i)}$

Доказательство. следствие формулы полной вероятности:

$$P(B_k|A) = \frac{P(AB_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum\limits_i P(A|B_i)P(B_i)}$$

□

Пример 4.2. При рентгеновском обследовании вероятность обнаружить заболевание N у больного равна 0.95, вероятность принять здорового человека за больного равна 0.05. Доля больных по отношению ко всему населению равна 0.01. Найти вероятность того, что человек здоров, если он был признан больным при обследовании.

Решение. Предположим, что:
$A =\{$человек здоров$\}$, $B =\{$человек болен$\}$,
$B_1 = \{$обнаружили заболевание$\}$, $B_2 =\{$не обнаружили заболевание$\}$.
Тогда из условия получаем:

$$P(B_1|B) = 0.95,\ P(B_1|A) = 0.05,\ P(B) = 0.01,\ P(A) = 0.99.$$

Вычислим полную вероятность признания больным:

$$P(B_1) = P(B_1|A)P(A) + P(B_2|B)P(B) = 0.99 \cdot 0.05 + 0.01 \cdot 0.95 = 0.059.$$

Вероятность «здоров» при диагнозе «болен» можно посчитать, применив формулу Байеса:

$$P(A|B_1) = \frac{P(A)P(B_1|A)}{P(A)P(B_1|A) + P(B)P(B_1|B)} = \frac{0.99 \cdot 0.05}{(0.99 \cdot 0.05 + 0.01 \cdot 0.95} = 0.839.$$

Таким образом, 83.9% людей, у которых обследование показало результат «болен», на самом деле здоровые люди. Удивительный результат возникает по причине значительной разницы в долях больных и здоровых. Болезнь N — редкое явление, поэтому и возникает такой парадокс Байеса. При возникновении такого результата лучше всего сделать повторное обследование.

Определение 4.1. Событие A *не зависит* от B, если $P(A|B) = P(A)$ (предполагается, что $P(B) \neq 0$).

Определение 4.2. События A и B *независимые* (или A не зависит от B и B не зависит от A), если $P(AB) = P(A)P(B)$.

Утверждение 4.1. *A не зависит от B тогда и только тогда когда A и B независимы или B невероятно.*

Доказательство.

$$\begin{array}{rcl} P(A|B) & = & P(A) \\ P(A|B) & = & \frac{P(AB)}{P(B)} \end{array} \iff P(A) = \frac{P(AB)}{P(B)} \iff P(AB) = P(A)P(B)$$

□

Пример 4.3. Из колоды, содержащей 36 карт, наудачу вынимается одна карта. Пусть $A = \{$достали туз$\}$, $B = \{$достали карту трефовой масти $\clubsuit\}$. Можно ли утверждать, что данные события независимы?

Решение. Из условия имеем: $|\Omega| = 36$, $|A| = 4$, $|B| = 9$, $|AB| = 1$.

$$P(A) = \frac{4}{36} = \frac{1}{9},\ P(B) = \frac{9}{36} = \frac{1}{4},\ P(AB) = \frac{1}{36}$$

Получаем, что $P(AB) = P(A)P(B) \implies$ по определению события независимы.

Определение 4.3. События A_z, $z \in T$ называются *попарно независимыми*, если $\forall s \neq t \in T$ $P(A_sA_t) = P(A_s)P(A_t)$.

Определение 4.4. События A_z, $z \in T$ называются *независимыми в совокупности*, если $\forall U \subset T$ $(|U| < |\infty|) : P(\bigcap\limits_{t \in U} A_t) = \prod\limits_{t \in U} P(A_t)$.

Пример 4.4. События попарно независимы, но зависимы с совокупности: Пусть $\Omega = \{1, 2, 3, 4\}$, $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{1, 4\}$.
Тогда $P(A_i) = \frac{2}{4} = \frac{1}{2}$ $\forall i = 1, \ldots, 3$ и $P(A_iA_j) = \frac{1}{4}$, $i \neq j$.

Получаем, что $P(A_iA_j) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_i)P(A_j) \implies$ события попарно независимы.
$P(A_1A_2A_3) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3) \implies$ события зависимы в совокупности.

Пример 4.5. Пример Бернштейна

Рассмотрим правильный тетраэдр, три грани которого окрашены соответственно в красный, синий, зелёный цвета, а четвёртая грань содержит все три цвета.
$A_1 = \{$выпала грань, содержащая красный цвет$\}$,
$A_2 = \{$выпала грань, содержащая синий цвет$\}$,
$A_3 = \{$выпала грань, содержащая зеленый цвет$\}$.

Лемма 4.1. *Пусть $A_1, \dots, A_n$ $(n \geq 2)$ независимы в совокупности и $B_k = A_k$ или A_k^C при $k = 1, \dots, n$. Тогда $B_1, \dots, B_n$ - независимые события.*

Доказательство. Достаточно рассмотреть случай, когда $B_i = A_i^C$ и $B_k = A_k \, \forall k \neq i$. Проверим определение:
$U \subseteq \{1, \dots n\}$
$S := \{1, \dots n\} \setminus U$
$i \notin U \implies P(\bigcap\limits_{j \in U} B_j) = \prod\limits_{j \in U} P(B_j)$, потому что каждое $B_i = A_i$, а A_i - независимы в совокупности.
$i \in U \implies P(\bigcap\limits_{j \in U} B_j) = P(\bigcap\limits_{j \in S} A_j) - P(A_i \cap (\bigcap\limits_{j \in S} A_j)) = P(\bigcap\limits_{j \in S} A_j) - P(A_i)P(\bigcap\limits_{j \in S} A_j)) = \prod\limits_{j \in S} P(A_j) - P(A_i) \prod\limits_{j \in S} P(A_j) = (1 - P(A_i)) \prod\limits_{j \in S} P(A_j) = P(A_i^C) \prod\limits_{j \in S} P(A_j) = \prod\limits_{j \in U} P(B_j)$ □

Теорема 4.2. *Формула Эйлера из теории чисел*

Пусть $\varphi : \mathbb{N} \to \mathbb{N}$, $n \mapsto \varphi(n)$ - количество чисел $k \in \{1, \dots, n\}$, которые взаимно простых с n. Тогда $\varphi(n) = n \prod\limits_{p \in J_n} (1 - \frac{1}{p})$, где J_n - множество всех простых делителей n (другими словами, если $n = p_1^{k_1} \dots p_m^{k_m}$, где простые числа $p_1 < \dots < p_m$ и $k_i \geq 1$ $i = 1, \dots, m$, $J_n = \{p_1, \dots, p_m\}$.

Доказательство. Введём $\Omega = \{1, \dots, n\}$, $\mathcal{F} = 2^\Omega$, пусть выполнено классическое определение вероятности $P(\{i\}) = \frac{1}{n}$ при $i = 1, \dots, n$ (можно сказать, что рассматривается случайный эксперимент, в котором наудачу выбирается одно число из множества от 1 до n).
Пусть событие $A = \{$выбранное число взаимно просто с $n\}$. Тогда $P(A) = \frac{\varphi(n)}{n}$.
Введём событие $A_i =\{$выбранное число делится на $p_i\} = \{p_i, 2p_i, \dots, \frac{n}{p_i} p_i\}$, $|A_i| = \frac{n}{p_i}$, поэтому $P(A_i) = \frac{\frac{n}{p_i}}{n} = \frac{1}{p_i}$ для $i = 1, \dots, m$.
Тогда $A_{i_1} \dots A_{i_k} = \{p_{i_1} \dots p_{i_k}, 2p_{i_1} \dots p_{i_k}, \dots, \frac{p_{i_1} \cdots p_{i_k}}{n} p_{i_1} \dots p_{i_k}\}$ $(1 \leq i_1 < \dots < i_k \leq m) \implies P(A_{i_1} \dots A_{i_k}) = \frac{1}{p_{i_1} \cdots p_{i_k}} = P(A_{i_k}) \dots P(A_{i_k})$, то есть события $A_{i_1}, \dots, A_{i_k}$ - независимы.

$$P(A) = P((\bigcup_{i=1}^m A_i)^C) = P(\bigcap_{i=1}^m A_i^C) \overset{\text{по лемме}}{=} \prod_{i=1}^m P(A_i^C) = \prod_{i=1}^m (1 - P(A_i)) = \prod_{i=1}^m (1 - \frac{1}{p_i})$$

□

Определение 4.5. Пусть даны системы событий $\mathcal{A}_t \subset \mathcal{F}$, $t \in T$, где T некоторое множество $|T| \geq 2$. Эти *системы независимы в совокупности*, если для любого конечного $U \subset T$ независимы события $A_t \in \mathcal{A}_t$, $t \in U$.

Теорема 4.3. *Пусть независимы* $\pi-$*системы* $\mathcal{A}_t \subset \mathcal{F}$, $t \in T$, *где* $\mathcal{F}$ *- это* $\sigma-$*алгебра. Тогда независимы* $\sigma\{\mathcal{A}_t\}$ $\sigma-$*алгебры, порождённые этими системами.*

Доказательство. Возьмём $\forall\{t_1, \dots, t_n\} \subset T$ и $A_{t_k} \in \mathcal{A}_{t_k}$, $k = 1, \dots, n$. Введём систему $\mathcal{K}_{t_1}$, состоящую из событий B таких, что $P(BA_{t_2} \dots A_{t_n}) = P(B)P(A_{t_2}) \dots P(A_{t_m})$.

Легко увидеть, что $\mathcal{K}_{t_1}$ - это λ-система:

- $\Omega \in \mathcal{K}_{t_1}$;
- $A, B \in \mathcal{K}_{t_1}, A \subseteq B \overset{?}{\Longrightarrow} B \setminus A \in \mathcal{K}_{t_1}$
  $P(B \dots A_{t_n}) - P(A \dots A_{t_m}) = (P(B) - P(A))P(A_{t_2}) \dots P(A_{t_n}) =$
  $= P(B \setminus A)P(A_{t_2}) \dots P(A_{t_n}) = P((B \setminus A)A_{t_2} \dots A_{t_n})$;
- Следует из непрерывности меры.

По теореме о π-λ - системах получаем, что $\sigma\{\mathcal{A}_{t_1}\} \subset \mathcal{K}_{t_1}$. Следовательно, соотношение $P(BA_{t_2} \dots A_{t_n}) = P(B)P(A_{t_2}) \dots P(A_{t_m})$ верно для $\forall B \in \sigma\{\mathcal{A}_{t_1}\}$.

Продолжая аналогичным образом, приходим к равенству $P(A_{t_1} \dots A_{t_n}) = P(A_{t_1}) \dots P(A_{t_n})$ для $\forall A_{t_k} \in \sigma\{\mathcal{A}_{t_k}\}$. □

Лемма 4.2. *(о группировке)*

Пусть $\mathcal{A}_t, t \in T-$ *семейство независимых* σ*-алгебр (то есть* $\forall t\, \mathcal{A}_t \subset \mathcal{F}$*). Возьмём* $\Lambda \subset T$ *такое, что* $\Lambda = \bigsqcup\limits_{\alpha \in \Gamma} \Lambda_\alpha$*. Определим* $\mathcal{F}_\alpha = \sigma\{\mathcal{A}_t,\, t \in \Lambda_\alpha\}$*. Тогда* $\mathcal{F}_\alpha$, $\alpha \in \Gamma$ *- семейство независимых* σ*-алгебр.*

Доказательство. $\forall n \in \mathbb{N}\ \forall \alpha_1, \dots, \alpha_n \in \Gamma$
$B_{\alpha_1} \in \mathcal{F}_{\alpha_1}, \dots, B_{\alpha_n} \in \mathcal{F}_{\alpha_n}$
$P(B_{\alpha_1} \dots B_{\alpha_n}) \overset{?}{=} P(B_{\alpha_1}) \dots P(B_{\alpha_n})$

Пусть $\mathcal{F}_{\alpha_1} = \sigma\{\mathcal{A}_t,\, t \in \Lambda_{\alpha_1}\}$, тогда $\mathcal{F}_{\alpha_1} = \sigma\{\overbrace{\mathcal{A}_{t_1} \cap \dots \cap \mathcal{A}_{t_m}}^{\text{система из таких - } \pi\text{-система (обозн. } M_{\alpha_1})} | \forall t_1, \dots, t_m \in \Lambda_{\alpha_1}\}$

$P(\overbrace{A_{t_1} \dots A_{t_n}}^{\in M_{\alpha_1}} \overbrace{B_{s_1} \dots B_{s_q}}^{\in M_{\alpha_2}} \dots) = P(A_{t_1}) \dots P(A_{t_m})P(B_{s_1}) \dots P(B_{s_q}) \dots =$
$= P(A_{t_1} \dots A_{t_m})P(B_{s_1} \dots B_{s_q}) \dots \Longrightarrow \{M_{\alpha_i}\}-$ независимые π-системы $\Longrightarrow$ независимое семейство σ-алгебр, порожденных π-системами.

□

Лемма 4.3. *(Борель-Кантели)*

Пусть A_n *- последовательность событий*

1. *Если* $\sum P(A_k) < \infty$, *то* $P($*произойдёт бесконечное число* A_k*-ых*$) = 0$
   (или $P(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k \geq n}^{\infty} A_k) = P(\limsup A_k)$*).*
2. *Если* $A_1, A_2, \dots$ *- независимы, то* $P($*произойдёт бесконечное число* A_k*-ых*$) = 1$,

Доказательство. 1. $A \subset \bigcup\limits_{k \geq n} A_k \, \forall n$, $A = \limsup A_k$

$0 \leq P(A) \leq P(\bigcup\limits_{k \geq n} A_k) \leq \sum\limits_{k=n}^{\infty} P(A_k) \to 0$

2. Проверим, что $P((\limsup A_k)^C) = 0$

   $A^C = (\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k \geq n}^{\infty} A_k)^C = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k \geq n}^{\infty} A_k^C = \liminf A_k^C$

   $P(A^C) \leq \sum\limits_{n=1}^{\infty} P(\bigcap\limits_{k \geq n} A_k^C)$ (субаддитивность)

   Фиксируем $n \in \mathbb{N}$. $P(\bigcap\limits_{k=n}^{N} A_k^C) \xrightarrow{N \to \infty} P(\bigcap\limits_{k=n}^{\infty} A_k^C)$ по непрерывности вероятностной меры.

   $P(\bigcap\limits_{k=n}^{N} A_k^C) = \prod\limits_{k=n}^{N} P(A_k^C)$ - из леммы $\{A_n, \ldots, A_N\}$ - независимы $\implies$ $\{A_n^C, \ldots, A_N^C\}$ - независимы.

   $P(\bigcap\limits_{k=n}^{N} A_k^C) = \prod\limits_{k=n}^{N} P(A_k^C) = \prod\limits_{k=n}^{N} (1 - P(A_k)) \overset{1-x \leq e^x}{\leq} \prod\limits_{k=n}^{N} e^{-P(A_k)} =$

   $= e^{-\sum\limits_{k=n}^{N} P(A_k)} \xrightarrow{N \to \infty} 0$ (так как ряд расходится).

□

5 Случайные величины и их распределения

5.1 Случайные элементы

Определение 5.1. Пусть $(V, \mathcal{A})$, $(S, \mathcal{B})$ - некоторые измеримые пространства. Отображение $f : V \to S$ называется *$\mathcal{A}|\mathcal{B}$-измеримым*, если $\forall B \in \mathcal{B} \, f^{-1}(B) \in \mathcal{A}$, то есть $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}$.

Если $(V, \mathcal{A}) = (\Omega, \mathcal{F})$ и $X \in \mathcal{F}|\mathcal{B}$, то X называют *случайным элементом.*

Если $S = \mathbb{R}^n$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^n)$, $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, то X называется

1. $n = 1$ при *случайной величиной*;

2. $n > 1$ при *случайным вектором.*

Пример 5.1. Индикатор $\mathbb{1}_A(x) = \begin{cases} 1, \, x \in A \\ 0, \, x \notin A \end{cases}$ - случайная величина, так как $\mathbb{1}^{-1}(\{1\}) = A \in \mathcal{F}$ и $\mathbb{1}^{-1}(\{0\}) = A^C \in \mathcal{F}$.

Лемма 5.1. *Пусть $f : V \to S$. Возьмём в S произвольную систему подмножеств M и рассмотрим в V систему $f^{-1}(M)$. Введём $\mathcal{B} := \sigma\{M\}$, $\mathcal{A} := \sigma\{f^{-1}(M)\}$. Тогда $f \in \mathcal{A}|\mathcal{B}$. Более того, $\sigma\{f^{-1}(M)\} = f^{-1}(\sigma\{M\})$.*

Доказательство.

$\boxed{\supset}$ Определим $\mathcal{D} := \{B \subset S : f^{-1}(B) \in \mathcal{A}\}$. Очевидно, что это σ-алгебра (так как $\mathcal{A}$ - σ-алгебра). По построению имеем $M \subseteq \mathcal{D}$ (поскольку $f^{-1}(B) \in \mathcal{A}$) $\implies \sigma\{M\} \subset \mathcal{D}$.

Поэтому $f^{-1}(\mathcal{B}) = f^{-1}(\sigma\{M\}) \subset f^{-1}(\mathcal{D}) \subset \mathcal{A} = \sigma\{f^{-1}(M)\}$.

$\boxed{\subset}$ Заметим, что $f^{-1}(M) \subset f^{-1}(\mathcal{B})$, поскольку $M \subset \mathcal{B}$. Таким образом, $\sigma\{f^{-1}(M)\} \subset \sigma\{f^{-1}(\mathcal{B})\} = f^{-1}(\mathcal{B}) = f^{-1}(\sigma\{M\})$ □

Следствие 5.1. *Пусть $f^{-1}(M) \subset \mathcal{C}$, где $\mathcal{C}$ – σ-алгебра. Тогда $f^{-1}(\sigma\{M\}) \subset \mathcal{C}$*

Доказательство. Действительно, $f^{-1}(M) \subset \mathcal{C} \implies \sigma\{f^{-1}M\} \subset \mathcal{C}$. Тогда $f^{-1}(\sigma\{M\}) = \sigma\{f^{-1}(M)\} \subset \mathcal{C}$. □

Для проверки $\mathcal{F}|\mathcal{B}$-измеримости f достаточно рассматривать прообразы лишь любой системы, порождающей $\mathcal{B}$.

Следствие 5.2. *Функция $X : \Omega \to \mathbb{R}$ является случайной величиной $\iff$ выполнено любое из следующих условий*

1. $\{\omega : X(\omega) \leq x\} \in \mathcal{F}\ \forall x \in \mathbb{R}$;
2. $\{\omega : X(\omega) < x\} \in \mathcal{F}\ \forall x \in \mathbb{R}$;
3. $\{\omega : a < X(\omega) < b\} \in \mathcal{F}\ \forall a, b \in \mathbb{R}\ (-\infty < a < b < \infty)$.

Доказательство. Вспомним, что $\mathcal{B}(\mathbb{R})$ порождается любой из систем вида $\{(-\infty, x], x \in \mathbb{R}\}$, $\{(-\infty, x), x \in \mathbb{R}\}$, $\{(a, b), x \in \mathbb{R}\}$. Достаточно требовать, чтобы прообразы множеств системы входили в σ-алгебру $\mathcal{F}$, что и означает, что X - случайная величина. □

Определение 5.2. Пусть V, S - топологические пространства, снабжённые соответственно системами открытых множества ν и τ. Отображение $f : V \to S$ называется *непрерывным*, если $f^{-1}(\tau) \subset \nu$.

Утверждение 5.1. *Пусть $f : V \to S$ непрерывное отображение топологических просторанств, тогда $f \in \mathcal{B}(V)|\mathcal{B}(S)$*

Доказательство. В силу непрерывности f имеем: $f^{-1}(\tau) \subset \nu \subset \sigma\{\nu\} = \mathcal{B}(V)$, а $\mathcal{B}(S) = \sigma\{\tau\}$. По следствию 1 получаем необходимое. □

Определение 5.3. Функция $f : V \to S$ называется *борелевской*, если $f \in \mathcal{B}(V)|\mathcal{B}(S)$.

Итак, любая непрерывная является борелевской. В частности, если $f : \mathbb{R}^m \to \mathbb{R}^n$ - непрерывна, то f - борелевская.

Определение 5.4. Функция $X : \Omega \to \overline{\mathbb{R}}$ называется *расширенной случайной величиной*, если $X \in \mathcal{F}|\mathcal{B}(\overline{\mathbb{R}})$, где $\overline{\mathbb{R}} = [-\infty, +\infty]$ и $\mathcal{B}(\overline{\mathbb{R}}) := \{B, B \cup \{+\infty\}, B \cup \{-\infty\}, B \cup \{+\infty\} \cup \{-\infty\} | B \in \mathcal{B}(\mathbb{R})\}$

Замечание 5.1. $\mathcal{B}(\overline{\mathbb{R}})$ порождается системой $\{[-\infty, x], x \in \mathbb{R}\}$ или $\{[-\infty, x), x \in \mathbb{R}\}$. Поэтому $X : \Omega \to \overline{\mathbb{R}}$ является расширенной случайной величной $\iff$ выполнено любое из следующих условий:

1. $\{\omega : X(\omega) \leq x\} \in \mathcal{F}\ \forall x \in \mathbb{R}$;
2. $\{\omega : X(\omega) < x\} \in \mathcal{F}\ \forall x \in \mathbb{R}$.

Любая случайная величина является расширенной случайной величиной.

5.2 Операции со случайными величинами

Теорема 5.1. *Пусть $(X_n)_{n\in\mathbb{N}}$ - последовательность (расширенных) случайных величин $X_n : \Omega \to \mathbb{R}$. Тогда (вообще говоря, расширенными) случайными величинами являются:*

1. $\sup\limits_n X_n$, $\inf\limits_n X_n$;
2. $\limsup\limits_n X_n$, $\liminf\limits_n X_n$;
3. $\lim X_n$, *если для* $\forall\omega \in \Omega$ *предел существует.*

Доказательство.

1. Достаточно убедиться, что $\forall x \in \mathbb{R}\ \{\omega : \sup\limits_n X_n(\omega) \le x\} \in \mathcal{F}$. Заметим, что $\{\omega : \sup\limits_n X_n(\omega) \le x\} = \bigcap\limits_n \{\omega : X_n(\omega) \le x\}$ и каждое $\{\omega : X_n(\omega) \le x\} \in \mathcal{F}$. Аналогично и для inf: $\{\omega : \inf\limits_n X_n(\omega) < x\} = \bigcup\limits_n \{\omega : X_n(\omega) < x\}$, где $\{\omega : X_n(\omega) < x\} \in \mathcal{F}$.
2. Заметим, что $\limsup\limits_n X_n = \inf\limits_n \sup\limits_{k\ge n} X_k$. По доказанному первому пункту $\sup\limits_{k\ge n} X_k$ является случайной величиной, а также inf от случайный величин - случайная величина.
3. Если $\exists X = \lim\limits_n X_n$ на Ω, то $\lim\limits_n X_n(\omega) = \limsup\limits_n X_n(\omega) = \liminf\limits_n X_n(\omega)$, а по доказанному второму пункту $\limsup\limits_n X_n(\omega)$ и $\liminf\limits_n X_n(\omega)$ являются случайными величинами.

□

Лемма 5.2. *Пусть имеются $(\mathcal{S}_k, \mathcal{B}_k)$ - измеримые пространства, отображение $f_k : \mathcal{S}_k \to \mathcal{S}_{k+1}$ при $k = 1, \dots, n-1$. Тогда $f := f_n \circ \dots \circ f_1$ будет $\mathcal{B}_1|\mathcal{B}_n$-измеримой.*

Доказательство. Достаточно рассмотреть случай $n = 3$.

$\forall B \in \mathcal{B}_3$ имеем $f^{-1}(B) = (f_2 \circ f_1)^{-1}(B) = f_1^{-1}(f_2^{-1}(B))$, где $f_2^{-1}(B) \in \mathcal{B}_2$, а значит $f_1^{-1}(f_2^{-1}(B)) \in \mathcal{B}_1$ □

Лемма 5.3. *Пусть $(\Omega, \mathcal{F})$ и $(\mathcal{S}_k, \mathcal{B}_k)$, где $k = 1, \dots, n$, - измеримые пространства. Рассмотрим $X_k : \Omega \to \mathcal{S}_k$. Отображение $X = (X_1, \dots, X_n) : \Omega \to \mathcal{S} = \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ будет $\mathcal{F}|\mathcal{B}$, где $\mathcal{B} = \sigma\{$брусы вида $B_1 \times \dots \times B_n$, $B_i \in \mathcal{B}_i\}$ $\iff$ $X_k \in \mathcal{F}|\mathcal{B}_k$ для $k = 1, \dots, n$*

Доказательство. Введём отображения "проектирования" $\pi_k : \mathcal{S} \to \mathcal{S}_k$, где $\pi_k(x_1, \dots, x_n) = x_k$. $\pi_k \in \mathcal{B}|\mathcal{B}_k$, так как прямоугольники порождают $\mathcal{B}$ и $\pi_k(B) = B_k \in \mathcal{B}_k$.

$\boxed{\Longrightarrow}$ Пусть $X = (X_1, \dots, X_n) \in \mathcal{F}|\mathcal{B}$. Тогда $X_k = \pi_k \circ X \in \mathcal{F}|\mathcal{B}_k$ - как композиция измеримых.

$\boxed{\Longleftarrow}$ Пусть $X_k \in \mathcal{F}|\mathcal{B}_k$ для $k = 1, \dots, n$. Тогда для $\forall$ прямоугольников $B = B_1 \times \dots B_n$ имеем $X^{-1}(B) = X_1^{-1}(B_1) \cap \dots \cap X_n^{-1}(B_n) \in \mathcal{F}$.

□

Теорема 5.2. *Пусть X и Y - случайные величины, то $X+Y$, $X-Y$, XY - случайные величины. Если $Y(\omega) \neq 0 \ \forall \omega \in \Omega$, то $\frac{X}{Y}$ - случайная величина.*

Доказательство.

- По лемме 5.3 отображение $(X, Y) : \Omega \to \mathbb{R}^2$ - случайная величина.
- Функция $h(x, y) := x+y$ непрерывна на $\mathbb{R}^2$, следовательно $h \in \mathcal{B}(\mathbb{R}^2)|\mathcal{B}(\mathbb{R})$. По лемме 5.2 отображение $X + Y = h(X, Y) \in \mathcal{F}|\mathcal{B}(\mathbb{R})$.
- Остальные случаи рассматриваются аналогично.

□

5.3 Распределение случайного элемента

Определение 5.5. Пусть $X : \Omega \to S$ - случайный элемент, то есть $\mathcal{F}|\mathcal{B}$-измеримое отображение. *Распределением (или законом распределения)* называется мера на пространстве $(S, \mathcal{B})$, задаваемая формулой $P_X(B) := P(X^{-1}(B))$, $B \in \mathcal{B}$.

Иногда используется следующее обозначение: $law(X)$.

Определение 5.6. *Функция распределения случайного вектора* $X : \Omega \to \mathbb{R}^n$ называется функция $F_X(x) := P(X \in (-\infty, x]) = P(x \in (-\infty, x_1] \times \cdots \times (-\infty, x_n]) = P(X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n)$.

Замечание 5.2. Пусть $(S, \mathcal{B})$-измеримое пространство, снабжённое вероятностной мерой Q. Тогда на некотором $(\Omega, \mathcal{F}, P)$ существует случайный элемент $X : \Omega \to S$ такой, что $Law(X) = Q$

Доказательство. Достаточно взять $(\Omega, \mathcal{F}, P) = (S, \mathcal{B}, Q)$ и $X := I$ - тождественное отображение ($I(\omega) = \omega$). □

Для случайного элемента $X : \Omega \to S$ ($X \in \mathcal{F}|\mathcal{B}$) введём σ-алгебру $\sigma\{X\} \subset \mathcal{F}$, где $\sigma\{X\} = \{X^{-1}(B) | \, B \in \mathcal{B}\} = X^{-1}(\mathcal{B})$.

Определение 5.7. Говорят, что случайные элементы $X_t : \Omega \to S_t$, где $(S_t, \mathcal{B}_t)$– измеримые пространства и $t \in T$, *независимы (в совокупности)*, если независимы порождённые ими σ-алгебры. Другими словами, если $\forall$ конечного множества $J \subset T$ и всех $B_t \in \mathcal{B}_t, t \in J$ выполняется $P(\bigcap\limits_{t \in J} \{X_t \in B_t\}) = \prod\limits_{t \in J} P(X_t \in B_t)$.

Следствие 5.3. *$X = (X_1, \ldots, X_n)$ – случайная величина, имеющая независимые компоненты $\iff F_X(x_1, \ldots, x_n) = \prod\limits_{k=1}^{n} F_{X_k}(x_k)$.*

Доказательство. Следует из утверждения про независимость σ-алгебр, порождённых π-системами. □

Теорема 5.3. *(Ломницкий - Улам)*
Пусть $\{(S_t, \mathcal{B}_t, Q_t), \, t \in T\}$ - произвольное семейство вероятностных пространств. Тогда на некотором $(\Omega, \mathcal{F}, P)$ существует семейство независимых случайных элементов $\{X_t, \, t \in T\}$ ($X_t : \Omega \to S_t, X_t \in \mathcal{F}|\mathcal{B}_t$) таких, что $P_{X_t} = Q_t$, $t \in T$.

Всегда можно на вероятностном пространстве построить семейство независимых случайных элементов с заданными распределениями.

Доказательство. Запись обсуждения этого с консультации вот тут.

Вводим $\prod\limits_{t\in T}\mathcal{B}_t$ и это есть наименьшая σ-алгебра, содержащая все цилиндрические множества (множества вида $B_{t_1}\times\ldots B_{t_n}$). Задаём меру на этих прямоугольниках как произведение мер (в силу независимости).

σ-алгебра, порождённая прямоугольниками, устроена следующим образом: если множество входит в это произведение, то найдётся счётная последовательность точек $t_1,\ldots,t_n,\, t_i\in T$, то множество войдет в счетное произведение σ-алгебр.

От мер на прямоугольниках нужно будет перейти к счётным произведениям и заодно понять, почему так заданная мера продолжается на σ-алгебру, описанную выше.

Примерно работа на полтора часа :))))) $\square$

6 Теорема Пуассона

6.1 Следствия леммы о группировке для семейства независимых случайных элементов

Следствие 6.1 (лемма о группировке)**.** *Пусть $\{X_t,\, t\in T\}$ – семейство независимых случайных элементов, где $X_t\colon\Omega\to S_t,\ X_t\in\mathcal{F}|\mathcal{B}_t$ для любого $t\in T$. Если $I(\alpha),\ \alpha\in\Lambda,$ – некоторое семейство попарно непересекающихся подмножеств в T, то системы $\{X_t,\, t\in I(\alpha)\},\ \alpha\in\Lambda,$ независимы.*

Следствие 6.2. *Пусть выполнены условия предыдущего следствия. Тогда измеримые (должным образом) функции от попарно непересекающихся наборов случайных элементов являются независимыми.*

Пример 6.1. Пусть независимы величины $X_1,\ \ldots,\ X_7$. Тогда будут независимыми $Y_1=(X_1,\, X_3),\ Y_2=X_2+e^{-X_5}+\sin X_7$, поскольку $I_1=\{1,\, 3\},\ I_2=\{2,\, 5,\, 7\}$ попарно непересекаются.

Замечание 6.1. Если $X_1,\ \ldots,\ X_n$ – независимые случайные величины, то величины $X_1+\ldots+X_{n-1}$ и X_n независимы.

6.2 Классическая теорема Пуассона

Теорема 6.1. *Пусть $p=p(n)>0,\ n\in\mathbb{N}$, причём $np(n)\to\lambda>0$ при $n\to\infty$. Тогда для всякого $m\in\mathbb{Z}_+:=\{0\}\cup\mathbb{N}$ верно соотношение*

$$C_n^m p(n)^m(1-p(n))^{n-m}\to\frac{\lambda^m}{m!}e^{-\lambda},\quad n\to\infty.$$

Доказательство. Имеем

$$C_n^m p(n)^m(1-p(n))^{n-m}=\frac{1}{m!}\frac{(n-m+1)\ldots n}{n^m}\cdot \\ \cdot(np(n))^m(1-p(n))^n(1-p(n))^{-m}.$$

По условию $np(n)\to\lambda,\ n\to\infty$. Из этого немедленно следует, что

$$(1-p(n))^{-m}\to 1,\quad n\to\infty.$$

Кроме того, справедливы соотношения

$$\frac{(n-m+1)\dots n}{n^m} = \left(1 - \frac{m-1}{n}\right)\dots\left(1 - \frac{1}{n}\right) \to 1, \quad n \to \infty,$$

и

$$(1-p(n))^n = \exp\{n \log(1-p(n))\} = \\ = \exp\left\{n\left(-p(n) + O\left(p(n)^2\right)\right)\right\} \to e^{-\lambda}, \quad n \to \infty.$$

В итоге получаем

$$C_n^m p(n)^m (1-p(n))^{n-m} \to \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \to \infty.$$

□

Переформулируем теорему Пуассона в вероятностных терминах. Предположим, что для любого $n \in \mathbb{N}$ на некотором вероятностном пространстве $(\Omega,\, \mathcal{F},\, P)$ задан набор случайных величин $X_{n,\,1},\, \dots,\, X_{n,\,n}$ таких, что

$$P(X_{n,\,k} = 1) = p(n), \quad P(X_{n,\,k} = 0) = 1 - p(n)$$

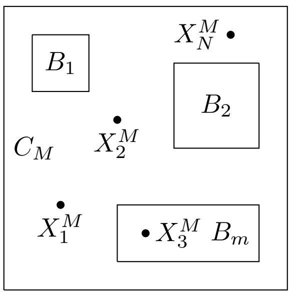
для всех $k = 1,\, \dots,\, n$. Положим $S_n := X_{n,\,1} + \ldots + X_{n,\,n}$. Тогда теорема Пуассона утверждает, что для любого $m \in \mathbb{N}$

$$P(S_n = m) \to P(Y = m), \quad n \to \infty,$$

где $np(n) \to \lambda$ при $n \to \infty$ и $Y \sim \mathrm{Pois}\,(\lambda)$.

6.3 Пространственный пуассоновский процесс

В пространстве $\mathbb{R}^d$ возьмём ограниченные борелевские множества $B_1,\, \dots,\, B_m$ такие, что $B_i \cap B_j = \varnothing$ при $i \neq j$, и положим $C_M := [-M/2,\, M/2]^d,\ M \in \mathbb{N}$. Ясно, что при достаточно большом M все $B_1,\, \dots,\, B_m$ будут содержаться в C_M. Поэтому далее считаем, что число M подобранно так, что $B_1,\, \dots,\, B_m \subset C_M$. Пусть случайные величины (точки) $X_1^M,\, \dots,\, X_N^M,\ N \in \mathbb{N}$, независимы и равномерно распределённы в кубе C_M.



Число точек, которые попали в некоторое борелевское множество $B \subset C_M$, можно подсчитать следующим образом:

$$Y_B^M := \sum_{k=1}^{N} \mathbb{1}(X_k^N \in B).$$

В силу развитой нами теории мы можем сказать, что при каждом борелевском $B \in \mathcal{B}(\mathbb{R}^d)$ индикатор множества B является измеримой функцией, а поскольку сумма и композиция измеримых функций есть измеримая функция, получаем, что Y_B^M – случайная величина.

По определению равномерного распределения

$$P(X_k^M \in B) := \frac{|B|}{|C_M|} =: p_k(M),$$

где $k = 1, \ldots, m$, а при $k = 0$ будем полагать, что

$$p_0(M) := 1 - \sum_{k=1}^{m} p_k(M).$$

Это равенство означает, что мы вводим

$$B_0 := C_M \setminus \left(\bigcup_{k=1}^{m} B_k \right).$$

Нас интересует вероятность того, что $Y_{B_1}^M = r_1, \ldots, Y_{B_m}^M = r_m$ для чисел $r_1, \ldots, r_m \in \mathbb{Z}_+$. Найдём эту вероятность:

$$\begin{aligned} P\left(Y_{B_1}^M = r_1 \ldots, Y_{B_m}^M = r_m\right) &= C_N^{r_0} C_{N-r_0}^{r_1} \ldots C_{N-r_0-r_1-\ldots-r_{m-1}}^{r_m} \cdot \\ &\cdot P\left(X_{i_1}^M \in B_0, \ldots, X_{i_{r_0}}^M \in B_0, \ldots, X_{j_1}^M \in B_m, \ldots, X_{j_{r_m}}^M \in B_m\right) = \\ &= \frac{N!}{r_0! r_1! \ldots r_m!} \left(p_0(M)\right)^{r_0} \left(p_1(M)\right)^{r_1} \ldots \left(p_m(M)\right)^{r_m}. \end{aligned}$$

где $r_0 := N-(r_1+\ldots+r_m)$ (учли, что величины $X_1^M, \ldots, X_N^M$ независимы).

Введём естественные физические условия для нашей модели. Будем считать, что с ростом M и N сохраняется плотность точек, т. е. при $M \to \infty$ и $N \to \infty$

$$\frac{N}{|C_M|} \to \lambda > 0.$$

Тогда

$$\left(N p_k(M)\right)^{r_k} = \left(\frac{N|B_k|}{C_M}\right)^{r_k} \to \left(\lambda |B_k|\right)^{r_k}, \quad n \to \infty,$$

для всех $k = 1, \ldots, m$. Кроме того, выполнено

$$\frac{N!}{r_0!} = \frac{N!}{\left(N - \sum\limits_{k=1}^{m} r_k\right)!} \sim N^{\sum\limits_{k=1}^{m} r_k}, \quad n \to \infty,$$

и

$$\left(p_0(M)\right)^{r_0} = \left(1 - \sum_{k=1}^{m} p_k(M)\right)^{N - \sum\limits_{k=1}^{m} r_k} \to e^{-\lambda(|B_1|+\ldots+|B_m|)}, \quad n \to \infty.$$

Таким образом, установлена

Теорема 6.2. *Для описанной модели случайных точек для любых попарно непересекающихся ограниченных борелевских множеств $B_1, \ldots, B_m$ и любых $r_1, \ldots, r_m \in \mathbb{Z}_+$ выполнено*

$$P\left(Y^m_{B_1} = r_1, \ldots, Y^M_{B_m} = r_m\right) \to \\ \to \frac{(\lambda|B_1|)^{r_1}}{r_1!} e^{-\lambda|B_1|} \ldots \frac{(\lambda|B_m|)^{r_m}}{r_m!} e^{-\lambda|B_m|}, \quad n \to \infty.$$

Пример 6.2. Представим себе кондитерское производство, на котором выпекаются булочки с изюмом. Есть большой чан, в котором находится тесто. Этот чан вращается и в него высыпают изюм. Далее всё тесто раскатывают, режут на булочки и выпекают. Возникает вопрос: взяв отдельную булочку, какова вероятность того, что в неё попадёт хотя бы одна изюминка? Чтобы решить поставленную задачу, воспользуемся только что доказанной теоремой.

Пусть всего было N изюминок и испекли s булочек. В условиях теоремы 6.2 возьмём $B = B_1$. Посмотрим на вероятность того, что булочка B не будет содрежать изюма. В силу теоремы 6.2 имеем

$$P(Y_B = 0) \approx e^{-\lambda|B|} = e^{-N/s},$$

где $|B|$ – объём булочки B. Таким образом, для того, чтобы булочка с вероятностью близкой к 1 содрежала изюм, нужно соотвествующим образом подобрать параметры N и s.

6.4 Оценка погрешности пуассоновской аппроксимации

Пусть $X_1, \ldots, X_n$ – независимые случайные величины такие, что

$$P(X_k = 1) = p_k, \quad P(X_k = 0) = 1 - p_k,$$

где $0 < p_k < 1$ и $k = 1, \ldots, n$ (не предполагаем, что величины одинаково распределены). Приведём без доказательства важный факт.

Теорема 6.3. *Для введённого набора случайных величин $X_1, \ldots, X_n$ справедливо неравенство*

$$\sup_{B \in \mathcal{B}(\mathbb{R})} |P(X_1 + \ldots + X_n \in B) - P(Y \in B)| \leqslant \sum_{k=1}^{n} p_k^2, \tag{6.1}$$

где $Y \sim \mathrm{Pois}(\lambda)$, $\lambda = p_1 + \ldots + p_n$.

Замечание 6.2. У этой теоремы есть три достоинства:

1) мы не предполагаем одинаковой распределённости слагаемых $X_1, \ldots, X_n$;

2) оценивается близость попадания $X_1 + \ldots + X_n$ в любое борелевское множество B, а не только в отдельную точку;

3) оценка 6.1 верна для любого конечного n.

Отметим, что из теоремы 6.3 вытекает теорема 6.1. Действительно, если $p_1 = \ldots = p_n = p(n)$ и $np(n) \to \lambda > 0$ при $n \to \infty$, то

$$\sum_{k=1}^{n} p_k^2 = n(p(n))^2 = n\left(\frac{\lambda}{n} + o\left(\frac{1}{n^2}\right)\right)^2 \to 0, \quad n \to \infty.$$

Лемма 6.1. *Пусть $Z_1,\ \ldots,\ Z_n$ – независимые случайные величины, причём $Z_k \sim \mathrm{Pois}(\lambda_k),\ \lambda_k > 0,\ k = 1,\ \ldots,\ n$. Тогда $Z_1 + \ldots + Z_n \sim \mathrm{Pois}(\lambda_1 + \ldots + \lambda_n)$.*

Доказательство. Пусть $m \in \mathbb{Z}_+$. Тогда

$$P(Z_1 + Z_2 = m) = \sum_{k=0}^{m} P(Z_1 = k,\, Z_2 = m-k) = \sum_{k=0}^{m} \frac{\lambda_1^k}{k!}\frac{\lambda_2^{m-k}}{(m-k)!}e^{-(\lambda_1+\lambda_2)} =$$
$$= \frac{1}{m!}e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{m} C_m^k \lambda_1^k \lambda_2^{m-k} = \frac{(\lambda_1+\lambda_2)^m}{m!}e^{-(\lambda_1+\lambda_2)}.$$

Тем самым для Z_1 и Z_2 утверждение верно. Осталось лишь применить индукцию к независимым случайным величинам $Z_1 + \ldots + Z_{n-1}$ и Z_n. □

Лемма 6.2. *Пусть X и Y – случайные величины. Тогда для каждого борелевского множества $B \in \mathcal{B}(\mathbb{R})$ имеет место неравенство*

$$|P(X \in B) - P(Y \in B)| \leqslant P(X \neq Y).$$

Пример 6.3. Имеется текст, содержащий $n = 10000$ знаков. Считаем, что при наборе этого текста вероятность опечатки равна $p = 0,0001$. Какова вероятность того, что число опечаток в тексте не превосходит 5? Количество опечаток может быть найдено по формуле

$$S_n = X_1 + \ldots + X_n,$$

где $X_1,\ \ldots,\ X_n$ – независимые величины, $P(X_k = 1) = p$ и $P(X_k = 0) = 1-p$ для всех $k = 1,\ \ldots,\ n$. Если воспользоваться схемой Бернулли, то точный ответ будет иметь следующий вид:

$$P(S_n \leqslant 5) = \sum_{k=0}^{5} C_n^k p^k (1-p)^{n-k} = \sum_{k=0}^{5} C_{10000}^k (0,0001)^k (0,9999)^{1000-k}.$$

Считать полученную сумму маленькое удовольствие, поэтому воспользуемся теоремой 6.3. Пусть $Y \sim \mathrm{Pois}(\lambda),\ \lambda = np = 1$. Тогда

$$|P(S_n \leqslant 5) - P(Y \leqslant 5)| \leqslant np^2 = 0,0001.$$

Так как $P(Y \leqslant 5) = 0,9996\ldots$, то мы можем утверждать, что $P(S_n \leqslant 5) \approx 0,9996$, причём погрешность составляет $0,0001$.

7 Интеграл Лебега

7.1 Построение интеграла Лебега по вероятностной мере

Пусть $(\Omega, \mathcal{F}, P)$ – вероятностное пространство, $X : \Omega \to \mathbb{R}$.

7.1.1 Интеграл для простых функций

Определение 7.1. Случайная величина X называется простой, если $X(\omega) = \sum\limits_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega)$, где $a_i \in \mathbb{R}$, а непустые A_i образуют разбиение Ω.

Определение 7.2. Интегралом Лебега от простой случайной величины X, задаваемой формулой $X = \sum\limits_{i=1}^{n} a_i \mathbb{1}(A_i)$, называется число $\int\limits_{\Omega} X dP = \sum\limits_{i} a_i P(A_i)$. В теории вероятностей интеграл Лебега по вероятностой мере принято обозначать $\mathbb{E}\, X$.

Пример 7.1. Есть N лотерейных билетов, причём на $m_1, \ldots, m_n$ из них приходятся соответственно выигрыши $a_1, \ldots, a_n$. Разыгрывается денежная сумма $S = a_1 m_1 + \ldots + a_n m_n$. Тогда средний выигрыш, приходящийся на 1 билет, равен $\frac{S}{N}$, т.е. $a_1 \frac{m_1}{N} + \ldots + a_n \frac{m_n}{N} = a_1 p_1 + \ldots + a_n p_n$, где $p_i = \frac{m_i}{N}$.

Пример 7.2. Пусть на прямой в точках $x_1, \ldots, x_n$ сосредоточены массы $p_1, \ldots, p_n$. Тогда центр тяжести этой системы $\bar{x} = \frac{x_1 p_1 + \ldots + x_n p_n}{p_1 + \ldots + p_n}$.

Пример 7.3. Если $X = c$, то $\mathbb{E}\, X = c$. Если $X = \mathbb{1}(A)$, то $\mathbb{E}\, X = P(A)$. Таким образом, для бернуллиевской сл.в. с параметром p $\mathbb{E}\, X = p$.

Утверждение 7.1. *Определение интеграла Лебега для простых случайных величин корректно, т.е. если* $X = \sum\limits_{i=1}^{n} a_i \mathbb{1}(A_i) = \sum\limits_{i=1}^{m} b_i \mathbb{1}(B_i)$, *то* $\sum\limits_{i=1}^{n} a_i P(A_i) = \sum\limits_{i=1}^{m} b_i P(B_i)$.

Доказательство.

$$
\begin{aligned}
&X = \sum_{i=1}^{n} \left(a_i \sum_{j=1}^{m} \mathbb{1}(A_i B_j) \right) = \sum_{j=1}^{m} \left(b_j \sum_{i=1}^{n} \mathbb{1}(A_i B_j) \right) = \sum_{(i,j) \in J} c_{ij} \mathbb{1}(C_{ij}), \\
&C_{ij} = A_i B_j, c_{ij} = a_i = b_j \quad \forall (i,j) : C_{ij} \neq \varnothing.
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{n} a_i P(A_i) = \sum_{i=1}^{n} \left(a_i \sum_{j=1}^{m} P(A_i B_j) \right) = \sum_{(i,j) \in J} a_i P(C_{ij}) = \\
= \sum_{(i,j) \in J} b_j P(C_{ij}) = \sum_{j=1}^{m} \left(b_j \sum_{i=1}^{n} P(A_i B_j) \right) = \sum_{j=1}^{m} b_j P(B_j)
\end{aligned}
$$

□

Теорема 7.1. *Пусть X, Y – простые сл.в. Тогда:*

1. $\mathbb{E}(cX) = c\,\mathbb{E}\, X$
2. $\mathbb{E}(X + Y) = \mathbb{E}\, X + \mathbb{E}\, Y$
3. *Если* $X \geq 0$, *то* $\mathbb{E}\, X \geq 0$
4. *Если* $Y \leq X$, *то* $\mathbb{E}\, Y \leq \mathbb{E}\, X$

Доказательство. 1. Очевидно, т.к. cX принимает на A_i значение ca_i.

2. Если X, Y допускают запись в виде линейной комбинации событий, образующих разбиение Ω, то утверждение тривиально. В общем случае переходим к разбиению A_iB_j как в доказательстве 7.1.

3. Очевидно из определения

4. $\mathbb{E}(Y - X) = \mathbb{E}(Y + (-X)) = \mathbb{E}\,Y - \mathbb{E}\,X$. $Y - X \leq 0 \implies \mathbb{E}\,Y \leq \mathbb{E}\,X$.

□

Следствие 7.1. *Если $\alpha, \beta \in \mathbb{R}$, а X, Y – пр.сл.в., то $\mathbb{E}(\alpha X + \beta Y) = \alpha\,\mathbb{E}\,X + \beta\,\mathbb{E}\,Y$.*

7.1.2 Интеграл для знакопостоянных функций

Определение 7.3. Для сл.в. $X \geq 0$ определим $\mathbb{E}\,X$ так:
$\mathbb{E}\,X = \sup\{\mathbb{E}\,Y : 0 \leq Y \leq X, Y \text{ – пр.сл.в.}\}$

Лемма 7.1. *Пусть сл.в. $X \geq 0$. Тогда найдётся последовательность пр.сл.в. X_n таких, что $0 \leq X_n \uparrow X$.*

Доказательство. Для $n \in \mathbb{N}, \omega \in \Omega$ определим

$$X_n(\omega) = \begin{cases} k2^{-n}, & \text{если } k2^{-n} \leq X_n(\omega) < (k+1)2^{-n}, k = 0, \ldots, n2^n - 1 \\ n, & \text{если } X(\omega) \geq n \end{cases}$$

Это же можно записать более лаконично:

$$X_n = \min\left(2^{-n}[2^n X], n\right)$$

Несложно убедиться, что такая последовательность является искомой.

□

Лемма 7.2. *Пусть сл.в. $X \geq 0$. Тогда для любой последовательности простых функций $(X_n)_{n\in\mathbb{N}}$ таких, что $0 \leq X_n \uparrow X$ верно соотношение $\mathbb{E}\,X = \lim\limits_{n\to\infty} \mathbb{E}\,X_n$.*

Доказательство. Обозначим $a = \lim\limits_{n\to\infty} \mathbb{E}\,X_n$. $\mathbb{E}\,X_n \leq \mathbb{E}\,X_{n+1} \leq \mathbb{E}\,X \implies a \leq \mathbb{E}\,X$. Покажем, что $a \geq \mathbb{E}\,X$. Иначе говоря, убедимся, что $\mathbb{E}\,Y \leq a$ для любой пр.сл.в Y такой, что $0 \leq Y \leq X$. Пусть Y принимает значения $0 = b_0, b_1, \ldots, b_k$ с вероятностями $p_0, \ldots, p_k$ соответственно. Если $Y \equiv 0$, то утверждение очевидно.

Введём $B_i = \{Y = b_i\}, i = 0, \ldots, k$. Тогда $\mathbb{E}\,Y = \sum\limits_{i=1}^{k} b_i P(B_i)$. Возьмём $\varepsilon \in (0, 1)$, и определим сл.в. $Y_n = (1-\varepsilon)Y\mathbb{1}((1-\varepsilon)Y \leq X_n)$. Тогда $Y_n \equiv 0$ на событии $B_0 = \{Y = 0\}$, а для $1 \leq i \leq k$ на $C_{i,n} = B_i \cap \{(1-\varepsilon)Y \leq X_n\}$ имеем $Y_n = (1-\varepsilon)Y = (1-\varepsilon)b_i$.

Итак, $Y_n \leq X_n$, значит, $\mathbb{E}\,Y_n \leq \mathbb{E}\,X_n \leq a$. $X_n \uparrow X \implies C_{i,n} \uparrow B_i$. Таким образом, $\mathbb{E}\,Y_n = (1-\varepsilon)\sum\limits_{i=1}^{k} b_i P(C_{i,n})$, а это стремится к $(1-\varepsilon)\sum\limits_{i=1}^{k} b_i P(B_i) = (1-\varepsilon)\,\mathbb{E}\,Y \leq a$. Так как ε был произвольный, можем написать $\mathbb{E}\,Y \leq a$. □

Следствие 7.2. *Пусть даны сл.в.* $X, Y \geq 0$, $c > 0$. *Тогда* $\mathbb{E}(cX) = c\,\mathbb{E}\,X$, *а* $\mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y$.

Доказательство. Возьмём X_n, Y_n из леммы, их сумма будет приближать $X+Y$. □

7.1.3 Интеграл для измеримых функций

Определение 7.4. $X : \Omega \to \mathbb{R}$. Тогда $X = X^+ - X^-$, где $X^+ = X\mathbb{1}(x \geq 0)$, $X^- = -X\mathbb{1}(x < 0)$. Положим $\mathbb{E}\,X = \mathbb{E}\,X^+ - \mathbb{E}\,X^-$. Если $\mathbb{E}\,X \in \mathbb{R}$, то говорят, что X интегрируема, или что $X \in L^1$.

Замечание 7.1. $\mathbb{E}\,X$ не определено, если $\mathbb{E}\,X^+ = \mathbb{E}\,X^- = +\infty$.

Замечание 7.2. Будем считать, что $\forall c \in \mathbb{R}$ $\infty - c = \infty, c - \infty = -\infty$.

Определение 7.5. Если $A \in \mathcal{F}$, то

$$\int\limits_A X dP := E(X\mathbb{1}(A)) = \int\limits_\Omega X\mathbb{1}(A) dP$$

Обозначается $\mathbb{E}(X; A)$

Замечание 7.3. Если $P(X = \infty) \neq 0$, то $\mathbb{E}\,X = \infty$. Действительно, достаточно взять $X_n = n\mathbb{1}(X = \infty) \leq X$.

Замечание 7.4. Если $\mathbb{E}\,|X| < \infty$, то $|X| < \infty$ на Ω' и $P(\Omega') = 1$. В таком случае говорят, что $|X| < \infty$ почти наверно (п.н.).

7.2 Свойства математического ожидания

Теорема 7.2. *Справедливы следующие утверждения:*

1. *Если* $0 \leq Y \leq X$ *и* $X \in L^1$, *то* $Y \in L^1$
2. $X \in L^1 \iff |X| \in L^1$, *причём* $|\mathbb{E}\,X| \leq \mathbb{E}\,|X|$
3. L^1 *– линейное пространство*
4. $\mathbb{E}$ *– линейный функционал на* L^1
5. *Если* $X, Y \in L^1$ *и* $X \leq Y$, *то* $\mathbb{E}\,X \leq \mathbb{E}\,Y$.

Доказательство.

1. Если $0 \leq Z \leq Y$, то $Z \leq X$, и когда возьмём супремум оценка сохранится.
2. Пусть $|X| \in L^1$. Понятно, что $X^+ \leq |X|$ и $X^- \leq |X|$, значит, $X^+, X^- \in L^1$, и $X \in L^1$.

   Обратно, пусть $X \in L^1$, т.е. $X^+, X^- \in L^1$. Тогда $|X| = X^+ + X^- \in L^1$ (по 7.2).

   Если $X \in L^1$, то $|\mathbb{E}\,X| = |\mathbb{E}\,X^+ + \mathbb{E}\,X^-| \leq |\mathbb{E}\,X^+| + |\mathbb{E}\,X^-| = \mathbb{E}\,X^+ + \mathbb{E}\,X^- = \mathbb{E}\,|X|$.
3. Пусть $X, Y \in L^1$,$\alpha, \beta \in \mathbb{R}$. Тогда $|\alpha X + \beta Y| \leq |\alpha||X| + |\beta||Y|$. В силу 7.2 и первых двух пунктов получаем, что $\alpha X + \beta Y \in L^1$, значит, L^1 – линейное пространство.

4. Пусть $X, Y \in L^1$. Покажем, что $\mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y$. Надо проверить, что $\mathbb{E}(X+Y)^+ - \mathbb{E}(X+Y)^- = \mathbb{E}\,X^+ - \mathbb{E}\,X^- + \mathbb{E}\,Y^+ - \mathbb{E}\,Y^-$. Все эти мат. ожидания конечны. Нужное нам равенство получается из равенства просто для чисел: $(x+y)^+ + x^- + y^- = (x+y)^- + x^+ + y^+$.

   То, что $\mathbb{E}(cX) = c\,\mathbb{E}\,X$, доказывается просто (рассмотреть случаи, когда $c \geq 0$ и $c < 0$.

5. $0 \leq Y - X \implies 0 \leq \mathbb{E}(Y - X) = \mathbb{E}\,Y - \mathbb{E}\,X$

$\square$

8 Свойства интеграла Лебега

Теорема 8.1. *(Неравенство Маркова) Пусть неубывающая функция $f : \mathbb{R}_+ \to \mathbb{R}_+$.*

Тогда для любой случайной величины X и $\forall t \geq 0$ верно неравенство: $Ef(|x|) \geq f(t)P(|x| \geq 0)$

Если $f(t) > 0$, то $P(|x| \geq t) \leq \frac{Ef(|x|)}{f(t)}$.

В частности, если $f(t) = t^u, t \geq 0, u \geq 0 : P(|x| \geq t) \leq \frac{E|x|^u}{t^u}$

Для $f(t) = exp(\lambda t), t \geq 0, \lambda \geq 0 : P(|x| \geq t) \leq e^{-\lambda t} E exp \lambda |x|$

Если $Ef(|x|) = \infty$, то неравенство тривиально.

Доказательство. При $t \geq 0$ имеем $x \geq t \implies f(x) \geq f(t)$. Следовательно $f(|x|) \geq f(|x|)I(|x| \geq t)$ так как индикатор принимает только два значения 0 и 1. В свою очередь, $f(|x)|I(|x| \geq t) \geq f(t)I(|x| \geq t)$. Теперь берем математическое ожидание : $Ef(|x|) \geq Ef(t)I(|x| \geq t) =$(выносим константу из-под знака математического ожидания) $= f(t)P(|x| \geq t)$ так как мат. ожидание от индикатора - вероятность того события, которое характеризует индикатор. $\square$

Лемма 8.1. *Если $x \in L^1$(то есть величина имеет конечное математическое ожидание), то :*

1. *$E|x| = 0 \iff x = 0$почти наверное*

2. *Если случайная величина $Y : Y = X$ почти наверное, то $Y \in L^1$ и $EY = EX$*

Доказательство.

1. $X = 0$ почти наверное $\iff |X| = 0$ почти наверное, поэтому далее достаточно рассмотреть только $x \geq 0$.

   Берем последовательность $0 \leq X_n \nearrow x$ (такая последовательность уже была построена: $[2^n]$) Тогда : $C = (w : x(w) = 0) \subset (X_n = 0) = (x \in [0, 2^{-n}])$

   По условию $P(C) = 1$, получается, что $EX_n = 0$ для каждой такой аппроксимирующей последовательности (потому что все значения, кроме нулевого эта величина X_n принимает с вероятностью нуль, а нулевое с вероятностью 1)

   Значит $EX = \lim EX_n = 0$. Если $E|x| = 0$, то $P(|x| > 0)$ событие $(x > 0) = \cup(|x| \geq \frac{1}{n})$ Значит по субаддитивности вероятностей $P(|x| > 0) \leq \sum\limits_{i=1}^{n}(P|x| \geq \frac{1}{n})$ (применяем нер-во Маркова) $\leq \sum\limits_{i=1}^{n} \frac{E(x)}{\frac{1}{n}}$ (суммируем) $= 0$

Отсюда получаем, что $|x| = 0$ почти наверное.

2. Заметим, что если $Y = X \to Y - X = 0$ почти наверное. Теперь $Y = (Y - X) + X$, но $Y - X = 0$ п.н., а $X \in L^1$.

   Если величина равна нулю почти наверное, то она интегрируема и ее математическое ожидание равно нулю. Значит $EY = E(Y-X)+EX = EX$.

□

8.1 Предельный переход под знаком интеграла Лебега.

Теорема 8.2. *(Леви, о монотонной сходимости) Пусть даны случайные величины X_n (необязательно простые такие, что $0 \le X_n \nearrow x$ при $n \to \infty$). Тогда $EX_n \nearrow EX, n \to \infty$*

Доказательство. Очевидно, $EX_n \le EX \forall n$ (так как X_n не убывая сходятся к X, а для мат. ожиданий тот же знак нер-ва).

Неубывающая последовательность всегда имеет предел, значит $\exists \lim_n EX_n \le EX$ $\forall n \in \mathbb{N}$ построим простые неотрицательные случайные величины $Y_{n,k} \nearrow X_n, k \to \infty$.

Определим $Z_k := \max(Y_{n,j})$ (максимум по $1 \le n, j \le k$), получаем неубывающую последовательность простых случайных величин.

Следовательно, $Z_k \nearrow z$. Итак на Ω $\forall k \in \mathbb{N}$, $n = 1, ..., k$ справедливо нер-во:

$0 \le Y_{n,k} \le Z_k \le X_k \le x$.

Берем предел при $k \to \infty$. Тогда :

$0 \le X_n \le z \le x$.

Берем предел при $n \to \infty$. Тогда $z = k$. Тогда $Z_k \nearrow z = x$ и $Z_k \le X_k$. Получаем $EZ_k \le EX_k, \forall k \in \mathbb{N}$, значит $Ex = \lim_n EZ_k \le \lim_n EX_n$ □

Следствие 8.1. *Выполняются следующие два утверждения:*

1. *Пусть сл. в. $X_n \ge 0, \forall n \in \mathbb{N}$. Тогда $E(\sum_{n=1}^{\infty} X_n) = \sum_{n=1}^{\infty} EX_n$. (*)*

2. *Если сл. в. X_n таковы, что $\sum_{n=1}^{\infty} E|X_n| \le \infty$. Тогда ряд $\sum_{n=1}^{\infty} |X_n|$ сходится п.н. и выполнено соотношение (*).*

Доказательство.

Введем $S_n := \sum_{n=1}^{\infty} X_n$. Тогда $S_n \nearrow S = (\sum_{k=1}^{\infty} X_k, n \to \infty$. Поэтому утверждение (1) вытекает из т. о монотонной сходимости.

(Упражнение. Надо рассмотреть $T_n := (\sum_{k=1}^{n} |X_k| \nearrow T := (\sum_{k=1}^{\infty} |X_k|, n \to \infty$, а дальше введем $S_n(+) := (\sum_{k=1}^{n} X_k^+)$ и $S_n(-) = (\sum_{k=1}^{\infty} X_k^-$. Используя эти вспомогательными величинами и пользуясь $0 \le |X_k^+| \le |X_k|$ и $0 \le |X_k^-| \le |X_k|$, можно доказать) □

Лемма 8.2. *Утверждение теоремы о монотонной сходимости сохранится, если рассмотреть последовательность* $X_n : Y \leq X_n \nearrow x, EY > -\infty$. *(Если $Y = 0$, то получаем теорему о монотонной сходимости)*

Лемма 8.3. *(Фату) Пусть X_n и Y сл.в., такие что $X_n \geq Y \forall n \in \mathbb{N}$, причем $EY > -\infty$. Тогда:*

1. $E(\liminf_{n\to\infty} X_n \leq \liminf_{n\to\infty} EX_n)$.
2. *Если* $X_n \leq Y \forall n$, *причем* $EY < \infty$, *то* $limsup EX_n \leq E(limsup X_n)$

Доказательство.

1. Положим $x := \liminf_{n\to\infty} X_n = \lim_{k\to\infty} \inf_{k\geq n} X_k$. Тогда $Y \leq Y_n := \inf_{k\geq n} X_k \nearrow x, k \to \infty$. Пользуясь упражнением, получаем $EX = \lim nEY_n$. Для $k \geq n$ имеем $Y_n \geq X_k$. Отсюда вытекает, что $EY_n \leq EX_k$. Следовательно $EY_n \leq \inf k \geq nEX_k$, что влечет утверждение 1).
2. $-X_n \geq -Y$,и тогда $E(-Y) > -\infty$. (к величинам $-X_n$ применяем пункт 1)

□

Теорема 8.3. *(Лебег, о мажорируемой сходимости) Пусть x, X_n, Y - сл.в. такие, что $X_n \to x$ при $n \to \infty$ (на всем Ω) и $|X_n| \leq Y$, где $Y \in L^1$. Тогда $x \in L^1$ и $EX_n = \lim nEX_n$.*

Доказательство. Имеем $|X_n| \leq Y$, где $Y \leq L^1$. Поэтому $-Y \leq X_n \leq Y$.

Значит $Y, -Y \in L^1$. По лемме Фату $E(\liminf_n X_n) \leq \liminf_n EX_n \leq \limsup_n EX_n \leq E(\limsup_n X_n)$. По условию $x = \lim X_n = limsup X_n$. Отсюда получаем искомое утверждение. □

8.2 Переход к интегрированию по новому пространству

Теорема 8.4. *Пусть $(S, \mathcal{B})$ измеримое пространство. $X : \Omega \to S$, $X \in F|\mathcal{B}$. Если $h : S \to \mathbb{R}$, $h \in \mathcal{B}|\mathcal{B}(\mathbb{R})$, то тогда :*

$Eh(x) = \int\limits_{\Omega} h(xX(w))P(dw) = \int\limits_{S} h(X)P_X(dx)$, *где P_X -распределение случайного элемента X. Точнее говоря, если существует хотя бы один интеграл по Ω или по S, то существует и другой, при этом выполняется равенство.*

Доказательство. Пусть в начале $h(x) = \mathbb{1}_{\mathcal{B}}(x), B \in \mathcal{B}, x \in S$. Тогда $h(X(w)) = 1$ для $w \in X^{-1}(\mathcal{B})$ и h(X(w)) = 0 в остальных случаях.

Поэтому $Eh(x) = P(X^{-1}(B)) = P_X(B) = \int\limits_S \mathbb{1}_{\mathcal{B}}(x)P(dx)$. Пользуясь свойством линейности интеграла видим, что утверждение теоремы справедливо для $\forall$ простой функции $h : S \to \mathbb{R}(h \in \mathcal{B}|\mathcal{B}(\mathbb{R}))$. Пусть $h \geq 0$. Возьмем последовательность простых функций $h_n : S \to \mathbb{R}$ таких, что $h_n \nearrow h$. По теореме о монотонной сходимости Eh(x) $= \lim_{n\to\infty} Eh_n(x) = \lim n \int\limits_S h_n(x)P_X(dx) = \int\limits_S (x)P_X(dx)$ (тоже по т. о монотонной сходимости). Этот результат можем применить к h^+ и h^-. Получаем $Eh^+(x) = \int\limits_S h^+(x)P_X(dx)$, $Eh^-(x) = \int\limits_S h^-(x)P_X(dx)$.

□

8.3 $L^P(\Omega, F, P), p > 0$

Рассмотрим сл. в. X : $E|x|^p < \infty$. Обозначим $\tilde{x}$ класс случайных величин, таких что они равны почти наверное. X и Y входят в один класс, если X = Y пости наверное. Возникает отношение эквивалентности.

Если X $\sim$ Y и $X \in L^1$, то $Y \in L^1$ и $EY = EX$. $L^1(\Omega, F, P)$ состоит из классов эквивалентных сл. величин, которые входят в $L^1 = L^1(\Omega, F, P)$.

Аналогично $L^P(\Omega, F, P)$ состоит из классов эквивалентных функций таких, что $|x|^p \in L^1$

Замечание. Пусть $Y \le X_n \nearrow x$ почти наверное при $n \to \infty$ и $X_n \le X_{n+1}$ почти наверное. Если $EY > -\infty$, тогда $EX_n \nearrow Ex$.

Лемма 8.4. *(КБШ) Пусть H - гильбертово пространство со скалярным произведением* $(\cdot, \cdot)$ *и нормой* $\|x\| := \sqrt{(x, x)}$. *Тогда* $\forall x, y \in H$ *верно нер-во:* $|(x, y)| \le ||x||||y||$.

Доказательство. $\forall t \in \mathbb{R}$,пользуясь симметрией и билинейностью скалярного произведения, имеем (tx + y,tx + y) $= t^2(x, x) + 2t(x, y) + (y, y) \ge 0$ $(x, y)^2 - (x, x)(y, y) \le 0$ □

Пространство $L^2(\Omega, F, P)$ является гильбертовым со скалярным произведением $(X, Y) = \int\limits_{\Omega} XY dp = E(XY)$. Таким образом $|E(XY)| \le \sqrt{Ex^2 Ey^2}$.

8.4 Математатическое ожидание произведения независимых случайных величин

Теорема 8.5. *Пусть X и Y интегрируемые независимые сл. в.. Тогда* $(XY) \in L^1$ *и* $EX \cdot EY$.

Доказательство. Пусть в начале X и Y неотр. сл. в.

Возьмем $X_n \nearrow X$ и $Y_n \nearrow Y$, причем $X = \sum\limits_{k=1}^{N_n} a_{n,k} \mathbb{1}_{X \in B_{n,k}}$ и $Y = \sum\limits_{k=1}^{M_n} b_{n,k} \mathbb{1}_{Y \in C_{n,m}}$. $B_{n,k}, C_{n,k}$ попарно непересекающиеся борелевские подмножества.

Тогда $EX_n \cdot Y_n = \sum\limits_{k=1}^{N_n} \sum\limits_{m=1}^{M_n} a_{n,k} b_{n,m} E(\mathbb{1}_{X \in B_{n,k}} \cdot \mathbb{1}_{Y \in C_{n,m}}) = \sum\limits_{k=1}^{N_n} \sum\limits_{m=1}^{M_n} a_{n,k} b_{n,m} P(X \in B_{n,k}, Y \in C_{n,m})$ = (в силу независимости) $= \sum\limits_{k=1}^{N_n} \sum\limits_{m=1}^{M_n} a_{n,k} b_{n,m} P(X \in B_{n,k}) P(Y \in C_{n,m}) = \sum\limits_{k=1}^{N_n} a_{n,k} P(X \in B_{n,k}) \sum\limits_{m=1}^{M_n} b_{n,m} P(Y \in C_{n,m}) = EX_n \cdot EY_n$

Очевидно $0 \le X_n Y_n \nearrow XY$. Поэтому по теореме о монотонной сходимости $EX_n \cdot EY_n \nearrow EX \cdot EY$, а кроме того $E(X_n Y_n) \nearrow EX \cdot EY$. Теперь для неотрицательных величин все доказано.

Если $X = X^+ - X^-$, $Y = Y^+ - Y^-$, тогда $E(XY) = (X^+ - X^-)(Y^+ - Y^-) = E(X^+Y^+) - E(X^+Y^-) - E(X^-Y^+ + E(X^-Y^-)$ = [Если $X, Y \in L^1 \to X^+, X^-, Y^+, Y^- \in L^1$.] $= (EX^+ - EX^-)(EY^+ - EY^-)$ (достаточно заметить, что $E(X^+Y^+) = E(X^+)E(Y^+)$ и пользуемся, что из независимости X и Y следует независимость X^+Y^+ и т.д. (Борелевские функции от независимых случайных величин - независимы)) □

8.5 Моменты, Дисперсия, Ковариация

Определение 8.1. Моменты вводятся формулой:

EX_n - обычный момент, $E|X_n|$ - абсолютный момент, $E(X - EX)^n$ - центральный момент, $E|X - EX|^n$ - центральный абсолютный момент.

Определение 8.2. Дисперсия $\operatorname{var} X := E(X - EX)^2$.

Если $X, Y \in L^1 \to E|XY| < \infty$, значит $varX < \infty$, если $X \in L^1$.

Определение 8.3. Ковариация задается формулой $\operatorname{cov}(X, Y) := E(X - EX)(Y - EY) = E(XY) - EXEY$. Значит ковариация конечна, если $X, Y \in L^1$.

Если X и Y независимы и EX и EY конечны, то ковариация равна нулю.

9 Сходимость случайных величин

9.1 Свойства дисперсии и ковариации

Вначале напомним, что по определению *дисперсией* случайной величины X называется следующее математическое ожидание:

$$\operatorname{var} X := \mathbb{E}(X - \mathbb{E}\, X)^2 \tag{9.1}$$

(по-английски пишут *variance*, не путать с вариацией, т. е. с *variation*!). Тем самым подразумевается, что $\mathbb{E}\, X$ конечно в формуле 9.1. Говорят, что дисперсия существует, если $\mathbb{E}(X - \mathbb{E}\, X)^2 < \infty$, т. е. $X - \mathbb{E}\, X \in L^2$. Случайная величина, равная константе, входит в пространство L^2. Поскольку L^2 – линейное пространство (достаточно учесть, что $|X + Y| \leqslant 2X^2 + 2Y^2$ и $|cX|^2 = c^2X^2$), то видим, что дисперсия конечна тогда и только тогда, когда $X \in L^2$. Поэтому для $X \in L^2$ выполнено

$$\operatorname{var} X = \mathbb{E}(X^2 - 2X\,\mathbb{E}\, X + (\mathbb{E}\, X)^2) = \mathbb{E}\, X^2 - 2\,\mathbb{E}\, X\,\mathbb{E}\, X + (\mathbb{E}\, X)^2 = \mathbb{E}\, X^2 - (\mathbb{E}\, X^2).$$

Учли, что интегрируемы X^2, $(-2X\,\mathbb{E}\, X)$ и $(\mathbb{E}\, X)^2$.

Пример 9.1. Пусть $X = \mathbb{1}(A)$, где $A \in \mathcal{F}$. Принимая во внимание, что $\mathbb{1}_A^2 = \mathbb{1}_A$, получаем $\mathbb{E}(\mathbb{1}(A))^2 = \mathbb{E}\,\mathbb{1}(A) = P(A)$. Следовательно,

$$\operatorname{var} \mathbb{1}(A) = P(A) - (P(A))^2 = P(A)(1 - P(A)).$$

Итак, если случайная величина X принимает значения 1 и 0 соответственно с вероятностями p и $1 - p$, то $\operatorname{var} X = p(1 - p)$.

Для случайной величины X, принимающей значения $x_1, \ldots, x_n$ соответственно с вероятностями $p_1, \ldots, p_n$ имеем

$$\operatorname{var} X = \sum_{k=1}^{n} (x_k - \overline{x})^2 p_k, \tag{9.2}$$

где $\overline{x} := \mathbb{E}\, X = \sum\limits_{i=1}^{n} p_i x_i$. Если на прямой дана система материальных точек $x_1, \ldots, x_n$ таких, что в точке x_k сосредоточена масса p_k ($k = 1, \ldots, n$), то, как известно из механики, формула 9.2 представляет собой *момент* *инерции* этой системы точек. Таким образом, можно сказать, что *дисперсия является мерой разброса масс относительно центра тяжести системы.*

Из определения 9.1 немедленно следует, что $\operatorname{var} X \geqslant 0$. Кроме того, для любой величины $X \in L^2$ и константы c верны равенства

$$\operatorname{var}(X + c) = \operatorname{var} X, \quad \operatorname{var}(cX) = c^2 \operatorname{var} X.$$

Определение 9.1. *Ковариацией* случайных величин X и Y называется следующее математическое ожидание

$$\operatorname{cov}(X, Y) := \mathbb{E}(X - \mathbb{E}\, X)(Y - \mathbb{E}\, Y), \tag{9.3}$$

если оно конечно. Таким образом, заведомо предполагается, что в формуле 9.3 конечны $\mathbb{E}\, X$ и $\mathbb{E}\, Y$.

Неравенство Коши – Буняковского – Шварца показывает, что если $X, Y \in L^2$, то $\operatorname{cov}(X, Y)$ существует и конечна. В то же время, если величины X, Y интегрируемы (входят в L^1, но необязательно в L^2) и независимы, то независимыми и интегрируемыми будут $X - \mathbb{E}\, X$ и $Y - \mathbb{E}\, Y$. Поэтому согласно теореме о математическом ожидании произведения независимых интегрируемых величин **??** имеем

$$\operatorname{cov}(X, Y) = \mathbb{E}(X - \mathbb{E}\, X)(Y - \mathbb{E}\, Y) = \mathbb{E}(X - \mathbb{E}\, X)\, \mathbb{E}(Y - \mathbb{E}\, Y) = 0.$$

Постройте пример зависимых случайных величин X и Y, для которых $\operatorname{cov}(X, Y) = 0$.

Если $X, Y \in L^2$, то легко видеть, что

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}\, X\, \mathbb{E}\, Y.$$

Из определения 9.1 немедленно вытекает, что $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$ и $\operatorname{cov}(X, X) = \operatorname{var} X$. Если существует $\operatorname{cov}(X, Y)$, то для любой константы c имеем $\operatorname{cov}(X + c, Y) = \operatorname{cov}(X, Y)$ и $\operatorname{cov}(cX, Y) = c \operatorname{cov}(X, Y)$. Легко видеть, что если существуют $\operatorname{cov}(X, Y)$ и $\operatorname{cov}(X, Z)$, то существует $\operatorname{cov}(X, Y + Z)$ и при этом $\operatorname{cov}(X, Y + Z) = \operatorname{cov}(X, Y) + \operatorname{cov}(X, Z)$. Иначе говоря, если для случайных величин X, Y, Z определены $\operatorname{cov}(X, Y)$ и $\operatorname{cov}(X, Z)$, то для любых констант $\alpha, \beta \in \mathbb{R}$ справедливо равенство

$$\operatorname{cov}(X, \alpha Y + \beta Z) = \alpha \operatorname{cov}(X, Y) + \beta \operatorname{cov}(X, Z).$$

Лемма 9.1. *Пусть* $X_1, \ldots, X_n \in L^2$. *Тогда*

$$\operatorname{var}\left(\sum_{k=1}^n X_k\right) = \sum_{i,\, j=1}^n \operatorname{cov}(X_i, X_j) = \sum_{i=1}^n \operatorname{var} X_i + 2 \sum_{1 \leqslant i < j \leqslant n} \operatorname{cov}(X_i, X_j). \tag{9.4}$$

Доказательство. Пользуясь свойством билинейности ковариации, имеем

$$\operatorname{var}\left(\sum_{k=1}^n X_k\right) = \operatorname{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i,\, j=1}^n \operatorname{cov}(X_i, X_j).$$

Осталось в последней двойной сумме выделить слагаемые, для которых $i = j$, и учесть, что $\operatorname{cov}(X_i, X_i) = \operatorname{var} X_i$. □

Следствие 9.1. *Пусть* $X_1, \ldots, X_n \in L^2$ *и независимы. Тогда*

$$\operatorname{var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{var} X_k. \tag{9.5}$$

Пример 9.2. Пусть $X_1, \ldots, X_n$ – независимые бернулевские величины (т. е. принимающие значения 1 и 0 соответственно с вероятностями p и $1-p$). Тогда, пользуясь следствием 9.1 и примером 9.1, для $S_n := X_1 + \ldots + X_n$ получаем

$$\operatorname{var} S_n = \sum_{k=1}^{n} \operatorname{var} X_k = np(1-p), \quad n \in \mathbb{N}. \tag{9.6}$$

Заметим, что развитая теория позволила фактически устно найти дисперсию величины S_n. Если бы мы считали $\operatorname{var} S_n$ непосредственно, то (вспоминая, что $\mathbb{E}\, S_n = \sum\limits_{k=1}^{n} \mathbb{E}\, X_k = np$) должны были бы преобразовать к виду 9.6 следующую сумму:

$$\operatorname{var} S_n = \sum_{k=0}^{n} (k-np)^2 C_n^k p^k (1-p)^{n-k}.$$

9.2 Интеграл Лебега по конечной и сигма-конечной мере

Для конечной меры μ на $(S, \mathcal{B})$ построение интеграла Лебега в точности повторяет три этапа построения интеграла Лебега по вероятностной мере P. При этом для (измеримой функции) $f\colon S \to \mathbb{R}$

$$\int_S f\, d\mu = \mu(S) \int_S f\, dP, \tag{9.7}$$

где вероятностная мера $P := \mu/\mu(S)$. Тривиальный случай, когда μ тождественно равна нулю, исключается из рассмотрения. Точнее говоря, оба интеграла в 9.7 существуют или не существуют одновременно, а если существуют, то равны.

Заметим, что если μ – конечная мера на $(S, \mathcal{B})$ и $B \in \mathcal{B}$, то для любой измеримой функции $f\colon S \to \mathbb{R}$

$$\int_B f\, d\mu := \int_S f \mathbb{1}_B\, d\mu = \int_B f|_B\, d\mu|_B, \tag{9.8}$$

где $f|_B$ – сужение функции f на множество B, а $\mu|_B$ – сужение меры μ на σ-алгебру $\mathcal{B}|_B := \mathcal{B} \cap B$. Другими словами, f рассматривается теперь только на B, и $f(x) = f|_B(x)$ для $x \in B$, а $\mu|_B(A) = \mu(A \cap B)$ для $A \in \mathcal{B}$. Равенство 9.8 очевидно, если f – простая функция. Затем берём неотрицательную функцию f и строим последовательность простых функций $0 \leqslant f_n \nearrow f$ при $n \to \infty$. Тогда для сужения f_n на B получаем, что (для простых на B функций) $0 \leqslant (f_n)|_B \nearrow f|_B$, $n \to \infty$. И в этом случае 9.8 имеет место. Далее стандартным образом рассматриваются знакопеременные f.

Определение 9.2. Мера μ, заданная на $(S, \mathcal{B})$ называется *σ-конечной*, если существует разбиение S последовательностью множеств $S_1, S_2, \ldots \in \mathcal{B}$ таких, что $\mu(S_n) < \infty$ при каждом $n \in \mathbb{N}$.

Чтобы при этом исключить из рассмотрения конечную меру μ, обычно предполагают, что $\mu(S) = \infty$. Если μ является σ-конечной мерой на $(S, \mathcal{B})$, а последовательность множеств $(S_n)_{n\in\mathbb{N}}$ задаёт разбиение S $(S_n \in \mathcal{B},\ n \in \mathbb{N})$, то для любого $B \in \mathcal{B}$

$$\mu(B) = \sum_{n=1}^{\infty} \mu(B \cap S_n) = \sum_{n=1}^{\infty} \mu_n(B_n), \tag{9.9}$$

где μ_n – сужение меры μ на $(S_n, \mathcal{B}_n)$, а σ-алгебра $\mathcal{B}_n := \mathcal{B} \cap S_n$. Формула 9.9 подсказывает, как можно задать σ-конечную меру, например, на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Мы видели, что по функции $F = F(x),\ x \in \mathbb{R}$, обладающей определёнными четырьмя свойствами, можно ввести меру Q на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, для которой F будет функцией распредления. Для $n \in \mathbb{Z}$ введём (непрерывную) функцию

$$F_n(x) := \begin{cases} 0, & x \in (-\infty, n], \\ x - n, & x \in (n, n+1], \\ 1, & x \in (n+1, \infty). \end{cases}$$

Очевидно, функция F_n отвечает вероятностная мера μ_n на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ такая, что $\mu_n(S_n) = 1$, где $S_n := (n, n+1],\ n \in \mathbb{N}$. Назовём *мерой Лебега* на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ меру $\mu := \sum\limits_{n=1}^{\infty} \mu_n$, т. е.

$$\mu(B) := \sum_{n=1}^{\infty} \mu_n(B), \quad B \in \mathcal{B}.$$

Мера μ является σ-конечной. Заметим, что $\mu((a, b]) = b - a$ для $-\infty < a \leqslant b < \infty$.

Определение 9.3. Пусть μ есть σ-конечная мера на $(S, \mathcal{B})$. Для измеримой функции $f\colon S \to \mathbb{R}_+$ *интеграл по мере* μ вводится формулой

$$\int_S f\, d\mu := \sum_{n=1}^{\infty} \int_{S_n} f_n\, d\mu, \tag{9.10}$$

где f_n – сужение f на S_n, а μ_n – сужение μ на $(S_n, \mathcal{B}_n)$, $\mathcal{B}_n := \mathcal{B} \cap S_n$. Если измеримая функция $f\colon S \to \mathbb{R}$, то, как обычно, полагаем

$$\int_S f\, d\mu := \int_S f^+\, d\mu - \int_S f^-\, d\mu,$$

причём в случае неопределённости $\infty - \infty$ говорим, что интеграл не существует.

В качестве несложного упражнения предлагается доказать, что правая часть формулы 9.10 не изменится при выборе иного разбиения множества S.

Как и для вероятностной меры, при $p > 0$ вводятся пространства $L^p(S, \mathcal{B}, \mu)$, состоящие из функций f (точнее говоря, классов эквивалентных функций), для которых

$$\int_S |f|^p\, d\mu < \infty.$$

Взятие интеграла обладает свойством линейности на пространстве $L^1(S, \mathcal{B}, \mu)$ не только для конечной, но и для σ-конечной меры μ.

9.3 Замена меры в интеграле Лебега

Определение 9.4. Пусть ν и μ – две σ-конечные меры на $(S, \mathcal{B})$. Говорят, что ν *абсолютно непрерывна* относительно μ (пишут $\nu \ll \mu$), если $\mu(B) = 0$ для $B \in \mathcal{B}$ влечёт, что $\nu(B) = 0$.

Сформулируем без доказательства следующий очень важный результат.

Теорема 9.1. *(Радон-Никодим) Пусть μ и ν – две σ-конечные меры на $(S, \mathcal{B})$. Соотношение $\nu \ll \mu$ равносильно тому, что найдётся измеримая функция $p \colon S \to \mathbb{R}_+$, называемая плотностью меры ν по мере μ (или производной Радона – Никодима меры ν по мере μ), такая, что*

$$\nu(B) = \int_B p(x)\mu(dx) \quad \text{для каждого } B \in \mathcal{B}. \tag{9.11}$$

При этом если для любого $B \in \mathcal{B}$ и некотрой измеримой функции $q \colon S \to \mathbb{R}_+$ справедливо равенство

$$\nu(B) = \int_B q(x)\mu(dx),$$

то $q(x) = p(x)$ для μ-почти всех $x \in S$ (иначе говоря, $\mu\{x \in S : q(x) \neq p(x)\} = 0$).

Очевидно, интеграл в 9.11 не изменится, если вместо p взять любую функцию f такую, что μ-почти всюду $f = p$. Таким образом, когда мы говорим о плотности меры ν по мере μ, то имеем в виду целый класс эквивалентных функций (и можем оперировать с любым представителем этого класса). Производную Радона – Никодима меры ν по мере μ обозначают $d\nu/d\mu$.

Теорема 9.2. *Пусть μ и ν – две σ-конечные меры на $(S, \mathcal{B})$, причём $\nu \ll \mu$ и $p = d\nu/d\mu$. Если $h \colon S \to \mathbb{R}$, то*

$$\int_S h(x)\nu(dx) = \int_S h(x)p(x)\mu(dx). \tag{9.12}$$

Точнее говоря, оба интеграла существуют или не существуют одновременно, а если существуют, то равны.

Доказательство. Пусть $h = \mathbb{1}_B$ для некоторого $B \in \mathcal{B}$. Тогда

$$\int_S h(x)\nu(dx) = \nu(B), \quad \int_S h(x)p(x)\mu(dx) = \int_B p(x)\mu(dx) = \nu(B).$$

Линейность интеграла обеспечивает равенство 9.12 для простых функций. Для неотрицательных функций h требуемый результат получается по теореме о монотонной сходимости, которая легко переносится на случай интеграла по σ-конечной мере. Для знакопеременной h достаточно отдельно рассмотреть интегралы от h^+ и h^-. □

Замечание 9.1. Нетрудно показать, что для непрерывных и кусочно- непрерывных функций, заданных на отрезке, интеграл Римана совпадает с интегралом Лебега (по мере Лебега). В то же время (несобственный) интеграл

Римана по всей прямой от функции f может быть конечным, а интеграл от $|f|$ бесконечным. Достаточно рассмотреть непрерывную функцию

$$f(x) = \begin{cases} \dfrac{\sin x}{x}, & x \in \mathbb{R} \setminus \{0\}, \\ 1, & x = 0. \end{cases}$$

В этом проявляется существенное отличие интеграла Римана от интеграла Лебега (функция f имеет конечный интеграл Лебега по σ-конечной мере тогда и только тогда, когда конечен интеграл функции $|f|$ по этой мере).

Пример 9.3. Пусть $X \sim N(a, \sigma^2)$, где $a \in \mathbb{R}$, $\sigma > 0$. Другими словами, у случайно величины имеется плотность

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

Точнее говоря, $p(x) = \dfrac{dP_X}{d\mu}(x)$, где P_X – распределение X, μ – мера Лебега на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Найдём $\mathbb{E}\, X$ и $\operatorname{var} X$.

Применяя теорему прошлой лекции о переходе от интеграла по мере P к интегралу по распределению случайного элемента, а также пользуясь формулой 9.12, имеем (далее пишем dx вместо $\mu(dx)$, где μ – мера Лебега)

$$\mathbb{E}\, X = \int_{\mathbb{R}} xp(x)\, dx = \int_{\mathbb{R}} (x-a)p(x) + a \int_{\mathbb{R}} p(x)\, dx = \\ = \sigma \int_{\mathbb{R}} u \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du + a = a,$$

где учтено, что интеграл от плотности вероятностной меры, взятый по $\mathbb{R}$, равен единице, а интеграл по всей прямой (который конечен) от нечётной функции равен нулю. Легко видеть, что функция $|x|p(x)$, $x \in \mathbb{R}$, интегрируема и по Риману, и по Лебегу, причём значения этих интегралов совпадают. Применяя интегрирование по частям для интеграла Римана, получаем

$$\operatorname{var} X = \int_{\mathbb{R}} (x-a)^2 p(x)\, dx = \sigma^2 \int_{\mathbb{R}} u^2 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du = \sigma^2.$$

Таким образом, в записи $X \sim N(a, \sigma^2)$ указывается математическое ожидание и дисперсия величины X.

9.4 Сходимость случайных величин

Ниже даются определения четырёх основных видов сходимости случайных величин X_n $(n \in \mathbb{N})$ к случайной величине X при $n \to \infty$.

Определение 9.5. Говорят, что $X_n \to X$ *почти наверное* (*п. н.*) или *с вероятностью единица*, если

$$P(\omega : X_n(\omega) \not\to X(\omega) \text{ при } n \to \infty) = 0$$

Определение 9.6. Запись $X_n \xrightarrow{P} X$ при $n \to \infty$ обозначает *сходимость по вероятности.* Она имеет место, если для любого $\varepsilon > 0$ выполнено

$$P(\omega : |X_n(\omega) - X(\omega)| \geqslant \varepsilon) \to 0, \quad n \to \infty.$$

Определение 9.7. Если $X,\ X_n \in L^p,\ n \in \mathbb{N}$, то пишут $X_n \xrightarrow{L^p} X$, когда $\mathbb{E}\,|X_n - X|^p \to 0$ при $n \to \infty$.

Определение 9.8. Пусть на некотором пространстве $(S, \mathcal{B})$ заданы вероятностные меры Q и Q_n, $n \in \mathbb{N}$. *Слабая сходимость* Q_n к Q (пишут $Q \Rightarrow Q$) означает, что для каждой непрерывной и ограниченной функции $h\colon S \to \mathbb{R}$

$$\int_S h(x) Q_n(dx) \to \int_S h(x) Q(dx), \quad n \to \infty. \tag{9.13}$$

Пусть случайные элементы $X\colon \Omega \to S$ и $X_n\colon \Omega \to S$ (важно, чтобы все элементы действовали в одно и то же пространство S, снабжённое σ-алгеброй $\mathcal{B}$). Говорят, что X_n *сходятся по распределению* к X при $n \to \infty$, если

$$P_{X_n} \Rightarrow P_X, \quad n \to \infty, \tag{9.14}$$

где P и P_{X_n} – соответственно распределения X и X_n. Наряду с 9.14 пишут $X_n \xrightarrow{law} X$ или $X_n \xrightarrow{\mathcal{D}} X$ (от слова *distribution*), $n \to \infty$.

Обратите внимание на то, что, в отличие от теорем о предельном переходе под знаком интеграла Лебега, в формуле 9.13 меняются не подынтегральные функции, а меры. Сопоставляя 9.13 и 9.14 и вспоминая формулу перехода от интеграла по мере P к интегралу по распределению случайного элемента, видим, что $X_n \xrightarrow{\mathcal{D}} X$ тогда и только тогда, когда для любой непрерывной и ограниченной функции $h\colon S \to \mathbb{R}$

$$\mathbb{E}\, h(X_n) \to \mathbb{E}\, h(X), \quad n \to \infty. \tag{9.15}$$

Теорема 9.3. *Между видами сходимости случайных величин существуют следующие взаимосвязи.*

1. *Сходимость п. н. влечёт сходимость по вероятности.*
2. *Сходимость в L^p $(p > 0)$ влечёт сходимость по вероятности.*
3. *Сходимость по вероятности влечёт сходимость по распределению.*

Предлагается построить примеры, показывающие, что других нетривиальных соотношений, вообще говоря, нет (тривиальными считается последовательное использование двух импликаций, например, сходимость п. н. влечёт сходимость по вероятности, а сходимость по вероятности обеспечивает сходимость по распределению).

Нам понадобится вспомогательный результат.

Лемма 9.2. *Справедливо следующее утверждение: $X_n \to X$ п. н. тогда и только тогда, когда для каждого $\varepsilon > 0$*

$$P\left(\sup_{k \geqslant n} |X_k - X| \geqslant \varepsilon\right) \to 0, \quad n \to \infty. \tag{9.16}$$

Доказательство. Для $\varepsilon > 0$ и $n \in \mathbb{N}$ введём событие $A_n^\varepsilon := \{|X_n - X| \geqslant \varepsilon\}$. Положим

$$A^\varepsilon := \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k^\varepsilon.$$

Тогда, очевидно,

$$\{\omega : X_n(\omega) \not\to X(\omega) \text{ при } n \to \infty\} = \bigcup_{\varepsilon>0} A^\varepsilon.$$

Действительно, если найдётся $\varepsilon > 0$ такое, что для любого n существует $k \geqslant n$ такое, что $|X_n(\omega) - X(\omega)| \geqslant \varepsilon$, то все такие точки ω и будут составлять множество, на котором $X_n(\omega)$ не сходится к $X(\omega)$. Несчётное объединение событий, вообще говоря, не обязано быть событием. Однако очевидно,

$$\bigcup_{\varepsilon>0} A^\varepsilon = \bigcup_{m=1}^{\infty} A^{1/m}.$$

(учитываем, что $A^\varepsilon \subset A^\delta$ при $\varepsilon > \delta$). Итак, $P\left(\bigcup\limits_{m=1}^{\infty} A^{1/m}\right) = 0$ тогда и только тогда, когда $P\left(A^{1/m}\right) = 0$ для каждого $m \in \mathbb{N}$. Последнее утверждение равносильно тому, что $P(A^\varepsilon) = 0$ для любого $\varepsilon > 0$. По свойству непрерывности вероятностной меры

$$P(A^\varepsilon) = \lim_{n\to\infty} P\left(\bigcup_{k\geqslant n} |X_k - X| \geqslant \varepsilon\right).$$

Остаётся заметить, что

$$\left\{\bigcup_{k\geqslant n} |X_k - X| \geqslant \varepsilon\right\} = \left\{\sup_{k\geqslant n} |X_k - X| \geqslant \varepsilon\right\}.$$

□

Доказательство теоремы 2.3. В силу леммы 9.2 утверждение 1 теоремы вытекает из того, что для каждого $\varepsilon > 0$ и любого $n \in \mathbb{N}$

$$\{|X_n - X| \geqslant \varepsilon\} \subset \left\{\sup_{k\geqslant n} |X_k - X| \geqslant \varepsilon\right\}.$$

Утверждение 2 следует из неравенства Маркова: для любого $\varepsilon > 0$ и $p > 0$ имеем

$$P(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}\,|X_n - X|^p}{\varepsilon^p}.$$

Установим утверждение 3, т. е. проверим выполнение 9.15. Функция $h(X) \in L^1$, так как по условию $|h(x)| \leqslant C = \text{const}$. Если $A \in \mathcal{F}$, то измеримая функция $h(X)\mathbb{1}_A \in L^1$, поскольку $|h(X)\mathbb{1}_A| \leqslant |h(X)|$ (также учли, что произведение измеримых функций измеримо). Это же относится и к функции $h(X_n)$. Поэтому для любого $\varepsilon > 0$ имеем

$$\begin{aligned}\mathbb{E}\,h(X) = \mathbb{E}\,h(X)\mathbb{1}(|X - X_n| < \varepsilon) + \mathbb{E}\,h(X)\mathbb{1}(|X - X_n| \geqslant \varepsilon),\\ \mathbb{E}\,h(X_n) = \mathbb{E}\,h(X_n)\mathbb{1}(|X - X_n| < \varepsilon) + \mathbb{E}\,h(X_n)\mathbb{1}(|X - X_n| \geqslant \varepsilon).\end{aligned}$$

Очевидно,

$$\begin{aligned}|\mathbb{E}\,h(X)\mathbb{1}(|X - X_n| \geqslant \varepsilon)| \geqslant CP(|X - X_n| \geqslant \varepsilon),\\ |\mathbb{E}\,h(X_n)\mathbb{1}(|X - X_n| \geqslant \varepsilon)| \geqslant CP(|X - X_n| \geqslant \varepsilon). \quad (9.17)\end{aligned}$$

В силу свойств функции распределения для любого $\gamma > 0$ найдётся $a = a(\gamma)$ такое, что $P(|X| > a) = P(X < -a) + P(X > a) \leqslant \gamma$. Поэтому аналогично 9.17 получаем

$$|\mathbb{E}\, h(X)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| > a)| \geqslant CP(|X| > a),$$
$$|\mathbb{E}\, h(X_n)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| > a)| \geqslant CP(|X| > a).$$

Осталось оценить близость $\mathbb{E}\, h(X)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| \geqslant a)$ к $\mathbb{E}\, h(X)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| \geqslant a)$. Возьмём $\varepsilon \in (0,\, 1)$. Тогда одновременное выполнение событий $\{|X - X_n| < \varepsilon\}$ и $\{|X| \geqslant a\}$ влечёт, что значения X и X_n принадлежат отрезку $[-a - 1,\, a + 1]$. Функция h непрерывна на $\mathbb{R}$. Следовательно, она равномерно непрерывна на $[-a - 1,\, a + 1]$. Поэтому для каждого $\gamma > 0$ мы можем найти такое $\varepsilon \in (0,\, 1)$, $\varepsilon = \varepsilon(\gamma)$, что если $|x - u| < \varepsilon$ и $x,\ u \in [-a - 1,\, a + 1]$, то $|h(x) - h(u)| < \gamma$. Таким образом,

$$\begin{aligned}
&|\mathbb{E}\, h(X)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| \leqslant a) - \mathbb{E}\, h(X_n)\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| \leqslant a)| \leqslant \\
&\quad \leqslant \mathbb{E}\, |h(X) - h(X_n)|\mathbb{1}(|X - X_n| < \varepsilon)\mathbb{1}(|X| \leqslant a) \leqslant \\
&\quad\quad \leqslant \gamma P(|X - X_n| < \varepsilon,\ |X| \leqslant a) \leqslant \gamma.
\end{aligned}$$

Итак, для любого $\gamma > 0$ существует $\varepsilon = \varepsilon(\gamma) \in (0,\, 1)$ такое, что выполнено неравенство

$$|\,\mathbb{E}\, h(X) - \mathbb{E}\, h(X_n)| \leqslant 2CP(|X - X_n| \geqslant \varepsilon) + 2C\gamma + \gamma.$$

Осталось найти $N_0 = N_0(\gamma)$ такое, что $P(|X - X_n| \geqslant \varepsilon) < \gamma$ при $n \geqslant N_0$. Тогда

$$|\,\mathbb{E}\, h(X) - \mathbb{E}\, h(X_n)| \leqslant \gamma(4C + 1)$$

при $n \geqslant N_0$. Поскольку γ – произвольное положительное число, приходим к 9.15. □

В качестве сложного упражнения предлагается самостоятельно доказать следующий результат.

Теорема 9.4. *Соотношение $X_n \xrightarrow{\mathcal{D}} X$ справедливо в том и только в том случае, когда $F_n(x) \to F(x)$ при $n \to \infty$ во всех точках x, являющихся точками непрерывности предельной функции F (F и F_n – соответсвенно функции распределения случайных величин X и X_n, $n \in \mathbb{N}$).*

10 Закон больших чисел

10.1 Простейший вариант закона больших чисел

Лемма 10.1. *(Неравенство Бьенеме - Чебышева) Если* $\operatorname{var} X < +\infty$, *то*

$$\forall \varepsilon > 0 \quad P(|X - \mathbb{E}\, X| \geq \varepsilon) \leq \frac{\operatorname{var} X}{\varepsilon^2}$$

Доказательство. Это неравенство является тривиальным следствием из неравенства Маркова (теорема 8.1) □

Теорема 10.1. *(Бернулли)*[1] *Пусть* $X_1, X_2, \ldots$ *— последовательность независимых случайных величин таких, что* $\exists\ C > 0 \colon \forall n\ \operatorname{var} X_n \leq C$. *Тогда*

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}\,X_k) \xrightarrow{P} 0$$

Если же $X_1, X_2, \ldots$ *— независимые одинаково распределенные случайные величины,* $\operatorname{var} X_1 < +\infty$, *то*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mathbb{E}\,X_1$$

В частности, если $\forall k\ X_k \sim Be(p)$, *то*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} p$$

Доказательство. Пусть $S_n := \sum_{k=1}^{n} X_k$. Из неравенства Чебышева (лемма 10.1) получаем

$$P\left(\left|\frac{S_n}{n} - \frac{\mathbb{E}\,S_n}{n}\right| \geq \varepsilon\right) = P\left(\left|\frac{S_n}{n} - \mathbb{E}\,\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\operatorname{var}\frac{S_n}{n}}{\varepsilon^2} = \frac{1}{\varepsilon^2 n^2}\operatorname{var} S_n =$$
$$= \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^{n}\operatorname{var} X_k \leq \frac{1}{\varepsilon^2 n^2} Cn = \frac{C}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0$$

□

Замечание 10.1. Из хода доказательства теоремы 10.1 видно, что достаточно требовать попарной независимости (X_k). Более того, можно требовать, чтобы $\forall\ i \neq j \quad \operatorname{cov}(X_i, X_j) = 0$

10.2 Вероятностное доказательство теоремы Вейерштрасса о равномерном приближении непрерывных на отрезке функций

Теорема 10.2. *(Вейерштрасс) Пусть* f *непрерывна на* $[a, b]$. *Тогда* $\forall \varepsilon > 0$

$$\exists P_n(x) \in \mathbb{R}[x], \deg P_n = n : \quad \sup_{[a,b]} |f - P_n| < \varepsilon$$

Доказательство. Понятно, что достаточно доказать теорему только для непрерывных на $[0, 1]$ функций. Введем многочлены Бернштейна следующим образом:

$$B_n(f; p) := \sum_{k=0}^{n} f(\tfrac{k}{n}) C_n^k p^k (1-p)^{n-k}, \quad B_n(f; 0) := f(0), B_n(f; 1) = f(1)$$

Покажем, что $B_n(f; p)$ сходится равномерно к f на $[0, 1]$ при $n \to +\infty$:

[1] На семинарах у некоторых групп это называется ЗБЧ в форме Чебышева, а случай н.о.р. бернуллиевских сл.в. – теоремой Бернулли

Для этого рассмотрим последовательность независимых одинаково распределённых случайных величин $X_1, X_2, \ldots\colon \forall k \in \mathbb{N}\ X_k \sim Be(p)$, $S_n := \sum_{k=1}^{n} X_k$ Знаем, что $P(S_n = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$. Рассмотрим $p \in (0,1)$:

$$|f(p) - B(f;p)| = \left|\sum_{k=0}^{n} \left(f(p) - f\left(\frac{k}{n}\right)\right) C_n^k p^k (1-p)^{n-k}\right|$$

$$f \in C[0,1] \implies f \in UC[0,1] \implies \exists\ M > 0\colon |f| \leq M \text{ на } [0,1]$$

Фиксируем $p \in [0,1]$:

$$|f(p) - B(f;p)| \leq \left|\sum_{k\colon |p-\frac{k}{n}|<\delta} \left(f(p) - f\left(\frac{k}{n}\right)\right) C_n^k p^k (1-p)^{n-k}\right| +$$

$$+\left|\sum_{k\colon |p-\frac{k}{n}|\geq\delta} \left(f(p) - f\left(\frac{k}{n}\right)\right) C_n^k p^k (1-p)^{n-k}\right| \leq \sum_{k\colon |p-\frac{k}{n}|<\delta} \frac{\varepsilon}{2} C_n^k p^k (1-p)^{n-k} +$$

$$+ 2M \sum_{k\colon |p-\frac{k}{n}|\geq\delta} C_n^k p^k (1-p)^{n-k} \leq \frac{\varepsilon}{2} + 2M \sum_{k\colon |p-\frac{k}{n}|\geq\delta} P(S_n = k) =$$

$$= \frac{\varepsilon}{2} + P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right)$$

Из неравенства Чебышева (лемма 10.1) получаем

$$\frac{\varepsilon}{2} + P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \leq \frac{\varepsilon}{2} + \frac{M}{2n\delta^2}$$

Выбираем $n > \frac{M}{\varepsilon\delta^2}$, получаем оценку $|f(p) - B(f;p)| < \varepsilon$. □

10.3 Усиленный закон больших чисел

Теорема 10.3. *(Райхман) Пусть $X_1, X_2, \ldots$ — последовательность таких случайных величин, что $\exists C > 0\colon \quad \forall k \in \mathbb{N}\ \operatorname{var} X_k \leq C$. Предположим, что $\forall i \neq j\ \operatorname{cov}(X_i, X_j) = 0$. Тогда для $S_n := X_1 + \cdots + X_n$ выполнено*

$$\frac{S_n - \mathbb{E}\, S_n}{n} \xrightarrow[n\to+\infty]{\text{п.н.}} 0$$

Доказательство. Без ограничения общности рассуждений рассматриваем центрированные случайные величины, т.е. $\forall k \quad \mathbb{E}\, X_k = 0$. Рассмотрим натуральную последовательность m_n :

$$m_n^2 \leq n < (m_n + 1)^2$$

Тогда

$$\frac{|S_n|}{n} \leq \frac{|S_{m_n^2}| + |Z_{m_n^2}|}{m_n^2}, \qquad Z_{m_n^2} := \max_{k=0,\ldots,2m-1} \left|X_{m_n^2+1} + \cdots + X_{m_n^2+1+k}\right|$$

Понятно, что достаточно проверить

$$\frac{S_{m_n^2}}{m_n^2} \xrightarrow{\text{п.н.}} 0, \qquad n \to +\infty$$
$$\frac{Z_{m_n^2}}{m_n^2} \xrightarrow{\text{п.н.}} 0, \qquad n \to +\infty$$

Берем произвольный $\varepsilon > 0$.

$$\sum_{m=1}^{+\infty} P\left(\frac{|S_{m^2}|}{m^2} \geq \varepsilon\right) \leq \sum_{m=1}^{+\infty} \frac{\operatorname{var} S_m}{\varepsilon^2 m^4} \leq$$
$$\leq \sum_{m=1}^{+\infty} \frac{Cm^2}{\varepsilon^2 m^4} = \sum_{m=1}^{+\infty} \frac{C}{\varepsilon^2 m^2} = \frac{C}{\varepsilon^2 m^2} \sum_{m=1}^{+\infty} \frac{1}{m^2} < \infty$$

По лемме 4.3 Бореля-Кантелли с вероятностью 0 произойдет бесконечное число событий $\left\{\frac{|S_{m^2}|}{m^2} \geq \varepsilon\right\}$. Следовательно, с вероятностью 1 произойдет кроме быть может конечного числа $\left\{\frac{|S_{m^2}|}{m^2} < \varepsilon\right\}$. Получаем, что

$$\frac{S_{m_n^2}}{m_n^2} \xrightarrow{\text{п.н.}} 0, \qquad n \to +\infty$$

Аналогичным путём получается оценка и для Z_{m^2} □

Лемма 10.2. *Для любой случайной величины X справедливо двойное неравенство:*

$$\sum_{n=1}^{+\infty} P(|X| \geq n) \leq \mathbb{E}|X| \leq 1 + \sum_{n=1}^{+\infty} P(|X| \geq n)$$

Доказательство.

$$\sum_{n=1}^{+\infty} P(|X| \geq n) = \sum_{n=1}^{+\infty} \sum_{k \geq n} P(k \leq |X| < k+1) =$$
$$= \sum_{k=1}^{+\infty} kP(k \leq |X| < k+1) = \sum_{k=1}^{+\infty} \mathbb{E}(k\mathbb{1}(k \leq |X| < k+1)) \leq$$
$$\leq \sum_{k=1}^{+\infty} \mathbb{E}(|X|\mathbb{1}(k \leq |X| < k+1)) = \mathbb{E}|X| = \ldots$$

□

Из курса математического анализа знаем, что справедлива

Лемма 10.3. *(Тёплиц) Пусть $a_1, a_2, \ldots$ такова, что $a_r \to a < \infty$. Тогда* $\frac{1}{n}\sum_{r=1}^{n} a_r \xrightarrow[n\to\infty]{} a$.

Теорема 10.4. *(Колмогоров) Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины, $S_n = X_1 + \cdots + X_n$. Тогда*

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{п.н.}} c \in \mathbb{R} \iff \exists\, \mathbb{E}\, X_1 = c.$$

Доказательство. Покажем достаточность. Без ограничения общности рассуждений рассмотрим неотрицательные случайные величины $X_k \in L^1$. Введём $Y_r := X_r \mathbb{1}(0 \le X_r \le r), \quad r \in \mathbb{N}$. Положим $T_n := \sum_{r=1}^{n} Y_r$. Пусть для $\alpha > 1 \quad k_n := [\alpha^n], n \in \mathbb{N}$. Покажем, что

$$\frac{T_{k_n} - \mathbb{E}\, T_{k_n}}{k_n} \xrightarrow[n\to\infty]{\text{п.н.}} 0. \tag{10.1}$$

Фиксируем некоторый $\varepsilon > 0$:

$$\sum_{n\in\mathbb{N}} P\left(\left|\frac{T_{k_n} - \mathbb{E}\, T_{k_n}}{k_n}\right| \ge \varepsilon\right) \overset{\text{н-во } 10.1}{\le} \sum_{n=1}^{+\infty} \frac{\operatorname{var} T_{k_n}}{k_n^2 \varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{\sum_{r=1}^{k_n} \operatorname{var} Y_r}{k_n^2} \le$$
$$\le \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{\sum_{r=1}^{n} \mathbb{E}\, Y_r^2}{k_n^2} \le \frac{1}{\varepsilon^2} \sum_{r=1}^{\infty} \mathbb{E}\, Y_r^2 \sum_{n\colon k_n \ge r} \frac{1}{k_n^2} \tag{10.2}$$

Поскольку $\forall n \quad \alpha^n - 1 < \underbrace{[\alpha^n]}_{=k_n} \le \alpha^n$, получаем

$$\sum_{n\colon k_n\ge r} \frac{1}{k_n^2} = \sum_{n\colon [\alpha^n]\ge r} \frac{1}{[\alpha^n]} \le \sum_{n\colon \alpha^n \ge r} \frac{1}{(\alpha^n - 1)^2} \le$$
$$\le C \sum_{n \ge \log_\alpha r} \frac{1}{\alpha^{2n}} \le C \left(\frac{1}{\alpha^2}\right)^{\log_\alpha r} \frac{1}{1 - \frac{1}{\alpha^2}} \le \frac{D}{(1+r)^2},$$

где C, D – какие-то положительные константы, зависящие только от α.

$$\mathbb{E}\, Y_r^2 = \mathbb{E}\, X_r^2 \mathbb{1}(0 \le X_r \le r) = \int_{[0,r]} x^2 P_{X_1}(dx) = \sum_{j=0}^{r-1} \int_{\Delta_j} x^2 P_{X_1}(dx),$$

где $\Delta_j := (j, j+1]$, $\Delta_0 := [0,1]$. Вернемся к (10.2) и получим:

$$\frac{1}{\varepsilon^2} \sum_{r=1}^{\infty} \mathbb{E}\, Y_r^2 \sum_{n\colon k_n \ge r} \frac{1}{k_n^2} \le D \sum_{r=1}^{\infty} \mathbb{E}\, Y_r^2 \frac{1}{(1+r)^2} =$$
$$= D \sum_{r=1}^{\infty} \frac{1}{(r+1)^2} \sum_{j=0}^{r-1} \int_{\Delta_j} x^2 P_{X_1}(dx) = D \sum_{j=0}^{\infty} \int_{\Delta_j} x^2 P_{X_1}(dx) \sum_{r=j+1}^{\infty} \frac{1}{(r+1)^2} \le$$
$$\le D \sum_{j=0}^{\infty} \frac{1}{j+1} \int_{\Delta_j} x P_{X_1}(dx) = D \int_{\mathbb{R}^+} x P_{X_1}(dx) = D \; \mathbb{E}\, X_1 < \infty$$

Итак, по лемме 4.3

$$\frac{T_{k_n} - \mathbb{E}\, T_{k_n}}{k_n} \xrightarrow[n\to\infty]{\text{п.н.}} 0.$$

По лемме 10.3 для $a_r := \mathbb{E}\, X_r \mathbb{1}(0 \le X_r \le r)$ имеем

$$\frac{\mathbb{E}\, T_{k_n}}{k_n} \xrightarrow[n\to\infty]{} \mathbb{E}\, X_1$$

и из этого и (10.1)

$$\frac{T_{k_n}}{k_n} \xrightarrow[n\to\infty]{\text{п.н.}} \mathbb{E}\, X_1.$$

Заметим, что

$$\sum_{r=1}^{\infty} P(Y_r \neq X_r) = \sum_{r=1}^{\infty} P(X_r > r) \leq \sum_{r=1}^{\infty} P(X_1 \geq r) \leq 1 + \mathbb{E}\, X_1 < \infty$$

По лемме 4.3 почти наверное выполняется $Y_r = X_r \quad \forall r > r_0(\omega)$. Следовательно, $\frac{S_{k_n}}{k_n} \xrightarrow[n\to\infty]{\text{п.н.}} \mathbb{E}\, X_1$

$\forall n \in \mathbb{N}$ найдем $j_n \colon j_n \in \{[\alpha^m]\,, m \in \mathbb{N}\}$ и $j_n \leq n < j_{n+1}$. Имеем

$$\frac{S_n}{n} \geq \frac{S_{j_n}}{n} = \frac{S_{j_n}}{j_n}\frac{j_n}{j_{n+1}}$$
$$\frac{S_n}{n} \leq \frac{S_{j_{n+1}}}{n} = \frac{S_{j_{n+1}}}{j_{n+1}}\frac{j_{n+1}}{j_n}$$

и, как следствие,

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \mathbb{E}\, X_1 \cdot \frac{1}{\alpha}$$
$$\limsup_{n\to\infty} \frac{S_n}{n} \leq \mathbb{E}\, X_1 \cdot \alpha$$

Устремляя $\alpha \to 1$ получаем равенство

$$\lim_{n\to\infty} \frac{S_n}{n} \overset{\text{п.н.}}{=} \mathbb{E}\, X_1$$

□

11 Характеристические функции

Определение 11.1. Пусть X — случайная величина. *Характеристической функцией случайной величины* X назовем функцию $\varphi_X(t) := \mathbb{E}\, e^{itX}, \quad t \in \mathbb{R}$

Утверждение 11.1. $\varphi_X(t) = \mathbb{E} \cos tX + i\, \mathbb{E} \sin tX$

Доказательство. Следует из формулы Эйлера

$$e^{i\varphi} = \cos\varphi + i \sin\varphi$$

и линейности интеграла Лебега. □

Определение 11.2. Функция $\psi : \mathbb{R} \to \mathbb{C}$ называется *неотрицательно определенной*, если

$$\forall n \in \mathbb{N},\ \forall t_1, \dots, t_n \in \mathbb{R},\ \forall z_1, \dots, z_n \in \mathbb{C} \quad \sum_{k,l=1}^{n} z_k \overline{z_l}\ \psi(t_k - t_l) \geq 0$$

Теорема 11.1. *Справедливы следующие утверждения:*

1. $\varphi_X(0) = 1$
2. $|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}$

3. $\varphi_X(.)$ *равномерно непрерывна на* $\mathbb{R}$

4. φ_X *является неотрицательно определенной функцией*

Доказательство.

1. $\varphi_X(0) = \mathbb{E}\, e^{i0X} = \mathbb{E}\, 1 = 1$

2. $|\mathbb{E}\, Z| \le \mathbb{E}\, |Z|$

3. Без ограничения общности рассуждений $h > 0$

$$|\varphi_X(t+h) - \varphi_X(t)| = \left|\mathbb{E}\, e^{i(t+h)X} - \mathbb{E}\, e^{itX}\right| = \\ = \left|\mathbb{E}\, e^{itX}\left(e^{ihX} - 1\right)\right| \le 2h \sup_{\mathbb{R}} |X|$$

4. $n \in \mathbb{N},\ t_1, \ldots, t_n \in \mathbb{R},\ z_1, \ldots, z_n \in \mathbb{C}$

$$\sum_{k,l=1}^{n} z_k \overline{z_l}\ \varphi_X(t_k - t_l) = \sum_{k,l=1}^{n} z_k \overline{z_l}\ \mathbb{E}(e^{i(t_k - t_l)X}) = \\ = \sum_{k,l=1}^{n} z_k \overline{z_l}\ \mathbb{E}(e^{it_k X} e^{-it_l X}) = \sum_{k,l=1}^{n} z_k \overline{z_l}\ \mathbb{E}(e^{it_k X} \overline{e^{it_l X}}) = \\ = \mathbb{E}\left(\sum_{k=1}^{n} z_k e^{it_k X} \cdot \overline{\sum_{l=1}^{n} z_l e^{it_l X}}\right) = \mathbb{E}\left|\sum_{k=1}^{n} z_k e^{it_k X}\right|^2 \ge 0$$

□

Приведем без доказательства следующий факт:

Теорема 11.2. *(Bohner - Хинчин)* φ *является характеристической функцией тогда и только тогда, когда выполнены следующие условия:*

1. $\varphi(0) = 1$*;*

2. $\varphi \in C(\mathbb{R})$*;*[2]

3. φ *неотрицательно определенная функция.*

Теорема 11.3. *(Формула обращения) Пусть* φ_X *— характеристическая функция некоторой случайной величины* X*,* F_X *— функция распределения* X*. Тогда* $\forall a, b \in C(F_X)$[3] *выполнено*

$$F_X(b) - F_X(a) = \text{v. p.}\, \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-iat} - e^{-ibt}}{it} \varphi_X(t) dt$$

Доказательство.

$$\int_{-C}^{C} \left|\frac{e^{-ita} - e^{-itb}}{it} e^{itx}\right| dx \le 2C(b-a) \implies$$

[2] Можно доказать, что достаточно непрерывности в нуле

[3] Множество точек непрерывности функции F_X

применима теорема Фубини о перестановке повторных интегралов. По теореме Фубини и определению 11.1 характеристической функции,

$$\frac{1}{2\pi}\int_{-C}^{C}\frac{e^{-iat}-e^{-ibt}}{it}\varphi_X(t)dt =$$
$$= \frac{1}{2\pi}\int_{-C}^{C}\frac{e^{-iat}-e^{-ibt}}{it}\left(\int_{-\infty}^{+\infty}e^{itx}dF_X(x)\right)dt =$$
$$= \frac{1}{2\pi}\int_{-C}^{C}\left(\int_{-\infty}^{+\infty}\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}dF_X(x)\right)dt =$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-C}^{C}\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}dt\right)dF_X(x) =$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-C}^{C}\frac{e^{i(x-a)t}-e^{i(x-b)t}}{it}dt\right)dF_X(x) =$$
$$= \int_{-\infty}^{+\infty}\frac{1}{\pi}\left(\int_{0}^{C(x-a)}\frac{\sin u}{u}du - \int_{0}^{C(x-b)}\frac{\sin u}{u}du\right)dF_X(x)$$

Обозначим $\psi_C(x) := \frac{1}{\pi}\left(\int_0^{C(x-a)}\frac{\sin u}{u}du - \int_0^{C(x-b)}\frac{\sin u}{u}du\right)$. Понятно, что

$$\lim_{C\to+\infty}\psi_C(x) =: \psi(x) = \begin{cases}\frac{1}{2}, & x = a \text{ или } x = b\\ 1, & x\in(a,b)\\ 0, & x\notin[a,b]\end{cases}$$

Тогда из теоремы 8.3 Лебега о мажорируемой сходимости получаем

$$\lim_{C\to+\infty}\int_{-\infty}^{+\infty}\psi_C(x)dF_X(x) = \int_{-\infty}^{+\infty}\psi(x)dF_X(x) =$$
$$= 1\cdot Q((a,b)) + \tfrac{1}{2}\cdot Q(\{a\}) + \tfrac{1}{2}\cdot Q(\{b\}) = Q((a,b)) = F_X(b)-F_X(a),$$

где Q — вероятностная мера, порожденная распределением F_X □

12 Слабая сходимость вероятностных мер и характеристические функции.

12.1 Оценка модуля интеграла комплекснозначной функции

Кратко обсудим оценку модуля интеграла комплекснозначной функции, которую мы уже использовали (*лекция 8*). Мы используем частный случай, рассматривая:

$$\int_a^b h(t)\,dt, \text{ где } h(t) = u(t)+iv(t),\ t\in[a,b],\ i^2=-1$$

Пусть $a<b$. Тогда этот интеграл мы понимаем, как:

$$\int_a^b h(t)\,dt = \int_a^b u(t)\,dt + i\int_a^b v(t)\,dt$$

Определение 12.1. Функция $h(t)$ *измерима по мере Лебега*, если измеримы функции u и v, т.е.

$$h \in L^1(\text{по мере Лебега}),\ \text{если}\ u, v \in L^1$$

Замечание 12.1. Если $a > b$, то, по определению

$$\int_a^b h(t)\,dt = -\int_a^b h(t)\,dt$$

Лемма 12.1. *Пусть* $h \in L^1$, $a < b$. *Тогда*

$$\left|\int_a^b h(t)\,dt\right| \leqslant \int_a^b |h(t)|\,dt$$

Т.е. перепишем, использую действительную и мнимую часть:

$$\left(\left(\int_a^b u(t)\,dt\right)^2 + \left(\int_a^b v(t)\,dt\right)^2\right)^{\frac{1}{2}} \leqslant \int_a^b \sqrt{u^2(t) + v^2(t)}\,dt$$

Доказательство. Рассмотрим 2 случая:

1. $z = \int_a^b h(t)\,dt = 0$. Тогда очевидно верно

2. $z \neq 0$. Тогда $z = re^{-i\theta},\ r > 0 \quad (\theta = arctg(z))$. Тогда

$$r = ze^{-i\theta} \implies r = e^{-i\theta}\int_a^b h(t)\,dt = \int_a^b e^{-i\theta}h(t)\,dt =$$
$$= \int_a^b Re(e^{-i\theta}h(t))\,dt + i\int_a^b Im(e^{-i\theta}h(t))\,dt$$

Поскольку $r > 0$, то

$$\int_a^b Im(e^{-i\theta}h(t))\,dt = 0 \implies r = \int_a^b Re(e^{-i\theta}h(t))\,dt$$

Очевидно, что $|Re(w)| \leqslant |w|$. Тогда, т.к. $w = v + is$

$$|z| = r \leqslant \int_a^b \left|Re(e^{-i\theta}h(t)\right|\,dt \leqslant \int_a^b |e^{-i\theta}h(t)|\,dt = \int_a^b |h(t)|\,dt$$

Это так, потому что $|e^{-i\theta}| = 1$, точка на окружности.

□

Следствие 12.1. $|e^{i\alpha} - 1| \leqslant |\alpha|, \quad \alpha \in \mathbb{R}$.

Доказательство.

$$\int_a^\alpha e^{it}\,dt = \int_0^\alpha \cos t\,dt + i\int_0^\alpha \sin t\,dt = \sin t\,\Big|_0^\alpha - i\cos t\,\Big|_0^\alpha = -i(e^{i\alpha} - 1)$$

Тогда
$$\left|\int_0^\alpha e^{it}\,dt\right| = \left|-i(e^{i\alpha}-1)\right| = \left|e^{i\alpha}-1\right|$$
По лемме 12.1 выше
$$\left|\int_0^\alpha e^{it}\,dt\right| = \left|\int_0^{|\alpha|} e^{it}\,dt\right| \le \int_0^{|\alpha|} \left|e^{it}\right|\,dt = |\alpha|$$
Итак, мы можем написать, что (по формуле Эйлера)
$$\left|\int_0^w e^{it}\,dt\right| = \left|\int_a^w \cos t\,dt + i\int_0^w \sin t\,dt\right| = \left|-i(e^{iw}-1)\right| \implies$$
$$\implies \left|\int_0^w e^{it}\,dt\right| = |e^{iw}-1| \implies \textit{(по лемме 12.1)}$$
$$\implies \left|\int_0^w e^{iw}\,dt\right| = \left|\int_0^{|w|} e^{it}\,dt\right| \leqslant \int_0^{|w|} \left|e^{iw}\right|\,dt = |w| \implies$$
$$\implies \left|e^{iw}-1\right| \leqslant |\alpha|, \quad \alpha\in\mathbb{R}$$
И это то неравенство, которое мы использовали раньше. □

12.2 Следствие формулы обращения.

У нас была формула, которая позволяла по характеристической функции восстанавливать меру. Напомним, в более общей постановке, что было сделано.

Если дано пространство $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ и мера Q на $\mathcal{B}(\mathbb{R}^d)$, то характеристической функцией $\varphi_Q(t)$ называется (по определению 11.1)
$$\varphi_Q(t) \overset{def}{=} \int\limits_{\mathbb{R}^d} e^{i(t,x)}\,Q(dx), \qquad t\in\mathbb{R}^d$$
Здесь и далее (в этой лекции), запись (t,x) означает *скалярное произведение*:
$$(t,x) \overset{def}{=} \sum_{k=1}^n t_k x_k$$
Замечание 12.2. Если $x=(x_1,\ldots,x_d)$ — случайный вектор, то
$$\varphi_X(t) \overset{def}{=} \varphi_{P_X}(t) = \int\limits_{\mathbb{R}^d} e^{i(t,x)}\,P_X(dx) = \mathbb{E}e^{i(t,x)}$$
Теорема 12.1 (Единственность). *Пусть $\varphi_X(t)=\varphi_Y(t)$ при всех $t\in\mathbb{R}$, $d=1$. Тогда функция*
$$F_X(x)=F_Y(x) \quad \forall x\in\mathbb{R}$$
Иначе говоря, между характеристическими функциями и функциями распределения (а значит, и распределениями) существует взаимно-однозначное соответствие.

Доказательство. Пусть $D = C(F_X) \cap C(F_Y)$, $C(F_X)$ — точки непрерывности. Множество точек разрыва любой монотонной функции — *не более чем счётно.*

Тогда D получается из $\mathbb{R}$ удалением не более чем счетного числа точек. Тогда, для a, $b \in D$, $a \leqslant b$, получаем

$$F_X(b) - F_x(a) = F_Y(b) - F_Y(a)$$

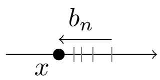
Это так по формуле обращения:

$$\lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-iat} - e^{-ibt}}{it} \varphi_X(t)\, dt, \quad \varphi_X(t) = \varphi_Y(t)$$

Устремляем $a \to -\infty$, $a \in D$. Тогда функция распределения сходится к нулю. Тогда

$$F_X(b) = F_Y(b) \quad \forall b \in D$$

Так как D получается выкидыванием не более чем счетного множества точек, то для любой точки x найдем последовательность точек $b_n \in D$, которая справа сходится к x.



Но функция распределения *непрерывна справа.* Значит,

$$F_X(x) = F_Y(x) \qquad \forall x \in \mathbb{R}$$

□

Теорема 12.2. *Если $\varphi_X(\cdot) \in L^1(\mathbb{R})$, т.е. $\int_{\mathbb{R}} |\varphi_X(t)| dt < \infty$, то случайная величина X имеет непрерывную плотность, которая задается с помощью обратного преобразования Фурье:*

$$p(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_x(t) dt, \quad x \in \mathbb{R}$$

Доказательство. Заметим, что p — непрерывная функция (это доказывается аналогично тому, как мы доказывали аналогичное утверждение для характеристической функции, когда вместо φ было P. Возьмем $a, b, \in C(F_X)$.

По теореме Фубини

$$\int_a^b p(x)\, dx = \int_a^b \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t)\, dt\, dx =$$
$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\varphi_X(t) \int_a^b e^{-itx} dx \right) dt = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) \frac{e^{-iat} - e^{-ibt}}{it}\, dt =$$
$$= \frac{1}{2\pi} \lim_{c\to\infty} \int_{-c}^{c} \varphi_X(t) \frac{e^{-iat} - e^{-ibt}}{it}\, dt = F_X(b) - F_X(a)$$

Пояснение.

$$\left| \varphi_X(t) \frac{e^{-ita} - e^{-itb}}{it} \right| \leqslant \underbrace{|\varphi_X(t)|}_{\leqslant 1} \left| \frac{e^{-ita} - e^{-itb}}{it} \right|$$

$$\left| \frac{e^{ita} = e^{-itb}}{it} \right| = \left| \frac{e^{iat}(1 - e^{-ibt+iat})}{t} \right| = \frac{\left| e^{-it(b-a)} - 1 \right|}{|t|} \leqslant \frac{|t(b-a)|}{|t|} = |b - a|$$

Выше мы применили лемму 12.1. Значит, функция интегрируема и применение теоремы Фубини законно. Итак, мы доказали, что

$$\forall a, b \in C(F_X): \quad \int_a^b p(x)dx = F_X(b) - F_X(a) \implies$$
$$\implies \forall x, y \in \mathbb{R} \quad \int_x^y p(s)ds = F_X(y) - F_X(x)$$

Замечание 12.3. $p(s) > 0$

Доказательство. Предположим, в некоторой точке x_0 $p(x_0) < 0$. Тогда в окрестности этой точки функция тоже будет строго меньше 0.

Возьмем $x, y, \ x < y$ из этой окрестности

$$\int_x^y p(s)ds) = F_X(y) - F_X(x) < 0$$

Это невозможно, так как $x < y$, функция $F_X(t)$ не убывает. □

Итак, по теореме о монотонной сходимости:

$$\int_{\mathbb{R}} p(s)ds = 1$$

Введем меру $Q(B) = \int_B p(s)\,dx$, где B — борелевское. Тем самым Q — вероятность. Значит

$$Q = P_X \text{ на } \{(a, b], -\infty < a \leqslant b < \infty\}$$

Но множество полуинтервалов— π-система $\implies Q = P_X$ на $\mathcal{B}(\mathbb{R})$ □

Замечание 12.4. Это формула, в отличие от общей, действует ***не всегда***, так как функция может быть и не интегрируема.

Если $d > 1$, $(a, b] = (a_1, b_1] \times \ldots \times (a_d, b_d]$. Исключим из $\mathbb{R}^d$ вдоль каждой оси не более чем счетное множество точек таких, что P_X от каждой гиперплоскости равно 0. Пусть D полученное множество точек для которых мера положительна. Тогда также получим

$$P_X((a, b]) = \lim_{c\to\infty} \frac{1}{(2\pi)^d} \int\limits_{[-c,c]} \prod_{k=1}^{d} \frac{e^{it_k a_k} - e^{-it_k b_k}}{it_k} \varphi_X(t)\,dt$$

Аналогично теореме 12.2:

Если $\varphi_X(t) \in L^1(\mathbb{R}^d)$, то $\exists$ непрерывная плотность

$$p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(tx)} \varphi_X(t)dt, \quad x \in \mathbb{R}^d$$

Теорема 12.3. *Для случайного вектора $X = (X_1, \ldots, X_d)$ формула*

$$\varphi_X(t) = \prod_{x=1}^{d} \varphi_{X_k}(t_k)$$

справедлива $\forall t = (t_1, \ldots, t_d) \in \mathbb{R}^d$ в том и только том случае,

когда $X_1, \ldots, X_d$ — независимые величины.

Доказательство. ($\Longrightarrow$) Пусть $z_1, \ldots, z_d$ — независимые комплекснозначные величины, т.е. $z_k = u_k + iv_k, \quad k = 1, \ldots, d$, и независимы $(u_1, v_1), \ldots, (u_d, v_d)$

Если $z_1, \ldots, z_d$ независимы и $z_k \in L^1(\Omega)$, $k = 1, \ldots, d$, то

$$\mathbb{E}(z_1, \ldots z_d) = \mathbb{E}((u_1 + iv_1) \cdot \ldots \cdot (u_d + iv_d)) = \mathbb{E}z_1 \cdot \ldots \cdot \mathbb{E}z_d$$

Если раскрыть скобки в произведении, то получатся слагаемые вида $w_1 \cdot \ldots \cdot w_d$, где $w_k = (u_k$ либо $iv_k)$. По лемме о группировке

$$\mathbb{E}(w_1 \cdot \ldots \cdot w_d) = \mathbb{E}w_1 \cdot \ldots \cdot \mathbb{E}w_d \qquad \mathbb{E}(v_1 \cdot \ldots \cdot v_d) = \mathbb{E}v_1 \cdot \ldots \cdot \mathbb{E}v_d \Longrightarrow$$
$$\Longrightarrow (\mathbb{E}u_1 + i\mathbb{E}v_1) \ldots (\mathbb{E}u_d + i\mathbb{E}v_d)$$

Независимость $x_1, \ldots, x_d$ влечет независимость величин

$$\cos t_1 x_1 + i \sin t_1 x_1, \ \ldots, \ \cos t_d x_d + i \sin t_d x_d$$

Тогда $\varphi_x(t) = \mathbb{E}e^{i(t,x)} = \mathbb{E}\left(\prod\limits_{k=1}^{d} e^{it_k x_k}\right) = \prod\limits_{k=1}^{d} \mathbb{E}e^{it_k x_k} = \prod\limits_{k=1}^{d} \varphi_{x_k}(t_k)$

($\Leftarrow$) Возьмем независимые $Y_1, \ldots, Y_d$ такие, что распределения совпадают:

$$P_{Y_k} = P_{x_k}, \ k = 1, \ldots, d$$

Это возможно по теореме Ламницкого-Улема. Тогда, так как распрделения совпадают:

$$\varphi_Y(t) = \prod_{k=1}^{d} \varphi_{Y_k}(t) = \prod_{k=1}^{d} \varphi_{x_k}(t) = \varphi_x(t)$$

По теореме единственности

$$P_X = P_Y = P_{Y_1} \times \ldots \times P_{Y_d} = P_{X_1} \times \ldots \times P_{X_d} \Longrightarrow x_1, \ldots x_d \text{ — независимы}$$

$\square$

Лемма 12.2. *Пусть $x_1, \ldots, x_n$ — независимые случайные векторы со значениями в $\mathbb{R}^m$. Тогда*

$$\varphi_{x_1 + \ldots + x_n}(t) = \prod_{1}^{n} \varphi_{x_k}(t), \quad t \in R^m$$

Доказательство. Сами. Доказательство аналогично доказательству предыдущей теоремы 12.3, экспонента распадается в произведение экспонент. $\square$

Замечание 12.5. Обратное не верно.

12.3 Фундаментальная теорема Прохорова

Определение 12.2. Семейство вероятностных мер $\{Q_n, n \in T\}$ заданных на метрическом пространстве $(S, \mathcal{B}(S))$ называется *слабо относительно компактным*, если из любой последовательности $(Q_n)_{n \in \mathbb{N}}$ можно извлечь слабо сходящуюся подпоследовательность $(Q'_n)_{n' \in \mathbb{N}}$ к некоторой мере Q.

Замечание 12.6. Эта мера Q автоматически вероятностная, но не обязана входить в рассматриваемое семейство.

Определение 12.3. Вероятностные меры Q_n *сходятся слабо* к Q ($Q_n \Rightarrow Q$) на метрическом пространстве $(S, \mathcal{B}(S))$, если $\forall f\colon S \to \mathbb{R},\ f \in \mathcal{B}(S)|\mathcal{B}(\mathbb{R})$, непрерывной и ограниченной:

$$\int_A f\, dQ_n \to \int_S f\, dQ$$

Определение 12.4. Семейство вероятностных мер $\{Q_n, n \in T\}$ называется *плотными*, если

$$\forall \epsilon > 0\ \exists \text{ компакт } K_\epsilon \subset S\colon Q_k(K_\epsilon) > 1 - \epsilon \quad \forall u \in T$$

Теорема 12.4 (Прохоров)**.** *Плотное семейство вероятностных мер является слабо отностительно компактным. Если пространство S польское (полное, сепарабельное), то любое слабо отностительно компактное семейство вероятностных мер является плотным.*

Таким образом, на польском пространстве семейство вероятностных мер плотно тогда и только тогда, когда оно слабо относительно компактно.

12.4 Критерий слабой сходимости вероятностных мер

Лемма 12.3. *Пусть $(Q_n)_{n\in\mathbb{N}}$ — плотное семейство вероятностных мер. Если каждая слабо сходящаяся подпоследовательность $(Q_{n'})$ сходится к одному и тому же пределу Q, то и $Q_n \Rightarrow Q,\ n \to \infty$.*

Доказательство. Предположим, что $Q_n \not\Rightarrow Q$. Тогда $\exists$ непрерывная и ограниченная $f\colon S \to \mathbb{R}$ такая, что

$$\int_S f\, dQ_n \not\to \int_S f\, dQ$$

Тогда существует $\epsilon >$ и подпоследовательность (n') такие, что

$$\left| \int_S f\, dQ_n - \int_S f\, dQ \right| > \epsilon$$

По теореме Прохорова 12.4 находим подпоследовательность $(n'') \subset (n')$, что

$$Q_{n''} \Rightarrow \widetilde{Q}$$

По условию, $\widetilde{Q} = Q$. Следовательно,

$$\left| \int_S f(x)\, Q_{n''}(dx) - \int_S f(x)\, Q(dx) \right| \to 0$$

Приходим к противоречию. □

Лемма 12.4. *Пусть (Q_n) — плотное семейство мер на $\{\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\}$ Тогда последовательность (Q_n) имеет слабый предел в том и только том случае, когда*

$$\forall t \in \mathbb{R}^d \quad \exists \lim_{n\to\infty} \varphi_n(t)$$

Доказательство. Если $Q_n \Rightarrow Q$, то мы уже видели, что $\varphi_n(t) \to \varphi(t)\ \forall t \in \mathbb{R}^d$

Обратно. Пусть $\varphi_n(t) \to \varphi(t) \quad \forall t \in \mathbb{R}^d,\ n \to \infty$.

Поскольку семейство Q_n плотно по условию, найдется подпоследовательность (n') такая, что $Q_{n'} \Rightarrow Q$, где Q некоторая вероятностная мера.

Предположим, что Q_n не имеет слабого предела. Тогда по лемме 12.3 найдется подпоследовательность $(n'') \subset (n')$ такая, что $Q_{n''} \Rightarrow \widetilde{Q}, \quad \widetilde{Q} \neq Q$.

Тогда по условию $\forall t \in \mathbb{R}^d \quad \exists \lim \varphi_n(t)$, поэтому:

$$\lim_{n'\to\infty} \int\limits_{\mathbb{R}^d} e^{i(t,x)}\, Q_{n'}(dx) = \lim_{n''\to\infty} \int\limits_{\mathbb{R}} e^{i(t,x)}\, Q_{n''}(dx) \implies$$

$$\implies \int_{\mathbb{R}^d} e^{i(t,x)}\, Q(dx) = \int_{\mathbb{R}^d} e^{i(t,x)}\, \widetilde{Q}(dx)$$

По теореме единственности 12.1, получаем, что $Q = \widetilde{Q}$. Противоречие. $\square$

Лемма 12.5. *Пусть φ — характеристическая функция меры Q на $\{\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\}$. Тогда $\forall a > 0$*

$$Q\left(\mathbb{R}^d \setminus \left[-\frac{2}{a}, \frac{2}{a}\right]^d\right) \leqslant \frac{2}{(2a)^d} \int\limits_{[-a,a]^d} (1 - Re(\varphi(t))\, dt$$

Доказательство. Будем доказывать для $d = 1$.

$$Re(\varphi(t)) = \int\limits_{\mathbb{R}} \cos tx\, Q(dx)$$

По теореме Фубини

$$\frac{1}{2a}\int_{-a}^{a} (1 - Re\,\varphi(t))\, dt = \frac{1}{2a} \int\limits_{[-a,a]} \int\limits_{\mathbb{R}} (1 - \cos tx)\, Q(dx)\, dt =$$

$$= \int\limits_{\mathbb{R}} \frac{1}{2a} \int_{-a}^{a} (1 - \cos tx)\, dt\, Q(dx)$$

$$\int_{-a}^{a} \cos tx\, dt = \frac{\sin tx}{x}\Bigg|_{-a}^{a} = \frac{2 \sin ax}{x} \qquad (x \neq 0)$$

Теорему Фубини применять корректно, так как подынтегральная функция ограничена, а значит — интегрируема.

Итак,

$$\frac{1}{2a}\int_{-a}^{a} (1 - Re\,\varphi(t))\, dt = \int\limits_{\mathbb{R}} \left(1 - \frac{\sin ax}{ax}\right) Q(dx) \geqslant$$

$$\geqslant \int\limits_{\mathbb{R}\setminus\left[-\frac{2}{a},\frac{2}{a}\right]} \left(1 - \frac{\sin ax}{ax}\right) Q(dx) \geqslant \frac{1}{2} Q\left(\mathbb{R} \setminus \left[-\frac{2}{a}, \frac{2}{a}\right]\right)$$

Пояснение

$$\left|\frac{\sin u}{u}\right| \leqslant \frac{1}{|u|} \leqslant \frac{1}{2}, \qquad |u| \geqslant 2$$

$\square$

Теорема 12.5 (непрерывность, Леви). *Пусть $(Q_n)_{n\in\mathbb{N}}$ последовательность вероятностных мер на $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $(\varphi_n)_{n\in\mathbb{N}}$ последовательность характеристических функций этих мер. Тогда справедливы следующие утверждения:*

1. *Если $Q_n \Rightarrow Q$, то $\varphi_n(t) \to \varphi(t)$, $n \to \infty \quad \forall t \in \mathbb{R}^d$, где φ — характеристическая функция Q*
2. *Если $\varphi_n(t) \to \varphi(t) \quad \forall t \in \mathbb{R}^d$, $n \to \infty$ и φ непрерывна в точке $0 \in \mathbb{R}^d$, то φ является характеристической функцией некоторой вероятностной меры Q и $Q_n \Rightarrow Q$, $n \to \infty$.*

Доказательство. Утверждение 1 очевидно, так как

$$\varphi_n(t) = \int_{\mathbb{R}^d} e^{i(t,x)}\, Q_n(dx) \to \int_{\mathbb{R}^d} e^{(t,x)}\, Q(dx) = \varphi(t) \qquad \forall t \in \mathbb{R}^d$$

Для доказательства второго утверждения понадобятся леммы 12.3, 12.4 и 12.5.

Все будем доказывать для случая $d = 1$.

$$\varphi_n(t)| \leqslant 1 \implies |\varphi| \leqslant 1$$

Заметим, что φ измеримо как предел измеримых функций. Значит, φ — интегрируема. φ непрерывно в точке 0. Значит,

$$\forall \varepsilon > 0\ \exists a = a(\varepsilon)\colon 0 \leqslant 1 - Re\,\varphi(t) \leqslant \frac{\varepsilon}{4}$$

Замечание 12.7. Комплекснозначная функция непрерывна тогда и только тогда, когда непрерывны её комплексная и вещественная часть.

Заметим, что $\varphi(0) = 1$, так как $\varphi_n(0) = 1$.

Тогда

$$\frac{1}{a}\int_{-a}^{a} \underbrace{(1 - Re\,\varphi(t))}_{\leqslant \frac{\epsilon}{4}}\, dt \leqslant \frac{\epsilon}{2}$$

По теореме Лебега о мажорируемой сходимости

$$\frac{1}{a}\int_{-a}^{a} (1 - Re\,\varphi_n(t))\, dt \to \frac{1}{a}\int_{-a}^{a} (1 - Re\,\varphi(t))\, dt$$

Значит, $\forall n \geqslant N_0(\epsilon)$

$$\frac{1}{a}\int_{-a}^{a} (1 - Re\,\varphi_n(t))\, dt \leqslant \epsilon$$

Тогда по лемме 12.5

$$Q\left(\mathbb{R} \setminus \left[-\frac{2}{a}, \frac{2}{a}\right]\right) < \epsilon, \qquad n > N_0$$

Для любой вероятностной меры на $\{\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\}$ верно

$$\forall \epsilon > 0\ \exists b = b(\epsilon)\colon P(\mathbb{R} \setminus [-b, b] < \epsilon$$

Здесь мы воспользовались тем, что $[-b,b] \to \mathbb{R}$ и непрерывностью вероятностной меры.

У нас есть конечное число мер $Q_{n_1}, \ldots, Q_{N_0}$. Найдем $b_n = b_n(\epsilon)$ такое, что

$$Q_n(\mathbb{R} \setminus [-b_n, b_n]) < \epsilon, \qquad n = 1, \ldots, N_0$$

Возьмем

$$b_0 = \max\left\{\frac{2}{a(\epsilon)}, b_1(\epsilon), \ldots, b_{N_0}(\epsilon)\right\}$$

Тогда $\forall n \in \mathbb{N} \quad Q_n(\mathbb{R} \setminus [-b_0, b_0]) < \epsilon$. Плотность семейства мер Q_n установлена.

По лемме 12.4 получаем, что существует слабый предел Q_n — Q. Тогда

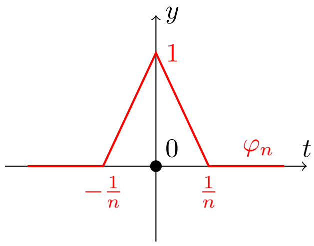
$$\psi(t) = \varphi(t), \qquad t \in \mathbb{R}$$

где ψ — характеристическая функция меры Q. Следовательно, φ — характеристическая функция меры Q и $Q_n \Rightarrow Q$ $\square$

Следствие 12.2. *Если Q, Q_n — вероятностные меры, то $Q_n \Rightarrow Q$ тогда и только тогда, когда $\varphi_n(t) \to \varphi(t) \quad \forall t \in \mathbb{R}$*

Пример 12.1. *Условие непрерывности в теореме Леви существенно.*

Возьмем характеристические функции такого вида (см. рис.). Это будут характеристические функции, так как:



Доказательство. Если бы φ_n *была* характеристической функций, то она была бы интегрируема. А значит, существует плотность (и она задается преобразованием Фурье). Эта плотность неотрицательна и интеграл равен 1. Значит, это плотность и мы нашли распределение, которому отвечает φ_n. $\square$

Заметим, что

$$\varphi_n(t) \to \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{if } t \neq 0 \end{cases}$$

Но предельная функция — не характеристическая, так как не непрерывна.

13 Центральная предельная теорема

13.1 Вспомогательные результаты

В дальнейшем будем пользоваться доказанными ранее неравенствами из леммы 12.1 и следствия ??.

При $a \leqslant b$ и $f : [a,b] \mapsto \mathbb{C}$

$$\left|\int\limits_a^b f(t)\,dt\right| \leqslant \int\limits_a^b |f(t)|\,dt$$

$\forall \alpha \in \mathbb{R}$

$$\left|e^{i\alpha} - 1\right| \leqslant |\alpha|$$

Также полезна следующая тривиальная оценка:

$$\left|e^{i\alpha} - 1\right| \leqslant \left|e^{i\alpha}\right| + |1| = 2$$

Лемма 13.1. $\forall u \in \mathbb{R}$, $\forall n \in \mathbb{Z}_+$ *верно неравенство:*

$$\left|e^{iu} - \sum_{k=0}^{n} \frac{(iu)^k}{k!}\right| \leqslant \frac{2|u|^n}{n!} \wedge \frac{|u|^{n+1}}{(n+1)!}$$

где $a \wedge b = \min\{a, b\}$.

Доказательство. Достаточно доказать, что модуль разности оценивается каждой из дробей в правой части неравенства. Докажем по индукции.
База: $n = 0$. Согласно неравенствам выше: $\left|e^{i\alpha} - 1\right| \leqslant 2 \wedge |\alpha|$.
Пусть верно для n, докажем для $n+1$. Обозначим $h_n(u) := e^{iu} - \sum\limits_{k=0}^{n} \frac{(iu)^k}{k!}$.
Тогда

$$\left|\int\limits_0^u h_{n-1}(s)\, ds\right| = \left|\frac{1}{i} h_n(u)\right| \implies |h_n(u)| = \left|\int\limits_0^u h_{n-1}(s)\, ds\right|$$

Применим предположение индукции и оценку модуля интеграла интегралом модуля:

$$|h_n(u)| \leqslant \int\limits_0^{|u|} |h_{n-1}(s)| \leqslant \int\limits_0^{|u|} \frac{s^n}{n!} = \left.\frac{s^{n+1}}{(n+1)!}\right|_0^{|u|} = \frac{|u|^{n+1}}{(n+1)!}$$

Для другой оценки:

$$\left|e^{iu} - \sum_{k=0}^{n} \frac{(iu)^k}{k!}\right| \leqslant \left|e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!}\right| + \left|\frac{(iu)^n}{n!}\right| \leqslant \frac{|u|^n}{n!} + \frac{|u|^n}{n!} = \frac{2|u|^n}{n!}$$

□

Следствие 13.1. *Пусть X — случайная величина, для которой* $\mathbb{E}\,|X| < \infty$. *Тогда $\forall t \in \mathbb{R}$, $\forall \varepsilon > 0$ справедливо неравенство:*

$$\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\,\mathbb{E}\,X^k\right| \leqslant$$
$$\leqslant \frac{2|t|^n}{n!}\,\mathbb{E}\Big(|X|^n\,\mathbb{1}\{|X| > \varepsilon\}\Big) + \frac{|t|^{n+1}}{(n+1)!}\,\mathbb{E}\left(|X|^{n+1}\,\mathbb{1}\{|X| \leqslant \varepsilon\}\right),$$

где $\varphi_X(t)$ — характеристическая функция случайной величины X.

Доказательство.

$$\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\,\mathbb{E}\,X^k\right| = \left|\mathbb{E}\left(e^{itX} - \sum_{k=0}^{n} \frac{(it)^k}{k!} X^k\right)\right| \leqslant \mathbb{E}\left|e^{itX} - \sum_{k=0}^{n} \frac{(it)^k}{k!} X^k\right| =$$
$$= \mathbb{E}\left(\left|e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!}\right| \mathbb{1}\{|X| > \varepsilon\}\right) + \mathbb{E}\left(\left|e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!}\right| \mathbb{1}\{|X| \leqslant \varepsilon\}\right)$$

Пользуясь леммой 13.1, выражения с модулем под знаком математического ожидания в обоих слагаемых можно оценить первой и второй дробью из доказанной оценки соответственно. □

Лемма 13.2. *$\forall z_1, \dots, z_m \in \mathbb{C}$, $\forall w_1, \dots, w_m \in \mathbb{C}$ таких, что $\forall k = 1, \dots m$ $|z_k| \leqslant 1$, $|w_k| \leqslant 1$ верно неравенство:*

$$|z_1 z_2 \dots z_m - w_1 w_2 \dots w_m| \leqslant \sum_{k=1}^{m} |z_k - w_k|$$

Доказательство. Докажем по индукции. При $m = 1$ неравенство обращается в тождество. Пусть верно для $m-1$, тогда докажем для m.

$$\begin{aligned}
&|z_1 z_2 \dots z_m - w_1 w_2 \dots w_m| \leqslant \\
&\leqslant |z_1 z_2 \dots z_m - w_1 z_2 \dots z_m| + |w_1 z_2 \dots z_m - w_1 w_2 \dots w_m| = \\
&= |z_1 - w_1||z_2 \dots z_m| + |w_1||z_2 \dots z_m - w_2 \dots w_m| \leqslant \\
&\leqslant |z_1 - w_1| + \sum_{k=2}^{m} |z_k - w_k| = \sum_{k=1}^{m} |z_k - w_k|
\end{aligned}$$

□

13.2 Теорема Линдеберга

Теорема 13.1. *(Центральная предельная теорема, Линдеберг) Пусть $\{X_{n,k},\, k = 1, \dots, m_n,\, n \in \mathbb{N}\}$ — массив независимых при каждом n случайных величин $X_{n,1} \dots, X_{n,m_n}$ (этот массив также называется схемой серий, а последовательность случайных величин при фиксированном n называется n-ой серией). Пусть также случайные величины $X_{n,k}$ центрированные. Предположим, что*

$$\sum_{k=1}^{m_n} \operatorname{var} X_{n,k} \to \sigma^2 \geqslant 0,\, n \to \infty$$

и выполнено условие Линдеберга на функцию Ляпунова $\mathcal{L}_n$:

$$\forall \varepsilon > 0\, \mathcal{L}_n(\varepsilon) := \sum_{k=1}^{m_n} \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) \to 0,\, n \to \infty.$$

Тогда $S_n := \sum\limits_{n=1}^{m_n} X_{n,k} \xrightarrow{\mathcal{D}} Z \sim N(0, \sigma^2)$

Доказательство. Пусть наряду с массивом $\{X_{n,k},\, k = 1, \dots, m_n,\, n \in \mathbb{N}\}$ задан массив случайных величин $\{Y_{n,k},\, k = 1, \dots, m_n,\, n \in \mathbb{N}\}$, где $\forall n\, Y_{n,1}, \dots Y_{n,m_n}$ — независимые случайные величины, при этом $Y_{n,k} \sim N(0, \sigma_{n,k}^2)$, где $\sigma_{n,k}^2 = \operatorname{var} X_{n,k}$. Такие случайные величины можно взять по теореме Ломницкого—Улама.

Определим $T_n := \sum\limits_{n=1}^{m_n} Y_{n,k}$. Тогда

$$T_n \sim N(0, \sum_{n=1}^{m_n} \sigma_{n,k}^2),$$

поскольку характеристическая функция T_n есть произведение характеристических функций $Y_{n,k}$, имеющих нормальное распределение, а характеристическая функция нормального распределения с параметрами a, σ^2 равна $\varphi(t) = \exp\left\{iat - \frac{\sigma^2 t^2}{2}\right\}$. Из взаимооднозначного соответствия между характеристическими функциями и случайными величинами следует, что T_n распределена нормально с параметрами 0 и $\sum\limits_{n=1}^{m_n} \sigma_{n,k}^2$. Более того:

$$\varphi_{T_n}(t) = \exp\left\{-\sum_{n=1}^{m_n} \frac{\sigma_{n,k}^2 t^2}{2}\right\} \underset{n\to\infty}{\to} e^{-\frac{\sigma^2 t^2}{2}}\ \forall t \in \mathbb{R}$$

Тогда по теореме Леви о непрерывности 12.5 $T_n \xrightarrow{\mathcal{D}} N(0, \sigma^2)$. Для доказательства теоремы достаточно показать, что $\varphi_{S_n}(t) \underset{n\to\infty}{\to} e^{-\frac{\sigma^2 t^2}{2}}\ \forall t \in \mathbb{R}$ и опять применить теорему о непрерывности. Для этого, в свою очередь, достаточно показать, что $\varphi_{S_n}(t) - \varphi_{T_n}(t) \underset{n\to\infty}{\to} 0\, \forall t \in \mathbb{R}$. Действительно:

$$|\varphi_{S_n}(t) - \varphi_{T_n}(t)| = \left|\prod_{k=1}^{m_n} \varphi_{X_{n,k}}(t) - \prod_{k=1}^{m_n} \varphi_{Y_{n,k}}(t)\right| \leqslant \sum_{k=1}^{m_n} \left|\varphi_{X_{n,k}}(t) - \varphi_{Y_{n,k}}(t)\right|$$

В силу независимости между собой $X_{n,k}$ и $Y_{n,k}$, а также леммы 13.2 ($|\varphi(t)| \leqslant 1\, \forall t \in \mathbb{R}$ по свойству характеристической функции).

$$\sum_{k=1}^{m_n} \left|\varphi_{X_{n,k}}(t) - \varphi_{Y_{n,k}}(t)\right| \leqslant$$
$$\leqslant \sum_{k=1}^{m_n} \left|\varphi_{X_{n,k}}(t) - 1 + \frac{\sigma_{n,k}^2 t^2}{2}\right| + \sum_{k=1}^{m_n} \left|\varphi_{Y_{n,k}}(t) - 1 + \frac{\sigma_{n,k}^2 t^2}{2}\right|$$

По лемме 13.1:

$$\sum_{k=1}^{m_n} \left|\varphi_{X_{n,k}}(t) - 1 + \frac{\sigma_{n,k}^2 t^2}{2}\right| + \sum_{k=1}^{m_n} \left|\varphi_{Y_{n,k}}(t) - 1 + \frac{\sigma_{n,k}^2 t^2}{2}\right| \leqslant$$
$$\leqslant \sum_{k=1}^{m_n} \left(t^2\, \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) + |t|^3\, \mathbb{E}\left(|X_{n,k}|^3\, \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right)\right) +$$
$$+ |t|^3 \sum_{k=1}^{m_n} \mathbb{E}\left(|Y_{n,k}|^3\right) \leqslant t^2 \mathcal{L}_n(\varepsilon) + \varepsilon\, |t|^3 \sum_{k=1}^{m_n} \sigma_{n,k}^2 + |t|^3 \sum_{k=1}^{m_n} \mathbb{E}\left(|Y_{n,k}|^3\right).$$

Последняя оценка получена подстановкой функции Ляпунова по определению и неравенством:

$$\mathbb{E}\left(|X_{n,k}|^3\, \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right) = \mathbb{E}\left(|X_{n,k}| \cdot X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right) \leqslant$$
$$\leqslant \mathbb{E}\left(\varepsilon X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right) = \varepsilon\, \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right) \leqslant \varepsilon\, \mathbb{E}\, X_{n,k}^2 = \varepsilon \sigma_{n,k}^2$$

Равенство в конце последней строчки выполнено, поскольку случайные величины центрированные.

Если $Z \sim N(0, \sigma^2)$, то

$$\mathbb{E}\,|Z|^3 = \int\limits_{-\infty}^{+\infty} \frac{|x|^3}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\,dx = 2\int\limits_{0}^{+\infty} \frac{x^3}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\,dx = 2\sigma^3 \int\limits_{0}^{+\infty} \frac{u^3}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\,du = \frac{4\sigma^3}{\sqrt{2\pi}}$$

Последние два перехода выполнены с помощью замены $u = \frac{x}{\sigma}$ и интегрирования по частям.
Отсюда получаем:

$$\sum_{k=1}^{m_n} \mathbb{E}\left(|Y_{n,k}|^3\right) \leqslant \frac{4}{\sqrt{2\pi}} \sum_{k=1}^{m_n} \sigma_{n,k}^3 \leqslant 2 \max_{k=1,\dots,m_n} \sigma_{n,k} \sum_{k=1}^{m_n} \sigma_{n,k}^2.$$

При этом $\sum\limits_{k=1}^{m_n} \sigma_{n,k}^2 \to \sigma^2$ по условию. Также

$$\sigma_{n,k}^2 = \mathbb{E}\,X_{n,k}^2 = \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) + \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| \leqslant \varepsilon\right\}\right) \leqslant$$
$$\leqslant \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) + \varepsilon^2 = \mathcal{L}_n(\varepsilon) + \varepsilon^2 \implies \max_{k=1,\dots,m_n} \sigma_{n,k} \leqslant \mathcal{L}_n(\varepsilon) + \varepsilon^2 \underset{n\to\infty}{\to} \varepsilon^2$$

Тогда для достаточно большого номера n верно, что $\max\limits_{k=1,\dots,m_n} \sigma_{n,k} \leqslant \frac{3}{2}\varepsilon^2$.
Итого получаем, что для любого фиксированного вещественного t и достаточно большого n верно неравенство:

$$|\varphi_{S_n}(t) - \varphi_{T_n}(t)| \leqslant t^2 \mathcal{L}_n(\varepsilon) + \varepsilon\,|t|^3 \sum_{k=1}^{m_n} \sigma_{n,k}^2 + 2|t|^3 \max_{k=1,\dots,m_n} \sigma_{n,k} \sum_{k=1}^{m_n} \sigma_{n,k}^2 \leqslant$$
$$\leqslant t^2 \mathcal{L}_n(\varepsilon) + \varepsilon\,|t|^3 \sum_{k=1}^{m_n} \sigma_{n,k}^2 + 3|t|^3 \varepsilon^2 \sum_{k=1}^{m_n} \sigma_{n,k}^2 \underset{n\to\infty}{\to} \varepsilon\,|t|^3\,\sigma^2 + 3|t|^3 \varepsilon^2 \sigma^2$$

В силу произвольности ε получаем требуемый поточечный предел:

$$|\varphi_{S_n}(t) - \varphi_{T_n}(t)| \underset{n\to\infty}{\to} 0$$

□

Теорема 13.2. *(Феллер)*
Пусть даны серии независимых случайных величин $X_{n,1}, \dots, X_{n,m_n}$,
$\forall k\ \mathbb{E}\,X_{n,k} = 0$, $\operatorname{var} X_{n,k} = \sigma_{n,k}^2$. *Пусть* $\sum\limits_{k=1}^{m_n} \sigma_{n,k}^2 \to \sigma^2 \geqslant 0$. *Предположим, что*

$$\boxed{\max_{k=1,\dots,m_n} \sigma_{n,k}^2 \to 0,\ n \to \infty}$$

Если $S_n \to N(0, \sigma^2)$, *то справедливо условие Линдеберга.*
Иначе говоря, для таких случайных величин, которые удовлетворяют условию пренебрежимой малости (обведено в рамку) выполнение условия Линдеберга необходимо и достаточно для справедливости Центральной предельной теоремы 13.1.

Теорема 13.3. *(Ляпунов)*
Пусть для массива независимых случайных величин $X_{n,1}, \ldots, X_{n,m_n}$
$\forall k\ \mathbb{E}\, X_{n,k} = 0,\ \mathbb{E}\, |X_{n,k}|^s < \infty$ *для некоторого* $s > 2$ *(в теореме 13.1* $s = 2$*).*
Как и ранее, предполагаем, что $\sum\limits_{k=1}^{m_n} \sigma^2_{n,k} \underset{n\to\infty}{\to} \sigma^2 \geqslant 0$*, где* $\sigma^2_{n,k} = \operatorname{var} X_{n,k}$*.*
Тогда из условия Ляпунова $\sum\limits_{k=1}^{m_n} \mathbb{E}\, |X_{n,k}|^s \to 0,\ n \to \infty$ *следует справедливость центральной предельной теоремы 13.1, то есть*

$$S_n \xrightarrow{\mathcal{D}} Z \ \sim\ N(0, \sigma^2).$$

Доказательство. По определению функции Ляпунова

$$\mathcal{L}_n(\varepsilon) = \sum_{k=1}^{m_n} \mathbb{E}\left(X^2_{n,k} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right).$$

Оценим каждое из слагаемых следующим образом

$$X^2_{n,k} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\} \leqslant \frac{|X_{n,k}|^{s-2}}{\varepsilon^{s-2}} X^2_{n,k} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\} = \frac{|X_{n,k}|^s}{\varepsilon^{s-2}} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\} \leqslant$$
$$\leqslant \frac{|X_{n,k}|^s}{\varepsilon^{s-2}} \implies \mathbb{E}\left(X^2_{n,k} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) \leqslant \frac{1}{\varepsilon^{s-2}}\, \mathbb{E}\, |X_{n,k}|^s$$

А тогда

$$\mathcal{L}_n(\varepsilon) \leqslant \frac{1}{\varepsilon^{s-2}} \sum_{k=1}^{m_n} \mathbb{E}\, |X_{n,k}|^s \underset{n\to\infty}{\to} 0$$

Применяя теорему 13.1, получаем требуемое. $\square$

Теорема 13.4. *Пусть* $X_1, X_2, \ldots$ *— последовательность независимых одинаково распределённых случайных величин таких, что*
$\mathbb{E}\, X_1 = a,\ \operatorname{var} X_1 = \sigma^2 \geqslant 0$*. Тогда*

$$\frac{S_n - na}{\sqrt{n}} \xrightarrow{\mathcal{D}} W \sim N(0, \sigma^2),\ n \to \infty,$$

где $S_n = X_1 + \ldots + X_n$*.*

Доказательство. Введём схему серий. $X_{n,k} := \frac{X_k - a}{\sqrt{n}},\ k = 1, \ldots, n$. Для такой схемы $m_n = n$. Тогда $\frac{S_n - na}{\sqrt{n}} = \sum\limits_{k=1}^{n} X_{n,k}$. Подстановкой $X_{n,k}$ по определению убеждаемся, что $\sum\limits_{k=1}^{n} \operatorname{var} X_{n,k} = \sigma^2$. Теперь достаточно проверить условие Линдеберга.

$$\mathcal{L}_n(\varepsilon) = \sum_{k=1}^{n} \mathbb{E}\left(X^2_{n,k} \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) = \mathbb{E}\left((X_1 - a)^2\, \mathbb{1}\left\{|X_1 - a| > \varepsilon\sqrt{n}\right\}\right)$$

Последнее равенство выполнено в силу одинаковой распределённости случайных величин и тождества $X^2_{n,k} = \left(\frac{X_k - a}{\sqrt{n}}\right)^2 = \frac{(X_k - a)^2}{n}$.
По теореме Лебега о мажорируемой сходимости

$$\mathbb{E}\left((X_1 - a)^2\, \mathbb{1}\left\{|X_1 - a| > \varepsilon\sqrt{n}\right\}\right) \underset{n\to\infty}{\to} 0$$

$\square$

Теорема 13.5. *Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин (не обязательно одинаково распределённых) таких, что $|X_n| \leqslant K$ для некоторого K и всех n. Предположим, что $\operatorname{var} S_n \to \infty$, $n \to \infty$, где $S_n = X_1 + \ldots + X_n$. Тогда*

$$\frac{S_n - \mathbb{E}\, S_n}{\sqrt{\operatorname{var} S_n}} \xrightarrow{\mathcal{D}} Z \sim N(0,1),\ n \to \infty$$

Доказательство. Из ограниченности случайных величин следует существование и конечность матожиданий, а также существование конечных дисперсий. В силу независимости

$$\operatorname{var} S_n = \sum_{k=1}^{n} \operatorname{var} X_k,$$

то есть для всякого n у величины S_n существует конечная дисперсия. Заметим, что случай, когда все X_n являются почти наверно константами невозможен, поскольку в этом и только в этом случае $\operatorname{var} S_n$ равен 0 для всякого n, что противоречит стремлению дисперсии к бесконечности. Поскольку есть X_N, не являющийся тождественной константой, получаем существование ненулевой дисперсии у S_n начиная с номера N.
Рассмотрим схему серий

$$X_{n,k} := \frac{X_k - \mathbb{E}\, X_k}{\sqrt{\operatorname{var} S_n}},\ k = 1, \ldots, n.$$

Проверим для неё выполнение условий теоремы 13.1.

$$\sum_{k=1}^{n} \operatorname{var} X_{n,k} = \frac{1}{\operatorname{var} S_n} \sum_{k=1}^{n} \operatorname{var}\left(X_k - \mathbb{E}\, X_k\right) = \frac{1}{\operatorname{var} S_n} \sum_{k=1}^{n} \operatorname{var} X_k = 1 = \sigma^2.$$

$$\mathcal{L}_n(\varepsilon) = \sum_{k=1}^{n} \mathbb{E}\left(X_{n,k}^2 \mathbb{1}\left\{|X_{n,k}| > \varepsilon\right\}\right) =$$

$$\frac{1}{\operatorname{var} S_n} \sum_{k=1}^{n} \mathbb{E}\left(\left(X_k - \mathbb{E}\, X_k\right)^2 \mathbb{1}\left\{|X_k - \mathbb{E}\, X_k| > \varepsilon\sqrt{\operatorname{var} S_n}\right\}\right) \leqslant$$

$$\leqslant \frac{1}{\operatorname{var} S_n} \sum_{k=1}^{n} \mathbb{E}\left(4K^2 \mathbb{1}\left\{|X_k - \mathbb{E}\, X_k| > \varepsilon\sqrt{\operatorname{var} S_n}\right\}\right) =$$

$$= \frac{4K^2}{\operatorname{var} S_n} \sum_{k=1}^{n} P\left(\left\{|X_k - \mathbb{E}\, X_k| > \varepsilon\sqrt{\operatorname{var} S_n}\right\}\right)$$

По неравенству Бьенеме—Чебышёва:

$$\frac{4K^2}{\operatorname{var} S_n} \sum_{k=1}^{n} P\left(\{|X_k - \mathbb{E}\, X_k| > \varepsilon \operatorname{var} S_n\}\right) \leqslant \frac{4K^2}{\operatorname{var} S_n} \sum_{k=1}^{n} \frac{\operatorname{var} X_k}{\operatorname{var} S_n \varepsilon^2} = \frac{4K^2}{\varepsilon^2 \operatorname{var} S_n} \to 0.$$

Таким образом выполнено условие теоремы Линдеберга, а поскольку $\sigma^2 = 1$, то получаем требуемое. $\square$

13.3 Неклассические условия ЦПТ

Рассмотренные выше условия предельной теоремы называются классическими. Главное из них — пренебрежимая малость слагаемых (дисперсий $\sigma_{n,k}$). Обсудим необходимые и достаточные условия без этого предположения.

Пусть $\{X_{n,k}\}$ и $\{Y_{n,k}\}$ — те же массивы, что и в доказательстве теоремы Линдеберга 13.1 и $\sum\limits_{k=1}^{m_n} \operatorname{var} X_{n,k} \to \sigma^2 \geqslant 0$.

Теорема 13.6. *(Золотарёв, Ротарь) Для описанных выше массивов случайных величин и во введённых ранее обозначениях*

$$S_n \xrightarrow{\mathcal{D}} Z \sim N(0, \sigma^2),\ n \to \infty \iff$$
$$\iff \forall \varepsilon > 0 \sum_{k=1}^{m_n} \int\limits_{\{|x| \geqslant \varepsilon\}} |x|\, |F_{n,k}(x) - \Phi_{n,k}(x)|\, dx \underset{n\to\infty}{\to} 0,$$

где $F_{n,k}(x)$ — функция распределения $X_{n,k}$, $\Phi_{n,k}(x)$ — функция распределения $Y_{n,k}$.

Доказательство данной теоремы сложно и объёмно, поэтому не приводится.

13.4 Метод Стейна

Метод Стейна на сегодняшний день является наиболее мощным методом доказательства не только предельных, но и многих других теорем.

Возьмём $h \in \mathrm{Lip}(1)$, то есть $|h(x) - h(y)| \leqslant |x - y|$.

Определение 13.1. *Уравнением Стейна* называется дифференциальное уравнение на функцию $f(x)$, $x \in \mathbb{R}$ следующего вида:

$$f'(x) - xf(x) = \underbrace{h(x) - \mathbb{E}\, h(W)}_{g(x)},$$

где $W \sim N(0, 1)$.

Решение уравнения Стейна — функция $f_h(x)$ — может быть найдена в явном виде и задаётся формулой

$$f_h(x) = e^{\frac{x^2}{2}} \int\limits_{-\infty}^{x} g(t) e^{-\frac{t^2}{2}}\, dt.$$

Действительно, продифференцируем эту функцию:

$$f_h'(x) = xe^{\frac{x^2}{2}} \int\limits_{-\infty}^{x} g(t) e^{-\frac{t^2}{2}}\, dt + e^{\frac{x^2}{2}} g(x) e^{-\frac{x^2}{2}} = xf_h(x) + g(x),$$

значит это решение.

Задача. Можно получить оценки для $\|f_h'(x)\|_\infty$ и $\|f_h''(x)\|_\infty$. В частности $\|f_h''(x)\|_\infty \leqslant 2$. Это не очень простая, но сугубо техническая работа.

Это нужно понимать на экзамене, но не обязательно уметь детально доказывать.

Само существование f_h следует из скорости роста h: эта функция растёт не быстрее линейной.

Теперь вместо x подставим в уравнение случайную величину X, например $X = \frac{S_n}{\sqrt{n}}$. Возьмём математическое ожидание от обеих частей и получим:

$$\mathbb{E}\left(f'\left(X\right) - Xf\left(X\right)\right) = \mathbb{E}\, h\left(X\right) - \mathbb{E}\, h\left(W\right)$$

Метод Стейна состоит в том, чтобы оценить разность функционалов от гауссовской случайной величины и случайной величины X с помощью оценки левой части в тождестве выше.

13.5 Доказательство ЦПТ методом Стейна

Далее рассматриваем в уравнении Стейна $h \in C^1\left(\mathbb{R}\right)$ и $\|h'\| \leqslant c < \infty$. Отсюда следует липшицевость h на $\mathbb{R}$.

Приведём нужную в дальнейшем лемму без доказательства.

Лемма 13.3. *Пусть $\|h'\| \leqslant c < \infty$ и f_h — соответвствующее решение уравнения Стейна. Тогда*

$$\|f_h\| \leqslant 2\|h'\|,\ \|f_h'\| \leqslant c_1\|h'\|,\ \|f_h''\| \leqslant c_2\|h'\|,$$

более того: $c_1 = \sqrt{\frac{2}{\pi}}$, $c_2 = 2$.

Для нас существенным является лишь ограниченность f_h и её производных некоторыми константами, а не их численные значения этих констант.

Теорема 13.7. *Пусть $X_1, X_2, \ldots$ – н.о.р. случайные величины,* $\forall k\ X_k \in L^3$, $\mathbb{E}\, X_k = 0, \operatorname{var} X_k = 1$. *Тогда* $\forall h\colon\ \|h'\| < \infty$

$$|\mathbb{E}\, h(T_n) - \mathbb{E}\, h(W)| \le \tilde{C}\frac{1 + \mathbb{E}\, X_1^3}{\sqrt{n}}, \qquad T_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}},\ W \sim N(0,1)$$

Доказательство. Из уравнения Стейна

$$|\mathbb{E}\, h(T_n) - \mathbb{E}\, h(W)| = |\mathbb{E}\left(f'(T_n) - T_n f(T_n)\right)|$$

Оценим правую часть.

$$T_n = \frac{X_1}{\sqrt{n}} + U_n, \qquad U_n := \frac{X_2 + \cdots + X_n}{\sqrt{n}}$$

и в силу независимости и одинаковой распределённости

$$\mathbb{E}(T_n f(T_n)) = \sqrt{n}\, \mathbb{E}(X_1 f(T_n))$$

$f \in C^2 \implies$ справедлива формула Тейлора:

$$f(T_n) = f(U_n) + f'(U_n)\frac{X_1}{\sqrt{n}} + f''(\xi_n)\frac{X_1^2}{2n},\ \xi_n \in [\min(U_n, T_n), \max(U_n, T_n)]$$

Ясно, что $\mathbb{E}\, f''(\xi_n)\frac{X_1^2}{n} < \infty$.

$$\sqrt{n}\left|\mathbb{E}\, X_1 f''(\xi_n)\frac{X_1^2}{2n}\right| \le \frac{\tilde{C}_1}{\sqrt{n}} \to 0$$

Из следствия леммы о группировке и н.о.р. (X_k)

$$\sqrt{n}\,\mathbb{E}\, X_1 f(U_n) = \sqrt{n}\,\mathbb{E}\, X_1\, \mathbb{E}\, f(U_n) = 0$$

Оценим оставшуюся часть:

$$|\mathbb{E}\, f'(T_n) - f'(U_n)| \le \mathbb{E}\left|\frac{f''(\xi_n)}{2}\frac{X_1^2}{n}\right| \to 0$$

Таким образом, $T_n \xrightarrow{\mathcal{D}} N(0,1)$ □

14 Многомерное нормальное распределение

Пусть $C \in \mathbb{R}^{n\times n}, C \ge 0, C = C^T, a \in \mathbb{R}^n$

Определение 14.1. Говорят, что случайный вектор X имеет нормальное распределение $N(a, C)$, если его характеристическая функция имеет вид

$$\varphi_X(t) = e^{i(a,t) - \frac{1}{2}(Ct,t)}$$

Аналогично с одномерным случаем,

$$\mathbb{E}\, X = a, \operatorname{Var} X = C$$

Теорема 14.1. *φ_X – х.ф.*

Доказательство. 1. $C > 0 \implies \det C \neq 0 \implies \exists A = C^{-1}$

$$f(x) := \frac{\sqrt{\det A}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(At,t)}$$

Покажем, что f – плотность.

Замена переменных: $x - a =: Bu, t =: Bv, B^T CB = diag(d_j) =: D$ То есть $B \in \mathcal{O}(R^n)$

$$i(t, x-a) - \frac{1}{2}(A(x-a), x-a) = i(u,v) - \frac{1}{2}(D^{-1}u, u)$$

Формула обращения + теорема Фубини:

$$\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det D}} \int_{\mathbb{R}^n} e^{i(v,u) - \frac{1}{2}(D^{-1}u,u)} |\det B| du =$$
$$= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi d_k}} \int_{\mathbb{R}} e^{iv_k u_k - \frac{u_k^2}{2d_k}} du_k = \prod_{k=1}^{n} e^{-\frac{v_k^2 d_k}{2}} = e^{-\frac{1}{2}(Ct,t)}$$

Замечание 14.1. $\varphi_{X_k}(v_k) = e^{-\frac{v_k^2 d_k}{2}} \implies X_k \sim N(0, d_k)$

2. $C \geq 0$. Введем $C_k := C + \frac{1}{k}\mathcal{I}$. Тогда $\forall k \; C_k > 0 \implies$ справедливы выше приведенные рассуждения и по теореме непрерывности предельная функция при $k \to \infty$ тоже будет характеристической.

$\square$

Следствие 14.1. *Пусть* $X \sim N(a, C)$. *Тогда*

$$X_1, \ldots, X_n - \textit{независимы} \iff C = diag$$

Теорема 14.2. *(ЦПТ) Пусть* $(X_{n,k})$ *– схема серий,* $\mathbb{E}\, X_{n,k} = 0$, $\sum_{k=1}^{m_n} \operatorname{Var} X_{n,k} \to C$ *и выполняется условие Линдеберга:*

$$\forall \varepsilon > 0 \sum_{k=1}^{m_n} \mathbb{E}\, \|X_{n,k}\|^2 \mathbb{1}(\|X_{n,k}\| > \varepsilon) \to 0$$

Тогда

$$\sum_{k=1}^{m_n} X_{n,k} \xrightarrow{\mathcal{D}} N(0, C)$$