

ТЕОРЕМ  
МАТСТАТИСТИКА  
и тщетные попытки понять что-нибудь из  
вышеперечисленного

Лектор: Р. С. Пусев  
Записал :taxus

18 января 2017 г.

# Оглавление

<b>1</b>	<b>Элементарная теория вероятностей</b>	<b>2</b>
§ 1	Аксиоматическое определение вероятности . . . . .	2
§ 2	Формула полной вероятности . . . . .	3
§ 3	Теорема Байеса . . . . .	3
§ 4	Независимые события . . . . .	4
§ 5	Случайные величины и их распределения . . . . .	4
§ 6	Моменты случайных величин . . . . .	5
§ 7	Характеристическая функция . . . . .	6
§ 8	Теорема Муавра-Лапласа . . . . .	7
<b>2</b>	<b>Матстатистика</b>	<b>7</b>
§ 9	Случайные векторы . . . . .	7
§ 10	Функция от случайного вектора . . . . .	8
§ 11	Матожидание и дисперсия суммы случайных величин . . . . .	9
§ 12	Матожидание функции случайной величины . . . . .	11
§ 13	Неравенство Шварца . . . . .	11
§ 14	Характеристическая функция суммы случайных величин . . . . .	11
§ 15	Суммирование большого числа случайных величин . . . . .	12
§ 16	Центральная предельная теорема . . . . .	13
§ 17	Обобщённая теорема Муавра-Лапласа . . . . .	13
§ 18	Метод моментов . . . . .	14
§ 19	Метод максимального правдоподобия . . . . .	15
§ 20	Лемма Фишера . . . . .	17
§ 21	Доверительны интервалы нормального распределения . . . . .	19
§ 22	Проверка гипотез по параметрам нормального распределения . . . . .	20
§ 23	Линейная регрессия . . . . .	21
§ 24	Теорема Гаусса-Маркова . . . . .	22
§ 25	Оценка дисперсии погрешностей . . . . .	22
§ 26	Критерий согласия Пирсона . . . . .	23
§ 27	Непараметрические критерии . . . . .	23
<b>3</b>	<b>Случайные процессы</b>	<b>24</b>
§ 28	Процессы с независимыми приращениями . . . . .	24
§ 29	Стационарные процессы . . . . .	25
§ 30	Цепи Маркова . . . . .	25
§ 31	Марковские процессы . . . . .	27
<b>A</b>	<b>Обозначения</b>	<b>28</b>
<b>B</b>	<b>Стандартные распределения</b>	<b>28</b>
	<b>Литература</b>	<b>28</b>

# Глава 1: Элементарная теория вероятностей

## § 1 Аксиоматическое определение вероятности

**Определение 1** ( $\sigma$ -алгебра). Алгеброй  $\mathcal{A}$  подмножеств множества  $\Omega$  называется такой набор его подмножеств с заданными операциями объединения, пересечения и дополнения множеств, что

1.  $\Omega \in \mathcal{A}$
2.  $X \in \mathcal{A} \Rightarrow \bar{X} \in \mathcal{A}$
3.  $X_1, X_2 \in \mathcal{A} \Rightarrow X_1 \cup X_2 \in \mathcal{A}$

Сигма-алгеброй подмножеств называется всё тоже самое, только можно объединять счётное число подмножеств.

**Определение 2** (Вероятностное пространство). Рассмотрим упорядоченную тройку  $(\Omega, \mathcal{F}, P)$ , где

$\Omega$  — Множество (элементарных исходов). Чисел, например.

$\mathcal{F}$  —  $\sigma$  — алгебра подмножеств  $\Omega$

$P$  — Собственно, вероятность

**Определение 3** (Вероятность).  $P: \mathcal{F} \rightarrow \mathbb{R}$  такая, что

1.  $\forall A \in \mathcal{F} \quad P(A) \geq 0$
2.  $\forall \{A_i\}: A_i \cap A_j = \emptyset \quad P(\bigcup_i A_i) = \sum_i P(A_i)$
3.  $P(\Omega) = 1$

Как видно, сильно похоже на площадь. Что впрочем неслучайно, вероятность — нормированная мера. Последнее условие как раз и означает нормированность.

**Определение 4** (Тривиальные события).  $\emptyset, \Omega$ .

**Утверждение 1.** *Свойства вероятности:*

1.  $P(\emptyset) = 0$
2.  $0 \leq P(A) \leq 1$
3.  $P(\bar{A}) = 1 - P(A)$
4.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

А это очень важное и будет в отдельном утверждении.

**Утверждение 2** (Непрерывность меры). Пусть  $A_1 \subset \dots \subset A_n \subset \dots$ ,  $\bigcup_i A_i = A$ . Тогда  $P(A) = \lim_{n \rightarrow \infty} P(A_n)$ .



Пусть

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ &\dots\dots\dots \end{aligned}$$

Тогда

$$A = \bigcup_{i=1}^{\infty} B_i; \quad A_n = \bigcup_{i=1}^n B_i$$

Следовательно,

$$P(A) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \lim_{n \rightarrow \infty} P(A_n)$$

Всё работает, потому что в  $\sigma$ -алгебре можно объединять счётное число множеств.



## § 2 Формула полной вероятности

**Определение 1** (Условная вероятность). Пусть  $A, B \in \mathcal{F}$ ,  $P(A) > 0$ <sup>1</sup>. Тогда

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

**Утверждение 1.** Пусть  $\{A_i, H\} \subset \mathcal{F}$  и

1.  $P(A_i) > 0$
2.  $A_i \cap A_j = \emptyset$
3.  $\bigcup_i A_i = \Omega$

(полная группа/система событий). Тогда

$$P(H) = \sum_i P(A_i) \cdot P(H \mid A_i)$$

## § 3 Теорема Байеса

**Теорема 1.** Пусть  $A_i$  — полная система событий,  $H \in \mathcal{F}$ :  $P(H) > 0$ . Тогда

$$P(A_k \mid H) = \frac{P(A_k) \cdot P(H \mid A_k)}{\sum_i P(A_i) \cdot P(H \mid A_i)}$$

---

<sup>1</sup>ну мы же делим, надо убедиться что неноль

## § 4 Независимые события

**Определение 1.** Пусть  $A, B \in \mathcal{F}$ . Они называются независимыми, если

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

**Утверждение 1.** События  $A, B$  независимы  $\Leftrightarrow P(A | B) = P(A) \vee P(B | A) = P(B)$  (в зависимости от того, что определено, вдруг там ноль где-нибудь).

**Утверждение 2.** Если  $A, B$  несовместны, то нетривиальные  $A, B$  — зависимы.

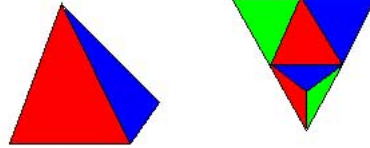
**Определение 2.** Случайные величины  $\{X_i\}$  попарно независимы, если

$$\forall i, j \quad P(X_i \cap X_j) = P(X_i) \cdot P(X_j)$$

**Определение 3.** Случайные величины  $\{X_i\}_{i=1}^n$  независимы по совокупности, если

$$\forall \{i_k \mid i_k, k \in (\mathbb{Z} \cap [1; n])\} \quad P\left(\bigcap_{i_k} X_{i_k}\right) = \prod_{i_k} P(X_{i_k})$$

*Замечание 1.* Определения 1.4.3 и 1.4.3 правда разные. Конечно попарная независимость следует из независимости по совокупности, но обратное неверно.



**Пример 1.** Тетраэдр Бернштейна: . Здесь вероятность выпадения всех 3 цветов —  $1/4$ , а через попарные —  $1/8$

## § 5 Случайные величины и их распределения

**Определение 1** (Случайная величина). Случайной величиной назовём произвольное хорошее отображение  $X: \Omega \rightarrow \mathbb{R}$ .

Тут нужно бы сказать про измеримость, это потребуется, чтобы говорить о вероятности попадания в интервал на прямой. Так что

$$X: (X^{-1}(B) = \{\omega \mid X(\omega) \in B\}) \in \mathcal{F},$$

где  $B$  — борелевское множество

**Определение 2.** Пусть  $B \subset \mathbb{R}$ ,  $B$  — промежуток, или дополнение в нему (борелевское множество).<sup>1</sup>

$$P(X \in B) = P(\{\omega \mid X(\omega) \in B\})$$

**Определение 3.** Случайная величина называется дискретной, если

$$\exists (\{a_i\} \sim \mathbb{N}): \left( \sum_i P(X = a_i) = 1 \right)$$

то есть

$$P(X \in B) = \sum_{\{i \mid a_i \in B\}} p_i, \quad p_i = P(X = a_i)$$

<sup>1</sup>тут вроде концы могут входить, так что точка — тоже борелевское множество.

**Определение 4.** Случайная величина называется непрерывной, если

$$\exists (f_X: B \rightarrow \mathbb{R}): \left( P(X \in B) = \int_B f_X(x) dx \right)$$

**Определение 5** (Распределение случайной величины).  $F(B) = P(X \in B)$

**Пример 1** (К непрерывному распределению). Пусть  $X(\omega) = \omega$ ,  $B = (0, 1)$ ,  $\Omega = (-1; 1)$ . Выберем  $f_X \equiv \frac{1}{2}$ .

$$F(B) = P(X \in B) = P(\{\omega \mid \omega \in (0, 1)\}) = P((0, 1)) = \frac{1-0}{1+1} = \frac{1}{2}$$

$$F(B) = \int_{(0,1)} \frac{1}{2} dx = \frac{1-0}{2} = \frac{1}{2}$$

Это всё верно, потому что на множестве интервалов вероятность — нормированная длина интервала.

**Определение 6** (Функция распределения).

$$F_X: \mathbb{R} \rightarrow [0; 1]: F_X(x) = P(X < x) = P(\{\omega \mid X(\omega) < x\})$$

**Утверждение 1.** Про  $F(x)$  верно следующее:

1.  $F \uparrow \mathbb{R}$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$
3.  $\lim_{x \rightarrow +\infty} F(x) = 1$
4.  $\lim_{x \rightarrow x_0 - 0} F(x) = F(x_0)$

**Утверждение 2.** Верно и обратное: если существует функция с указанными свойствами, под неё найдётся случайная величина.

*Замечание.* Если рассматривать обобщённые функции, то любое распределение запишется как

$$\int_{-\infty}^x f_X(u) du$$

## § 6 Моменты случайных величин

**Определение 1.** Пусть  $X$  — случайная величина,

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$$

Тогда

$$\langle x \rangle \equiv \bar{x} \equiv M X = \int_{-\infty}^{\infty} x f_X(x) dx$$

**Утверждение 1.** Свойства матожидания:

1.  $M < \infty$

2.  $M(aX + bY) = aMX + bMY$

3.  $P(X \geq 0) = 1 \Rightarrow MX \geq 0$

4. 
$$\begin{cases} P(X \geq 0) = 1 \\ MX = 0 \end{cases} \Rightarrow P(X = 0) = 1$$

5. если  $X, Y$  — независимы, то  $M(XY) = MX \cdot MY$

**Определение 2.** Момент  $k$ -ого порядка относительно начала  $a$ :

$$\lambda_{k,a} = \int_{-\infty}^{\infty} (x - a)^k f(x) dx$$

(если есть абсолютная сходимость)

**Определение 3.** Начальный момент:  $\nu_k = \lambda_{k,0}$

**Определение 4.** Центральный момент:  $\mu_k = \lambda_{k,\bar{x}}$

**Утверждение 2.** 
$$\nu_k = \sum_{i=0}^k C_k^i a^i \lambda_{k-i,a}$$

**Определение 5** (Дисперсия).  $DX = M(X - MX)^2$ ,  $\sigma = \sqrt{DX}$  — среднее квадратичное отклонение.

**Утверждение 3.**

$$D(aX + bY) = a^2 DX + b^2 DY$$

если  $X, Y$  — независимы, то  $D(XY) = DX \cdot DY$

$$D(X + C) = DX$$

## § 7 Характеристическая функция

**Определение 1** (Характеристическая функция).  $\Phi(t) = Me^{itx}$

**Утверждение 1.** Свойства характеристической функции:

1. Всегда существует и  $|\Phi(t)| \leq 1$ .

2. 
$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Phi(t) dt$$

3.  $\Phi_{a+Xb}(t) = e^{ita} \Phi_X(tb)$

4. Если  $X, Y$  — независимы, то  $\Phi_{X+Y}(t) = \Phi_X(t) \Phi_Y(t)$

5. Если  $M|X|^n < \infty$ , то  $\Phi^{(k)}(0) = i^k MX^k$

**Определение 2** (Сходимость по распределению).  $X_n \xrightarrow{d} X \Leftrightarrow F_n(x) \rightarrow F(x)$

**Теорема 2** (О непрерывном соответствии).  $X_n \xrightarrow{d} X \Leftrightarrow \Phi_{X_n}(t) \xrightarrow{t} \Phi_X(t)$ . Здесь на самом деле всё сложно. В разных книжках пишут равномерную сходимость на разных интервалах. У нас вроде было на всей вещественной прямой, но тогда это слишком сильное утверждение. К слову так же в [1] и [2]. А вот в [3] требуют только сходимости в каждом конечном интервале. Можно взять экспоненту, и понять, что эти условия разные.

## Глава 2: Матстатистика

### § 9 Случайные векторы

**Определение 1** (Случайный вектор). Случайным вектором назовём произвольное хорошее отображение  $X: \Omega \rightarrow \mathbb{R}^n$ . См. примечание про измеримость в § 5. Борелевские множества можно рассматривать и в  $\mathbb{R}^n$ , как наименьшую сигма-алгебру, содержащую все полуоткрытые параллелепипеды.

Так же можно считать, что случайный вектор — набор случайных величин.

**Определение 2** (Функция распределения).

$$F_X: \mathbb{R}^n \rightarrow [0; 1]: F_X(x) = P(X^1 < x^1, \dots, X^n < x^n)$$

То есть вероятность попадания в параллелепипед, уходящий в бесконечность.

*Замечание 1.* Тут как раз используется, что борелевские множества «прямоугольные».

**Утверждение 1.** Про  $F(x)$  верно следующее:

1.  $F$  не убывает по каждому аргументу.
2.  $\lim_{x_i \rightarrow -\infty} F(x) = 0$
3.  $\lim_{x \rightarrow +\infty} F(x) = 1$
4.  $\lim_{x \rightarrow x_0 - 0} F(x) = F(x_0)$  (по совокупности переменных). Это просто следует из непрерывности меры.

тоже важно, так что отдельно

**Утверждение 2.** Пусть  $a^1 < b^1, \dots, a^n < b^n$ , тогда работает формула включений и исключений

$$F(b^1, \dots, b^n) - \sum_i F(b^1, \dots, a^i, \dots, b^n) + \dots + F(a^1, \dots, a^n) = P(x \in [a^1, b^1) \times \dots \times [a^n, b^n))$$

По сути следствие формулки про вероятность объединения.

**Определение 3.** Векторная случайная величина называется дискретной, если

$$\exists (\{a_i \mid a_i \in \mathbb{R}^n\} \sim \mathbb{N}): \left( \sum_i P(X = a_i) = 1 \right)$$

то есть

$$P(X \in B) = \sum_{\{i \mid a_i \in B\}} p_i, \quad p_i = P(X = a_i)$$



**Определение 4.** Векторная случайная величина называется непрерывной, если

$$\exists (f_X: B \rightarrow \mathbb{R}): \left( P(X \in B) = \int_B f_X(x^1, \dots, x^n) dx^1 \cdots dx^n \right)$$

*Замечание 1.* Для функций распределения:

$$F(x^1, \dots, x^n) = \int_{-\infty}^{x^n} \cdots \int_{-\infty}^{x^1} f_X(x^1, \dots, x^n) dx^1 \cdots dx^n$$

**Утверждение 3.** Пусть  $X, Y$  — независимы. Тогда  $p_{X+Y}(x, y) = p_X(x) \cdot p_Y(y)$

▼

По определению функции распределения

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy$$

Из независимости  $X, Y$

$$\begin{aligned} F_{X,Y}(x, y) &= P(X < x, Y < y) = P(\omega \mid X(\omega) \in (-\infty; x] \cap \omega \mid Y(\omega) \in (-\infty; y]) \\ &= P(\omega \mid X(\omega) \in (-\infty; x]) \cdot P(\omega \mid Y(\omega) \in (-\infty; y]) \\ &= P(X < x) \cdot P(Y < y) = F_X(x) \cdot F_Y(y) \end{aligned}$$

А тогда из независимости подынтегральных функций

$$F_X(x) \cdot F_Y(y) = \int_{-\infty}^x p_X(x) dx \cdot \int_{-\infty}^y p_Y(y) dy = \int_{-\infty}^x \int_{-\infty}^y p_X(x) \cdot p_Y(y) dx dy$$

А дальше можно заметить, что нам неважно по какому множеству интегрировать.

$$\int_B (p(x, y) - p_X(x) \cdot p_Y(y)) dx dy = 0$$

Здесь правда всё ломается на отсутствии непрерывности у  $p$ . Но если она есть, то дальше стандартное рассуждение в окрестности точки где не 0.

▲

## § 10 Функция от случайного вектора

**Определение 1** (Функция от случайного вектора).  $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

Здесь нужно снова говорить про измеримость  $g$  — прообраз борелевского множества должен быть борелевским множеством. Иначе  $g(X)$  может не получиться случайной величиной. Но всякие мерзкие отображения всё равно никому не нужны ☺.

**Утверждение 1.** Пусть  $X$  — дискретная случайная величина,  $f$  — обратима,  $Y = g(X)$ ,  $b_j = f(a_j)$ . Тогда  $P(Y^i = b_j^i) = P(X^i = a_j^i)$ .

▼

$$P(Y^i = b_j^i) = P(f(X^i) = f(a_j^i)) = P(\omega \mid f(X^i(\omega)) = f(a_j^i))$$

Поскольку  $f$  — обратима, она биективна. Значит  $f(X) = f(a_j) \Leftrightarrow X = a_j$ . Собственно, всё.

▲

**Утверждение 2.** Пусть  $X$  — непрерывная случайная величина,  $f$  — обратима,  $Y = g(X)$ ,  $f^{-1} = g$ . Тогда  $p_Y(y) = p_X(g(y)) \left| \frac{dg}{dy} \right|$ .

▼

Пусть  $D = f(B)$ . Тогда  $P(Y \in D) = P(X \in B)$  опять-таки в силу биективности  $f$ . Ну, ничего нового туда попасть не может и у всего есть прообраз. Так что (здесь будем рисовать один значок интеграла и дифференциала из экономии размера пдф-ки, хотя в этом замечании данных может и больше)

$$\int_D p_Y(y) dy = \int_B p_X(x) dx = \int_D p_X(g(y)) \left| \frac{dg}{dy} \right|$$

Якобиан тут под модулем, так как множество неориентированное. Я верю, что нам ещё про это расскажут на матане.

▲

## § 11 Матожидание и дисперсия суммы случайных величин

**Утверждение 1.** Пусть  $X, Y, n \in (A \sim \mathbb{N})$  — две дискретные независимые случайные величины. Тогда

$$P(X + Y = n) = \sum_{k \in A} P(X = n - k) \cdot P(Y = k)$$

▼

Из формулы полной вероятности ( $Y = k, k \in A$  правда полная группа)

$$P(X+Y = n) = \sum_{k \in A} P(X+Y = n \mid Y = k) \cdot P(Y = k) = \sum_{k \in A} P(X = n-k \mid Y = k) \cdot P(Y = k)$$

А вот тут уже поможет независимость  $X, Y$ .

$$\dots = P(X = n - k) \cdot P(Y = k)$$

▲

**Утверждение 2.** Пусть  $X_1, X_2, n \in \mathbb{R}$  — две непрерывные независимые случайные величины. Тогда

$$p_{X_1+X_2}(y) = \int_{\mathbb{R}} p_1(y-t)p_2(t) dt$$



Пусть  $Y = X_1 + X_2$ . Тут видно, что нет биекции, придется руками что-то делать.

$$F_Y(y) = \iint_{x_1+x_2 < y} p(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{y-x_1} p(x_1, x_2) dx_2 = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^y p(x_1, u-x_1) du$$

Переменные независимы, так что можно поменять местами интегралы (ещё же по области интегрируем, неважно как)<sup>1</sup> А тогда, убирая внешний интеграл из определения функции распределения, получаем

$$p_Y(y) = \int_{-\infty}^{\infty} p(t, y-t) dt$$

Поскольку  $X_1, X_2$  независимы, то  $p(t, y-t) = p_1(t) \cdot p_2(y-t)$ . Нумерация никого не интересует, так что

$$p_Y(y) = \int_{-\infty}^{\infty} p_1(y-t) \cdot p_2(t) dt$$



Теперь можно перейти и к содержанию билета

**Утверждение 3.**  $M(\sum_i X_i) = \sum_i M X_i$ . Да и вообще оно линейно.



Пусть  $f(X, Y) = X + Y$

$$\begin{aligned} M(X+Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) p(x, y) dx dy = \int_{-\infty}^{+\infty} x p(x, y) dx dy + \int_{-\infty}^{+\infty} y p(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} x \left( \int_{-\infty}^{+\infty} p(x, y) dy \right) dx + \int_{-\infty}^{+\infty} y \left( \int_{-\infty}^{+\infty} p(x, y) dx \right) dy \\ &= M X + M Y \end{aligned}$$

Покажем, что  $\int_{-\infty}^{+\infty} p(x, y) dy = p_X(x)$

$$\begin{aligned} \int_{-\infty}^x \left( \int_{-\infty}^{+\infty} p(x, y) dx \right) dy &= F(x, +\infty) = P(X < x, Y < +\infty) \\ &= P(\omega \mid X(\omega) \in (-\infty, x], Y \in \mathbb{R} \cup \{+\infty\}) \\ &= P(X < x) = F_X(x) = \int_{-\infty}^x p_X(x) dx \end{aligned}$$

Опять-таки интервал можно сжать как угодно, правда снова проблемы с непрерывностью.

Часть про константу слишком очевидна, не будем её доказывать.




---

<sup>1</sup>слишком много раз пользовались на физике, так что оставим на 4 семестр

**Утверждение 4** (Дисперсия суммы).  $D(\sum_i X_i) = \sum_i D X_i$

▼

Сначала заметим, что  $D X = M(X - M X)^2$ ,  $M(X - M X) = M X - M X = 0$

$$\begin{aligned} D(X + Y) &= M(X + Y - M(X + Y))^2 = M((X - M X) + (Y - M Y))^2 \\ &= M(X - M X)^2 + M(Y - M Y)^2 + 2 M(X - M X) M(Y - M Y) \\ &= D X + D Y \end{aligned}$$

▲

**Утверждение 5.** Если  $X, Y$  — независимы, то  $M X Y = M X M Y$

## § 12 Матожидание функции случайной величины

**Определение 1** ( $\ddot{\sim}$ ). Пусть  $f(X)$  — функция от случайной величины. Тогда

$$M f(X) = \int_{-\infty}^{+\infty} f(x) p(x) dx.$$

В случае чего он многомерный, просто прикидывается. Существует, если есть абсолютная сходимость.

*Замечание.* я ещё подумаю, может это всё же утверждение.

**Утверждение 1.** Матожидание функции линейно

**Утверждение 2.** Если  $X, Y$  — независимы, то  $M f_1(X_1) f_2(X_2) = M f_1(X_1) M f_2(X_2)$

## § 13 Неравенство Шварца

**Утверждение 1.**  $(M X Y)^2 \leq M X^2 M Y^2$

▼

$M(X + tY)^2 = t^2 M Y^2 + 2t M X Y + M X^2 \geq 0$  из свойств матожидания. Ну там и подынтегральная функция положительна. Тогда квадратное уравнение в правой части может иметь не более одного корня.

$$(2 M X Y)^2 - 4 M X^2 M Y^2 \leq 0 \Leftrightarrow (M X Y)^2 \leq M X^2 M Y^2$$

▲

## § 14 Характеристическая функция суммы случайных величин

**Утверждение 1.** Пусть  $X, Y$  — независимые случайные величины. Тогда

$$\Phi_{X+Y}(t) = \Phi_X(t) \cdot \Phi_Y(t)$$

▼

Из 2.14.1

$$\Phi_{X+Y}(t) = M e^{itX} e^{itY} = M e^{itX} \cdot M e^{itY} = \Phi_X(t) \cdot \Phi_Y(t)$$

▲

**Следствие 1.** Если все величины одинаково распределены, то  $\Phi_{X_1+\dots+X_n}(t) = (\Phi(t))^n$ ,

$$p_{X_1+\dots+X_n} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\Phi(t))^n dt$$

## § 15 Суммирование большого числа случайных величин

:set aflameX~

**Теорема 1** (ЦПТ Линдберга-Леви-Агекияна ~). Пусть  $X_1, \dots, X_n$  — независимые одинаково распределённые случайные величины. Пусть к тому же  $S_n = X_1, \dots, X_n$ ,  $0 < D X_k < \infty$ . Пусть  $M X_k = a$ ,  $D X_k = \sigma$ . Тогда при  $n \rightarrow \infty$   $Z_n \sim N(0, 1)$ , в вариации из Агекияна  $S_n \sim N(na, n\sigma^2)$

□ Пусть  $M X_k = a$ ,  $D X_k = \sigma^2$ . Рассмотрим характеристическую функцию  $\Phi(t) = M e^{itX_k}$ . Введём замену (которая z-преобразование.):

$$z_n = \frac{S_n - an}{\sigma\sqrt{n}}$$

Давайте ещё немного схитрим и положим  $X_k \leftarrow X_k - a$ . А то потом будет много возни с бедным  $a$ . При этом  $z_n = \frac{S_n}{\sigma\sqrt{n}}$  Тогда

$$\Phi_{z_n}(t) = M \left( e^{\frac{itS_n}{\sigma\sqrt{n}}} \right) = \left( \Phi \left( \frac{t}{\sigma\sqrt{n}} \right) \right)^n$$

А характеристическая функция дифференцируема дважды из существования дисперсии.

$$\Phi'(0) = 0$$

$$\Phi''(0) = -\sigma^2$$

$$\Phi \left( \frac{t}{\sigma\sqrt{n}} \right) = \Phi(0) + \Phi'(0) \frac{t}{\sigma\sqrt{n}} + \Phi''(0) \frac{t^2}{2\sigma^2 n} + o \left( \frac{1}{n} \right) = 1 - \frac{1}{2} \frac{t^2}{n} + o \left( \frac{1}{n} \right)$$

А при  $n \rightarrow \infty$

$$\Phi \left( \frac{t}{\sigma\sqrt{n}} \right)^n = \left( 1 - \frac{1}{2} \frac{t^2}{n} + o \left( \frac{1}{n} \right) \right)^n \Rightarrow e^{-t^2/2}$$

Здесь сходимость есть на любом конечном интервале, но вот про всю прямую этого уже не скажешь. Так что снова поднимается вопрос какой теоремой о непрерывном соответствии пользоваться. Но если ей воспользоваться (тут потихому применили обратное преобразование Фурье), то

$$p_{z_n} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \frac{-s^2 + s^2 - 2its - t^2}{2} = \frac{e^{-s^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left( - \left( \underbrace{\frac{t+is}{\sqrt{2}}}_{\eta} \right)^2 \right) d\eta = \frac{e^{-s^2/2}}{\sqrt{2\pi}}$$

$$F_{Z_n}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

Дальше — вариация из Агекияна. Используя утверждение 2.10.2 про замену переменной как раз получаем нормальное распределение. Только тут нужно поменять в процессе  $\sigma$

$$Z_n = \frac{S_n}{\sigma\sqrt{n}}$$

$$F_{S_n}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{n}} e^{-\frac{u^2}{2\sigma^2 n}} du$$

Вернёмся обратно к ненулевому  $a$

$$S_n \leftarrow S_n - na$$

$$F_{S_n} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{n}} \int_{-\infty}^x e^{-\frac{(u-na)^2}{2\sigma^2 n}} du$$

$$S_n \sim N(na, n\sigma^2)(n \rightarrow \infty)$$

■

## § 16 Центральная предельная теорема

**Теорема 1** (ЦПТ Ляпунова). Пусть  $\{X_k\}$  — независимые случайные величины (тут нет одинаковости распределений!). Введём горю обозначений:

$$S_n = \sum_i x_i$$

$$a_k = M X_k \quad \sigma_k^2 = D x_k \quad \gamma_k = M |X_k - a_k|^3$$

$$A_n = \sum_{k=1}^n a_k \quad B_n^2 = \sum_{k=1}^n \sigma_k^2 \quad C_n = \sum_{k=1}^n \gamma_k$$

Тогда

$$\frac{C_N}{B_n^3} \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \frac{S_n - A_n}{B_n} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)$$

*Замечание.* Тут какая-то жесть. Она мало где формулируется и нигде не доказывается. Что-то есть тут: [3], а здесь [1] так другую теорему обозвали

:set aflame

## § 17 Обобщённая теорема Муавра-Лапласа

**Определение 1.** Пусть  $X_1, \dots, X_n \sim \mathcal{N}(0, 1)$  и независимы. Тогда говорят, что случайная величина  $\chi_n^2 = \sum_{k=1}^n X_k^2$  имеет распределение  $\chi^2$  с  $n$  степенями свободы.

**Утверждение 1.**  $p_{\chi^2}(z) = \frac{1}{2^{n/2} \cdot \Gamma(\frac{n}{2})} z^{n/2-1} e^{-z/2}$

▼

Характеристическая функция  $\chi$  может быть найдена из 1 Найдём сначала характеристическую функцию  $X_k^2$ . Для этого было бы недурно найти плотность соответствующего распределения

$$P(y < X^2 < y + dy) = P(\sqrt{y} < X < \sqrt{y + dy}) + P(-\sqrt{y} > X > -\sqrt{y + dy})$$

$$= \frac{2}{\sqrt{2\pi}} \int_{\sqrt{y}}^{\sqrt{y+dy}} e^{-u^2/2} du = \sqrt{\frac{2}{\pi}} \frac{e^{-y/2}}{2\sqrt{y}}$$

А теперь можно и фурье-образ найти

$$\int_{-\infty}^{+\infty} \frac{e^{-y/2}}{2\sqrt{y}} dy = \int_0^{+\infty} \frac{e^{-y/2}}{2\sqrt{y}} dy = \int_0^{+\infty} \exp\left(\frac{-\eta^2(1-2it)}{2}\right) d\eta = (1-2it)^{-1/2} \sqrt{\frac{\pi}{2}}$$

Вспоминая про коэффициент получим  $\Phi_k(t) = (1 - 2it)^{-1/2}$ .

$$\Phi_\chi(t) = (\Phi_k(t))^n = (1 - 2it)^{-n/2}$$

Тогда

$$p_\chi(z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (1 - 2it)^{-n/2} e^{-itz} dt$$

Дальше немного жёсть

$$\begin{aligned} \int_{-\infty}^{+\infty} (1 - 2it)^{-n/2} e^{-itz} dt &= \int_{-\infty}^{+\infty} e^{-l} \left(1 - 2 \frac{l}{z}\right)^{-n/2} \frac{1}{iz} dl = 2 \cdot \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}} \int_0^\infty s^{-n/2} e^{-s} ds \\ &= 2 \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}} \cdot \Gamma\left(1 - \frac{n}{2}\right) \end{aligned}$$

Из правила отражения для  $\Gamma$ -функции  $\Gamma(1 - n/2) \Gamma(n/2) = \frac{\pi}{\sin \frac{\pi n}{2}}$ . А это почти что надо. ✂там надо интеграл поаккуратнее брать.

▲

**Теорема 2** (Обобщённая теорема Муавра-Лапласа). Пусть  $X_1, \dots, X_n$  — независимые случайные величины с дискретным распределением:

$$X_k: \frac{1}{p_1} \mid \dots \mid \frac{r}{p_r}$$

Рассмотрим  $\nu_k = \#\{1 \leq i \leq n \mid X_i = k\}$ ,  $1 \leq k \leq r$ . Тогда

$$\sum_{k=1}^r \left( \frac{\nu_k - np_k}{\sqrt{np_k}} \right) \xrightarrow{d} \chi_{r-1}^2$$

## § 18 Метод моментов

Здесь походу нужно все статистические определения в одном параграфе :set aflame

**Вводные слова** В отличие от теорвера, матстатистике неизвестно распределение случайной величины. И нужно придумать, как его восстановить по конкретной реализации. Пусть  $X$  — та самая величина, про распределение которой очень хочется узнать

Главная героиня матстатистики — выборка

**Определение 1.** Выборка объёма  $n$  —

1.  $n$  независимых случайных величин, распределённых так же, как и  $X$
2. набор чисел  $X_i(\omega)$ ,  $\omega \in \Omega$  — какой-то исход. Ещё называется реализацией выборки.

Собственно, первое определение — это до испытания, а второе — уже после.

**Основные задачи** Пусть  $X_1, \dots, X_n \sim F(x, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}^d$  — множество параметров.

1. Оценивание параметров:

- Точечные оценки:  $\hat{\theta} = T(X_1, \dots, X_n)$
- Доверительные интервалы:  $P_\alpha(T_1 < \theta < T_2) = \alpha$

2. Проверка гипотез

Пусть  $\Theta = \Theta_0 \cup \Theta_1$ . А мы хотим узнать чему принадлежит  $\theta$ .

$H_0$ :  $\theta \in \Theta_0$  — основная гипотеза

$H_1$ :  $\theta \in \Theta_1$  — альтернативная гипотеза

**Выборочные характеристики** Выберем реализацию случайной величины (выборку во втором смысле)  $X_1, \dots, X_n$ . Рассмотрим новую случайную величину:

$$\tilde{X}: \begin{array}{c|c|c} X_1 & \cdots & X_n \\ \hline \frac{1}{n} & \cdots & \frac{1}{n} \end{array}$$

Ну а дальше все выборочные характеристики определяются уже для этой случайной величины. Напомним следующее

**Определение 2** (Индикатор).  $I_A(X) = \begin{cases} 1, & X \in A \\ 0, & X \notin A \end{cases}, I(X < x) = \begin{cases} 1, & X < x \\ 0, & X \geq x \end{cases}$

**Определение 3.** Если  $X_1, \dots, X_n$  можно упорядочить, то  $X_{(1)} \leq \dots \leq X_{(n)}$  называется вариационным рядом.

## ✂Свойства оценок

### Метод моментов

**Определение 4.** Пусть  $F(x, \theta)$  — семейство распределений,  $m(x) = M g(x)$  — какой-то момент этого распределения. Пусть известно, что  $h(\theta) = m(x)$ . Тогда собственно сам метод состоит в том, чтобы оценить  $\theta$  как решение уравнения выше.

$$\hat{\theta} = h^{-1}(m(x))$$

В случае чего там векторы, но особо не страшно.

**Пример 1.** <+примеры про непрерывные распределения+>

## § 19 Метод максимального правдоподобия

**Определение 1.** За  $p(x, \theta)$  обозначим плотность функции распределения  $F(x, \theta)$  в точке  $x$  в случае непрерывного распределения и  $P(X = x)$  в случае дискретного.

**Определение 2.** Пусть  $\{X_k\}$  —  $n$  независимых случайных величин. Тогда  $L(\theta) := \prod_{k=1}^n f_\theta(X_k)$  — функция правдоподобия. Ещё берут её логарифм.



Генеральная совокупность		Выборка	
Матожидание	$M X$	$\bar{X} = \frac{1}{n} \sum_k X_k$	Выборочное среднее
Дисперсия	$D X$	$S^2 = \frac{1}{n} \sum_k (X_k - \bar{X})^2$	Выборочная дисперсия
Момент порядка $l$	$M X^l$	$m_l = \frac{1}{n} \sum_k X_k^l$	
Ковариация	$M(X - M X)(Y - M Y)$	$\frac{1}{n} \sum_k (X_k - \bar{X})(Y_k - \bar{Y})$	
Ассиметрия ( $\gamma_3$ )	$M(X - M X)^3 / \sigma^3$	$\frac{1}{n} \sum (X - \bar{X})^3 / S^3$	
Эксцесс ( $\gamma_4$ )	$\frac{M(X - M X)^4}{\sigma^4} - 3$		
Функция распределения		$\frac{1}{n} \sum_k I(X_k < y)$	эмпирическая
Квантиль порядка $p \in (0; 1)$	$\sup\{x \mid F(x) \leq p\}$	$X_{([pn]+1)}$	член вариационного ряда

**Определение 3.** Сам метод состоит в следующем:

$$\hat{\theta}: L(\hat{\theta}) = \max_{\theta} L(\theta)$$

Однако проще искать максимум у  $\ln L(\theta)$ . Так можно в силу монотонности логарифма.

$$(\ln L(\theta))' = \frac{L'(\theta)}{L(\theta)}$$

<+гора примеров+>

### §\* Эффективные оценки

**Утверждение 1** (Неравенство Рао-Крамера). Пусть  $\theta, \hat{\theta}$  — параметр и его оценка,  $b(\theta) = M(\hat{\theta} - \theta)$  — смещение оценки,  $I(\theta)$  — информация Фишера,  $F(x, \theta)$  — параметрическое семейство распределений.

$$I(\theta) = M \left( \frac{\partial}{\partial \theta} \ln p(X, \theta) \right)^2,$$

где  $p(X, \theta)$  из определения 2.19.1. Если выполнены условия регулярности

1. Существует  $C \subset \mathbb{R}: \forall \theta \in \Theta P(X_1 \in C) = 1$  и  $\forall y \in C \sqrt{p(X, \theta)} \in C_{\theta}^1(\Theta)$

2.  $I(\theta) \in C_{\theta}(\Theta), I \geq 0$

и  $D \hat{\theta}$  ограничена на любом компакте  $\subset \Theta$ ,

то

$$M(\hat{\theta} - \theta)^2 \geq \frac{(1 + b'(\theta))}{nI(\theta)} + b^2(\theta)$$

*Замечание 1.* Всякая регулярность нужна, чтобы можно было законно записать производную по параметру под интеграл.

**Определение 4.** Оценка называется эффективной, если для неё неравенство Рао-Крамера обращается в равенство.

## § 20 Лемма Фишера

**Определение 1.** Пусть  $X_1, \dots, X_n$  — независимые случайные величины с распределением  $\mathcal{N}(0, 1)$ . Тогда  $\sum_i X_i^2$  имеет распределение  $\chi^2$  с  $n$  степенями свободы. Ещё так обозначается:  $K_n$

**Утверждение 1.** Плотность распределения  $\chi_n^2$  ищется по формуле

$$k_n = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2} \Gamma\left(\frac{n}{2}\right)}$$

**Определение 2.** Пусть  $X_1, \dots, X_n$  — независимые случайные величины с распределением  $\mathcal{N}(0, 1)$ . Тогда  $\frac{x}{\sqrt{\frac{1}{n} \sum_i X_i^2}}$  имеет распределение Стьюдента с  $n$  степенями свободы.

**Утверждение 2.** Плотность распределения  $T_n$  ищется по формуле

$$t_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

**Определение 3.** Пусть  $C: V_1 \rightarrow V_1$ . Тогда  $C$  — ортогональный, если  $CC^T = E$

**Следствие 1.**  $\det C = 1$

**Следствие 2.**  $\|Cx\| = \|x\|$

**Утверждение 3.** Оператор ортогональный  $\Leftrightarrow$  строки его матрицы (как векторы линейного пространства наборов чисел) образуют ортонормированный базис.

**Утверждение 4.** Пусть  $\{X_i\} \sim \mathcal{N}(0, 1)$ ,  $C$  — ортогональный линейный оператор. Тогда и  $Y = CX \sim \mathcal{N}(0, 1)$ .

▼

Докажется через утверждение о преобразовании плотности при замене переменных 2.10.2 и следствие 2. Независимость получится просто из того, что вышло нормальное распределение. А  $\sigma$ -у можно сначала засунуть в случайные величины, а потом вытащить обратно.

▲

**Лемма 5 (Фишера).** Пусть  $X_1, \dots, X_n$  независимы и  $X_i \sim \mathcal{N}(\theta, \sigma^2)$ . Тогда

$$1. \sqrt{n} \frac{\bar{X} - \theta}{\sigma} \sim \mathcal{N}(0, 1)$$

$$2. \bar{X}, S^2 \text{ независимы }^1$$

---

<sup>1</sup>Здесь смещённая дисперсия

$$3. \frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$4. \sqrt{n-1} \frac{\bar{X} - \theta}{S} \sim t_{n-1}$$

□

1. Заменим  $Z_i = \frac{X_i - \theta}{\sigma}$ ,  $Z_i \sim \mathcal{N}(0, 1)$  Найдем распределение  $\sum_i Z_i = \Sigma$ .

$$\Phi_i(t) = e^{-t^2/2}$$

$$\Phi_{S_n}(t) = \frac{1}{n} \cdot \exp\left(-n \frac{t^2}{2}\right) = \exp\left(-\left(\frac{\sqrt{n}t}{\sqrt{2}}\right)^2\right) \Rightarrow p_{\Sigma}(z) = \frac{1}{\sqrt{2\pi}\sqrt{n}} \exp\left(-\frac{z^2}{2n}\right)$$

$$\frac{S_n - n\theta}{\sigma} \sim \mathcal{N}(0, \sqrt{n}) \Rightarrow \frac{n}{\sqrt{n}} \frac{\bar{X} - \theta}{\sigma} = \sqrt{n} \frac{\bar{X} - \theta}{\sigma} \sim \mathcal{N}(0, 1)$$

2. Пусть

$$C = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \dots\dots\dots & & \end{pmatrix}$$

Первая строка как вектор имеет норму 1. А значит можно вспомнить процесс Грамма-Шмидта и набрать ортонормированный базис. Тогда  $C$  будет ортогональным. Пусть  $Y = CX$ . Из горы утверждений выше мы получим:

$$Y_1 = \frac{X_1 + \cdots + X_n}{\sqrt{n}} = \bar{X}\sqrt{n}$$

$$\|Y\| = \|X\| \Rightarrow \sum_i X_i^2 = \sum_i Y_i^2$$

$$S^2 = \frac{1}{n} \sum_i X_i^2 - (\bar{X})^2 = \frac{1}{n} \sum_{i=2}^n Y_i^2$$

А дальше надо честно посчитать  $\text{cov}\left(\frac{Y_1}{\sqrt{n}}, \sum_{i=2}^n Y_i\right)$ . Правда ноль получается. Если что,

$$\text{cov}(X, Y) = M(X - M X)(Y - M Y) = M(XY) - M X M Y$$

$$3. \frac{nS^2}{\sigma^2} = \sum_{i=2}^n \left(\frac{Y_i}{\sigma}\right)^2 = \chi_{n-1}^2$$

4. Мы там порешили, что  $\theta = 0$ , так что

$$\sqrt{n-1} \frac{\bar{X}}{S} = \frac{Y_1}{\sqrt{\frac{n}{n(n-1)} \sum_{i=2}^n Y_i^2}} = \frac{Y_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n Y_i^2}} \sim t_{n-1}$$

■

## § 21 Доверительны интервалы нормального распределения

Здесь собственно перешли к более интересной части — от точечных оценок параметров к построению для доверительных интервалов.

**Определение 1.**  $(T_1, T_2)$  — доверительный интервал уровня  $\gamma$ , если  $P(T_1 < \theta < T_2) = \gamma$

Дальше всюду ведутся рассуждения про доверительные интервалы (уровня  $\gamma$ ) параметров нормального распределения.

**Утверждение 1.** Доверительный интервал для  $\theta$  при известном  $\sigma$  равен  $\left( \bar{X} - \sigma \frac{z_{(1+\gamma)/2}}{\sqrt{n}}; \bar{X} + \sigma \frac{z_{(1+\gamma)/2}}{\sqrt{n}} \right)$

▼

Интервал ищем явно симметричный, так что посчитаем

$$P \left( \sqrt{n} \frac{|\bar{X} - \theta|}{\sigma} < z \right)$$

Поскольку величина внутри подчиняется стандартному нормальному распределению,

$$P \left( -z < \sqrt{n} \frac{(\bar{X} - \theta)}{\sigma} < z \right) = F_n(z) - F_n(-z) = 2F_n(z) - 1 = \gamma$$

По дороге сделали замену переменной, не пугайтесь. А дальше  $z = F_n(\frac{1+\gamma}{2})$ , что как раз соответствует определению  $\frac{1+\gamma}{2}$  квантили. Ну а дальше всё уже очевидно из преобразования неравенства выше. Мы умеем это делать, можно порассматривать  $\{\omega \mid X(\omega) \cdots\}$  как уже делали раньше.

▲

**Утверждение 2.** Доверительный интервал для  $\theta$  при неизвестном  $\sigma$  равен  $\left( \bar{X} - S \frac{t_{n-1, (1+\gamma)/2}}{\sqrt{n}}; \bar{X} + S \frac{t_{n-1, (1+\gamma)/2}}{\sqrt{n}} \right)$

▼

аналогично 2.21.4, только пользуемся 4 пунктом леммы Фишера 2.20.5.

▲

**Утверждение 3.** Доверительный интервал для  $\sigma^2$  при неизвестном  $\theta$  равен  $\left( \frac{nS^2}{v^2}; \frac{nS^2}{u} \right)$ . Чиселки  $u, v$  определяются с помощью  $\chi^2$ .

**Утверждение 4.** Доверительный интервал для  $\sigma^2$  при неизвестном  $\theta$  нормально не выражается. Проще численно.

$$1. \frac{\sum_{i=1}^n (X_i - \theta)^2}{\sigma^2} \sim \chi_n^2$$

$$2. \frac{n(\bar{X} - \theta)^2}{\sigma^2} \sim \chi_1^2$$

## § 22 Проверка гипотез по параметрам нормального распределения

Будем рассматривать здесь простую гипотезу:

$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$

1. Пусть  $X_1, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\sigma^2$  известно. Примем  $H_0: \theta = \theta_0$ . Но тогда

$$\sqrt{n} \frac{\bar{X} - \theta_0}{\sigma} \sim \mathcal{N}(0, 1)$$

из 1 пункта леммы Фишера (2.20.5).

Рассмотрим  $\alpha$  — уровень значимости — какое-нибудь маленькое число. Часто берут 0.05.

$$P\left(\sqrt{n} \frac{|\bar{X} - \theta_0|}{\sigma} > z\right) = \alpha \Leftrightarrow \sqrt{n} \frac{|\bar{X} - \theta_0|}{\sigma} > z_{1-\alpha/2}$$

Таким образом можно найти границы критической области.

Если  $H_0$  верна, то  $P\left(\sqrt{n} \frac{|\bar{X} - \theta|}{\sigma} > z_{1-\alpha/2}\right)$  мала. Можно выбрать, меньше чего мы хотим её сделать, и объявить сие критерием проверки. Собственно, так и делали на практике.

2.  $\sigma^2$  неизвестна. Здесь всё то же самое, только с распределением Стьюдента.

А здесь такую

$$H_0: \theta_1 = \theta_2$$

$$H_1: \theta_1 \neq \theta_2$$

Будем считать, что  $X_i, Y_i$  независимы, и нормально распределены:

$$X_1, \dots, X_{n_1} \sim \mathcal{N}(\theta_1, \sigma_1^2)$$

$$Y_1, \dots, Y_{n_2} \sim \mathcal{N}(\theta_2, \sigma_2^2)$$

1.  $\sigma_1^2, \sigma_2^2$  — известны.

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

2.  $\sigma_1^2 = \sigma_2^2$ , но не известны

$$\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sim S_{n_1 + n_2 - 2}$$

Это как раз тот самый **t-тест**

3.  $\sigma_1^2, \sigma_2^2$  неизвестны. Тут вообще ничего не понятно. (Проблемы Беренса-Фишера)

## § 23 Линейная регрессия

**Определение 1** (Регрессия). Пусть  $Y, X_1, \dots, X_m$  — случайные векторы. Тогда если определено уравнение  $y(x_1, \dots, x_m) = M(Y \mid X_1 = x_1, \dots, X_m = x_m)$ , то  $y$  называется регрессией  $Y$  по  $X_1, \dots, X_m$ .

**Определение 2** (Линейная регрессия). Пусть  $Y, X_1, \dots, X_m$  — случайные векторы. Тогда если определено уравнение  $y(x) = M(Y \mid X_i = x_i \forall i)$ , и  $y(x) = x \cdot \theta$  то  $y$  называется линейной регрессией  $Y$  по  $X$ .

Здесь  $x$  — матрица  $n \times m$ ,  $\theta \in R^m$ ,  $\epsilon \in R^n$ ,  $y \in R^n$

*Замечание 1.* Можно с тем же успехом написать  $Y = y(X) + \epsilon$ , если  $M\epsilon = 0$

**Определение 3.**  $Y$  называется откликом,  $X$  — регрессоры (предикторы),  $\epsilon$  — шум,  $\theta$  — параметры.

Основной метод поиска оптимальных параметров — по функции максимального правдоподобия. Если не сильно расписывать, то это выльется в  $\arg \min_{\hat{\theta}} (Y - X\hat{\theta})^T (Y - X\hat{\theta})$

А если расписывать, то получается следующее: раз все  $\epsilon_i$  независимы и нормально распределены, то функция максимального правдоподобия будет выглядеть так:

$$L(\theta_1, \dots, \theta_m) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left( -\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \sum_{k=1}^m x_{jk}\theta_k)^2 \right)$$

Как видно, условие максимума такой функции совпадает с минимумом суммы квадратов.

При этом нужны условия Гаусса-Маркова:

1.  $X^T X$  обратима
2.  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$  и независимы

**Утверждение 1.** Явное выражение для  $\hat{\theta}$  при минимизации выражения выше выглядит так:  $\hat{\theta} = (X^T X)^{-1} X^T Y$

▼

Пусть

$$Q(\theta) = (Y - X\theta)^T (Y - X\theta) = Y^T Y - \theta^T X^T Y - Y^T X \theta + \theta^T X^T X \theta \in R$$

В координатах это переписывается так (как обычно, суммирование по повторяющимся индексам)

$$Q(\theta) = y_i y_i - 2 y_j \theta_i X_{ji} + (X_{si} \theta_i) (X_{sj} \theta_j)$$

Тогда можно и продифференцировать

$$\begin{aligned} \frac{\partial Q}{\partial \theta_j} &= 2X_{si} X_{sj} \theta_j - 2X_{ji} y_j = 0 \Leftrightarrow 2X^T X \theta - 2X^T Y = 0 \Leftrightarrow \hat{\theta} = (X^T X)^{-1} X^T Y \\ \frac{\partial^2 Q}{\partial \theta_j \partial \theta_i} &= X_{si} X_{sj} \delta_{ij} \end{aligned}$$

Как видно, там и правда минимум. Здесь второй дифференциал просто сразу приведён к диагональному виду, и все числа на диагонали его матрицы положительны.

▲

## § 24 Теорема Гаусса-Маркова

**Определение 1.** Ковариационная матрица случайных векторов  $X, Y$  — матрица ковариаций их компонент

$$\text{cov}(X, Y)_{ij} = \text{cov}(X_i, Y_j) = M(X - M X)(Y - M Y)^T$$

**Определение 2.**  $\text{cov } X = \text{cov}(X, X)$

**Определение 3** (Эффективность оценок). Оценка параметра  $\hat{\theta}_1$  эффективнее оценки  $\hat{\theta}_2$ , если матрица  $\text{cov } \hat{\theta}_1 - \text{cov } \hat{\theta}_2$  отрицательно определена.

**Теорема 1.** Пусть

1.  $X^T X$  обратима
2.  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$  и независимы

Тогда

1.  $\hat{\theta} = (X^T X)^{-1} X^T Y$  — несмещённая оценка  $\theta$ ,
2.  $\hat{\theta}$  — наиболее эффективная среди линейных несмещённых оценок

□

1.  $\hat{\theta} = (X^T X)^{-1} X^T (X\theta + \varepsilon) = \theta + (X^T X)^{-1} X^T \varepsilon$ . Так как  $M \varepsilon = 0$ ,  $\varepsilon$  не зависит от  $X$ , последнее слагаемое обращается в ноль
2. Пусть  $\tilde{\theta} = HY$ . Такая оценка тоже будет несмещённой, если  $HX = E$ . Тогда

$$\text{cov } HY = (M HY - \theta)(HY - \theta)^T = M((HX - E)\theta + \varepsilon)((HX - E)\theta + \varepsilon)^T = M H \varepsilon \varepsilon^T H^T = \sigma^2 H H^T$$

Для изначальной оценки  $H_0 = (X^T X)^{-1} X^T$ , так что  $H_0 H_0^T = (X^T X)^{-1} X^T X (X^T X)^{-1} = (X^T X)^{-1}$ .

Покажем, что матрица  $H H^T - (X^T X)^{-1}$  положительно определена. Пусть  $C = H - (X^T X)^{-1}$ . Тогда

$$\begin{aligned} CX &= CH - E = 0 \\ HH^T &= (C + (X^T X)^{-1} X^T)(C^T + X (X^T X)^{-1}) = CC^T + (X^T X)^{-1} \end{aligned}$$

А матрицы вида  $CC^T$  обычно (над  $\mathbb{R}$ ) положительно определены.

■

## § 25 Оценка дисперсии погрешностей

**Теорема 1.** Пусть  $S^2 = \frac{1}{n-m} (Y - X\hat{\theta})(Y - X\hat{\theta})^T$ . Тогда  $S^2$  — несмещённая оценка  $\sigma^2$

□

$$Y - X\hat{\theta} = (E_n - X(X^T X)^{-1} X^T)(X\theta + \varepsilon) = (E_n - X(X^T X)^{-1} X^T)\varepsilon = B\varepsilon$$

Тогда  $S^2 = \frac{1}{n-m} \varepsilon^T B^T B \varepsilon$ , причём  $B$  — симметрична. Так что  $B^T B = B^2 = B$  (лень расписывать). Так что

$$M S^2 = \frac{1}{n-m} M(\varepsilon^T B \varepsilon) = \frac{\sigma^2}{n-m} \text{Tr } B$$

потому что всё, что не на главной диагонали обратится в ноль в силу независимости  $\varepsilon_i$ .

Осталось доказать, что  $\text{Tr}(X(X^T X)^{-1} X^T)$ . Покажем, что след произведения матриц не зависит от порядка сомножителей.

$$\text{Tr } AB = (ab)_i = a_{is} b_{si} = b_{si} a_{is} = (ba)_s$$

Так что просто переставим  $X^T$  в самое начало и получим  $E_m$ . ■

## § 26 Критерий согласия Пирсона

**Определение 1.** Пусть  $X_1, \dots, X_n \sim F$ . Будем проверять гипотезу

$$H_0: F = F_0 \quad H_1: F \neq F_0$$

Пусть существует функция  $\rho(X)$ , такая что

1.  $H_0$  верна  $\Rightarrow \rho(X) \xrightarrow{d} G$ ,  $G$  — некоторое непрерывное распределение.
2. если  $H_0$  неверна, то  $|\rho(X)| \xrightarrow{p} \infty$

Выберем критическую область по распределению  $G$  из равенства  $P(|xg| \geq C) = \gamma$ ,  $g \sim G$ . Тогда критерий введём так: будем отвергать  $H_0$ , если  $|\rho(X)| \geq C$ .

**Определение 2** (Критерий согласия Пирсона). Разобьём всю область значений  $X_i$  на интервалы  $I_i$ , с заданными вероятностями —  $p_i$ ,  $\sum_i p_i = 1$ . Будем проверять гипотезу  $\forall i P(X_1 \in I_i) = p_i$ . Этакая дискретизация распределения.

Ещё обозначим  $\nu_k = \#\{i \mid X_i \in I_k\}$ . Пусть

$$\rho = \sum_{k=1}^r \frac{(\nu_k - np_k)^2}{np_k}$$

А сам критерий основывается на сходимости распределения выше к распределению  $\chi^2$ , и критическая область выбирается из этого распределения.

**Теорема 1.** При справедливости  $H_0$   $\rho(X) \xrightarrow[n \rightarrow \infty]{d} \chi_{r-1}^2$ .

## § 27 Непараметрические критерии

Условия на  $X_1, \dots, X_n$  те же



**Утверждение 1** (Критерий Колмогорова-Смирнова). Пусть  $F_n$  — эмпирическая функция распределения,

$$D_n^+ = \sup_x (F_n(x) - F_0(x))$$

$$D_n^- = \sup_x (F_0(x) - F_n(x))$$

$$\text{Тогда } P(\sqrt{n}D_n^+ < z) \rightarrow \begin{cases} \sum_{k \in \mathbb{Z}} (-1)^k e^{-2k^2 z^2}, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

Здесь функция  $\rho = D_n^+$ . Чтобы доказать, что она — критерий согласия, можно воспользоваться теоремой Гливенко-Кантелли

$$P(|F_n(x) - F_0(x)| \rightarrow 0) = 1$$

Распределение в котором всё сошлось — распределение Колмогорова.

**Утверждение 2** (Критерий Смирнова).  $\rho = \max(D_n^+, D_n^-)$

**Утверждение 3** (Критерий Койгера).  $\rho = (D_n^+ + D_n^-)$

**Утверждение 4** (Критерий Крамера-фон-Мизеса).  $\rho = \omega^2 = \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 dF_0(x)$

**Утверждение 5** (Критерий Андерсона-Дарлинга).  $\rho = \int_{-\infty}^{+\infty} \frac{(F_n(x) - F_0(x))^2}{F_n(x)(1 - F_0(x))} dF_0(x)$

Здесь ещё была теорема почему они непараметрические, но на неё похоже забыли.

## Глава 3: Случайные процессы

### § 28 Процессы с независимыми приращениями

**Определение 1.** Случайный процесс — измеримое отображение  $X: \Omega \rightarrow L(T)$ ,  $L(T)$  — пространство функций над  $T$

**Пример 1.**  $T = t_0$ , тогда  $X$  — случайная величина

**Определение 2.**  $X$  — процесс с независимыми приращениями, если  $\forall t_0 < \dots < t_n \in T$  случайные величины

$$X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$$

**Пример 1.**  $T = \mathbb{N}$ , а сам независимый процесс  $X(n) = \sum_{i=1}^n Y_i$ ,  $Y_i$  — независимы.

**Определение 3.** Пусть  $T = \mathbb{R}$  (время) и при этом:

$$X(0) = 0$$

$$X(t) - X(s) \sim \Pi(\lambda(t - s))$$

$X$  — процесс с независимыми приращениями

**Определение 4** (Винеровский процесс(броуновское движение)). Пусть  $T = \mathbb{R}$  (время) и при этом:

$$\begin{aligned} X(0) &= 0 \\ X(t) - X(s) &\sim \mathcal{N}(0, t - s) \\ X &\text{ — процесс с независимыми приращениями} \end{aligned}$$

## § 29 Стационарные процессы

**Определение 1** (Стационарные в узком смысле).  $X(t)$  называется стационарным в узком смысле, если

$$\begin{aligned} \forall t_1, \dots, t_n \in T, \forall \tau > 0 \quad (t_1 + \tau, \dots, t_n + \tau \in T) \\ \Rightarrow (X(t_1 + \tau), \dots, X(t_n + \tau)) \stackrel{d}{=} (X(t_1), \dots, X(t_n)) \end{aligned}$$

Короче, можно двигать начало отсчёта времени, распределение не изменится.

**Следствие 1.**

$$\begin{aligned} m(t) &= M X(t) = \text{const} \\ \sigma^2(t) &= D X(t) = \text{const} \\ \text{cov}(X(s), X(t)) &= \text{cov}(X(0), X(t - s)) =: R(t - s) \end{aligned}$$

Здесь определена величина  $R(t - s)$ , если что.

**Определение 2** (Стационарные в широком смысле).  $X(t)$  называется стационарным в широком смысле, если  $M X(t) = \text{const}$  и  $\text{cov}(X(s), X(t)) = R(t - s)$ .

Сделаем теперь из процессов (пока любых) линейное пространство со скалярным произведением.

$$H_0 = \left\{ \sum_{k=1}^n c_k X(t_k) \mid n \in \mathbb{N}, t_i \in T, c_i \in \mathbb{R} \right\} \quad (3.1)$$

$$\langle X(s), X(t) \rangle = \text{cov}(X(s), X(t)) \quad (3.2)$$

Теперь сделаем из него гильбертово пространство, пополнив по метрике, соответствующей скалярному произведению. Далее надо бы доказать, что оно вообще расстояние

<+Здесь дальше какая-то жесть про спектральную меру. Я её не понимаю+>

## § 30 Цепи Маркова

**Определение 1.** Пусть  $\mathfrak{X}$  — дискретное множество состояний. Тогда  $\xi_i$  образуют цепь Маркова, если

$$P(\xi_n = i_n \mid \xi_{n-1} = i_{n-1}, \xi_1 = i_1) = P(\xi_n = i_n \mid \xi_{n-1} = i_{n-1})$$

(Цепь помнит только свое предыдущее состояние)

**Определение 2.** Если известны  $q_i = P(\xi_1 = i)$  и  $p_{ij} = \frac{1}{n} P(\xi_n = j \mid \xi_{n-1} = i)$ , то цепь называется полностью определённой.  $P = (p_{ij})$  ещё называется стохастической матрицей.

**Определение 3.** Если  $np_{ij}$  не зависит от  $n$ , то цепь называется однородной.

**Определение 4.**  $p_{ij}^{(n)} = P(\xi_{k+n} = j \mid \xi_k = i)$  (вероятность перейти за  $n$  шагов).

**Утверждение 1.** Для однородной цепи:

$$p_{ij}^{(n+m)} = \sum_k p_{ik}^{(m)} \cdot p_{kj}^{(n)} P(n) = P^n$$

Оно всё следует из независимости от старых состояний и формулы полной вероятности 1.2.1.

**Пример 1.** Если  $\xi_i$  независимы, то они образуют цепь Маркова

**Пример 2.** Если  $\xi_i$  независимы, то  $\eta_n: \eta_n = f(\eta_{n-1}, \xi_n)$  образуют цепь Маркова.

**Определение 5** (Случайные блуждания). Случайное блуждание — процесс с дискретным временем вида  $\eta_0 + \sum_i \xi_i$ , где  $\xi_n \in \mathbb{R}^d$  — независимые случайные величины,  $\xi_k$  принимает значения  $\pm e_i$  — ортонормированный базис  $\mathbb{Z}^d$

Случайные блуждания тоже можно считать цепью Маркова

**Определение 6.**  $f_{ii}^{(n)} = P(\xi_{n+1} = i \mid \xi_n \neq i, \dots, \xi_1 = i)$ . Состояние  $i$  возвратное, если  $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$  и невозвратное иначе.

Но надо ещё подумать над вероятностью вернуться в какое-то состояние, идя не важно как.

$$p_{ii}^{(n)} = f_{ii}^{(n)} + f_{ii}^{(n-1)} p_{ii} + \dots + f_{ii}^{(1)} p_{ii}^{(n-1)}$$

С таким можно разобраться при помощи производящих функций

$$P(z) = \sum_{n=0}^{\infty} p_{ii}^{(n)} z^n, \quad p_{ii}(0) = 1$$

$$F(z) = \sum_{n=1}^{\infty} f_{ii}^{(n)} z^n$$

$$P(z) = P(z)F(z) + 1$$

Получить последнюю формулу можно честно перемножив два ряда. Только нужно ещё не забыть, что один из них начинается с  $z^1$ . А дальше член при  $z^n$  как раз и оказывается суммой выше

**Утверждение 2** (Критерий возврата).  $i$  — возвратное  $\Leftrightarrow \sum_{n=0}^{\infty} p_{ii}^{(n)} = +\infty$

▼

Из формулы выше

$$P(z) = \frac{1}{1 - F(z)}$$

а  $F(1) = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$

▲

### § 31 Марковские процессы

Здесь будем брать и сурово строить сигма-алгебры на заданных множествах.

**Определение 1.**  $\sigma(Y_1, \dots, Y_n)$  — наименьшая сигма-алгебра, относительно которой  $Y_i$  измеримы. Короче, порождённая набором  $\{Y_i\}$ .

*Замечание.* Индекс  $i$  может быть и непрерывным:  $\sigma(X(t) \mid t \in T)$

**Пример 1.**  $F_a^b = \sigma(X(t), t \in [a, b])$  — сигма-алгебра интервала.

**Пример 2.**  $F_\infty^t = \sigma(X(\zeta), \zeta \in (-\infty, t])$  — сигма-алгебра прошлого. Можно так же и будущего. И настоящего ( $\zeta \in \{t\}$ ).

**Определение 2** (Марковский процесс).  $X(t)$  — марковский процесс, если

$$\forall A \in F_{-\infty}^t, B \in F_t^\infty \quad P(AB) = P(A)P(B)$$

Короче говоря, прошлое не зависит от будущего.

## Глава А: Обозначения

$\Omega$  — пространство элементарных исходов

$\omega$  — элемент пространства элементарных исходов

$\mathcal{A}$  — алгебра множеств

$\mathcal{F}$  —  $\sigma$ -алгебра множеств (случайные события)

$P$  — вероятность

$p(x_1, \dots, x_n)$  — плотность вероятности

$X(\omega), Y(\omega), \xi(\omega)$  — случайная величина

$F_X(x_1, \dots, x_n)$  — функция распределения случайной величины  $X$

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

✂ — ещё правится. Впрочем, относится почти ко всему.

□...■ — начало и конец доказательства теоремы

▼...▲ — начало и конец доказательства более мелкого утверждения

⋈ — кривоватая формулировка

:set aflame — набирающему зело не нравится билет

<+что-то+> — тут будет что-то, но попозже

## Глава В: Стандартные распределения

Распределение	Плотность
Бернулли	$P(x = 1) = p, P(x = 0) = 0$
Пуассона	$P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k \in \mathbb{N} \cup 0$
Биномиальное	$P(x = k) = C_n^k p^k (1 - p)^{n-k}$

## Литература

- [1] Чернова Н.И. Теория вероятностей.
- [2] Ширяев А.Н. *Вероятность*. М.: Наука, 1980.
- [3] Пономаренко Л.С. Прохоров Ю.В. *Лекции по теории вероятности и математической статистике*. 2-е изд., испр. и доп. М.: Издательство Московского университета, 2012.