ТЕОРВЕR
МАТСТАТИSТИКА
и тщетные попытки понять что-нибудь из вышеперечисленного

Лектор: Р. С. Пусев
Записал :taxus

19 января 2017 г.

Оглавление

Глава 1: Элементарная теория вероятностей

§ 1 Аксиоматическое определение вероятности

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

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

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

Ω — Множество (элементарных исходов). Чисел, например.

$\mathcal{F}$ — σ — алгебра подмножеств Ω

P — Собственно, вероятность

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

1. $\forall\, A \;\; F(A) \geqslant 0$
2. $\forall \{A_i\}\colon A_i \cap A_j = \varnothing \;\; P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$
3. $P(\Omega) = 1$

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

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

Утверждение 1. *Свйоства вероятности:*

1. $P(\varnothing) = 0$
2. $0 \leqslant P(A) \leqslant 1$
3. $P(\overline{A}) = 1 - P(A)$
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* $A_1 \subset \cdots A_n \subset \cdot$, $\bigcup\limits_i A_i = A$. *Тогда* $P(A) = \lim\limits_{n\to\infty} P(A_n)$.

▼

Пусть

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

Тогда

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

Следовательно,

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

Всё работает, потому что в σ-алгебре можно объединять счётное число множеств.

▲

§ 2 Формула полной вероятности

Определение 1 (Условная вероятность)**.** Пусть $A, B \in \mathcal{F}$, $P(A) > 0$[1]. 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. $A_i \cap A_j = \varnothing$
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*

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

[1] ну мы же делим, нажо убедиться что неноль

§ 4 Независимые события

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

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

Утверждение 1. *События A, B независимы $\Leftrightarrow P(A \mid B) = P(A) \vee P(B \mid A) = P(B)$ (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 события $\{A_i\}$ попарно независимы, если

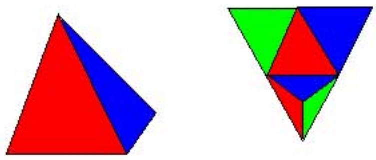
$$\forall\, i, j \;\; P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

.

Определение 3. Случайные события $\{A_i\}_{i=1}^{n}$ независимы по совокупности , если

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

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 (Случаная величина)**.** Случайной величиной назовём произвольное хорошее отображение $X\colon \Omega \to \mathbb{R}$.

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

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

где B — борелевское множество

Определение 2. Пусть $B \subset \mathbb{R}$, B — промежуток, или дополнение в нему (борелевское множество).[1]

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

Определение 3. Случайная величина называется дискретной, если

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

то есть

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

[1] тут вроде концы могут входить, так что точка — тоже борелевское множество.

Определение 4. Случайная величина называется непрерывной, если

$$\exists\,(f_X\colon B \to \mathbb{R})\colon \left(P(X \in B) = \int\limits_B f_X(x)\,\mathrm{d}x\right)$$

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 $f_X \equiv \frac{1}{2}$.

$$F(B) = P(X \in B) = P(\{\omega \mid \omega \in (0,1)\}) = P\big((0,1)\big) = \tfrac{1-0}{1+1} = \tfrac{1}{2}$$
$$F(B) = \int\limits_{(0,1)} \tfrac{1}{2}\,\mathrm{d}x = \tfrac{1-0}{2} = \tfrac{1}{2}$$

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

Определение 6 (Функция распределения)**.**

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

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

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

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

$$\int\limits_{-\infty}^{x} f_X(u)\,\mathrm{d}u$$

§ 6 Моменты случайных величин

Определение 1. Пусть X — случайная величина,

$$\int\limits_{-\infty}^{\infty} |x| f_X(x)\,\mathrm{d}x < \infty$$

Тогда

$$\langle x \rangle \equiv \bar{x} \equiv \mathrm{M}\,X = \int\limits_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$

Утверждение 1. *Свойства матожидания:*

1. $\mathrm{M} < \infty$
2. $\mathrm{M}(aX + bY) = a\,\mathrm{M}\,X + b\,\mathrm{M}\,Y$
3. $P(X \geqslant 0) = 1 \Rightarrow \mathrm{M}\,X \geqslant 0$
4. $\begin{cases} P(X \geqslant 0) = 1 \\ \mathrm{M}\,X = 0 \end{cases} \quad \Rightarrow P(X = 0) = 1$
5. *если* X, Y *— независимы, то* $\mathrm{M}(XY) = \mathrm{M}\,X \cdot \mathrm{M}\,Y$

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

$$\lambda_{k,a} = \int\limits_{-\infty}^{\infty} (x-a)^k f(x)\,\mathrm{d}x$$

(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. $\nu_k = \sum\limits_{i=0}^{k} C_k^i a^i \,\lambda_{k-i,a}$

Определение 5 (Дисперсия)**.** $\mathrm{D}\,X = \mathrm{M}(X - \mathrm{M}\,X)^2$, $\sigma = \sqrt{\mathrm{D}\,X}$ — среднеквадратичное отклонение.

Утверждение 3.
$\mathrm{D}(aX + bY) = a^2\,\mathrm{D}\,X + b^2\,\mathrm{D}\,Y$
если X, Y *— независимы, то* $\mathrm{D}(XY) = \mathrm{D}\,X \cdot \mathrm{D}\,Y$
$\mathrm{D}(X + C) = \mathrm{D}\,X$

§ 7 Характеристическая функция

Определение 1 (Характеристическая функция)**.** $\Phi(t) = \mathrm{M}\,e^{itx}$

Утверждение 1. *Свойства характеристической функции:*

1. *Всегда существует и* $|\Phi(t)| \leqslant 1$.
2. $f(x) = \dfrac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx}\Phi(t)\,\mathrm{d}t$
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* $\mathrm{M}\,|X|^n < \infty$, *то* $\Phi^{(k)}(0) = i^k\,\mathrm{M}\,X^k$

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

Теорема 2 (О непрерывном соответствии)**.** $X_n \xrightarrow{d} X \Leftrightarrow \Phi_{X_n}(t) \underset{t}{\rightrightarrows} \Phi_X(t)$. *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.*

§ 8 Теорема Муавра-Лапласа

Глава 2: Матстатистика

§ 9 Случайные векторы

Определение 1 (Случайный вектор)**.** Случайным вектором назовём произвольное хорошее отображение $X\colon \Omega \to \mathbb{R}^n$. 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)**.**

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

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 следущее:*

1. *F не убывает по каждому аргументу.*
2. $\lim\limits_{\exists\, i\ x_i \to -\infty} F(x) = 0$
3. $\lim\limits_{\forall\, i\ x_i \to +\infty} F(x) = 1$
4. $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$ *(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*

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

По сути следствие формулки про вероятность объединения.

Замечание 1. Если для многомерной функции верны 3 свойства и ещё вероятность попасть в прямоугольничек вообще определена как вероятность ($P \in [0, 1]$).

Определение 3. Векторная случайная величина называется дискретной, если

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

то есть

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

Определение 4. Векторная случайная величина называется непрерывной, если

$$\exists\,(f_X\colon B \to \mathbb{R})\colon \left(P(X \in B) = \int\limits_B p_X(x^1, \ldots, x^n)\,\mathrm{d}x^1 \cdots \mathrm{d}x^n\right)$$

Замечание 1. Для функций распределения:

$$F(x^1, \ldots, x^n) = \int\limits_{-\infty}^{x^n} \cdots \int\limits_{-\infty}^{x^1} p_X(x^1, \ldots, x^n)\,\mathrm{d}x^1 \cdots \mathrm{d}x^n$$

Утверждение 3. $p(x_1, \ldots, x_{n-1}) = \int\limits_{-\infty}^{+\infty} p(x_1, \ldots, x_{n-1}, x_n)\,\mathrm{d}x_n$

▼

Покажем, что $\int\limits_{-\infty}^{+\infty} p(x, y)\,\mathrm{d}y = p_X(x)$

$$\begin{aligned}
\int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} p(x, y)\,\mathrm{d}x\right)\mathrm{d}y = F(x, +\infty) &= P(X < x, Y < +\infty) \\
&= P(\{\omega \mid X(\omega) \in (-\infty, x), Y \in \mathbb{R}\}) \\
&= P(X < x) = F_X(x) = \int\limits_{-\infty}^{x} p_X(x)\,\mathrm{d}x
\end{aligned}$$

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

▲

Определение 5 (Независимые случайные величины)**.** Слйчаные величины $\{X_i\}$ независимые, если

$$\forall\,\{B_i\}\;\; P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$$

Утверждение 4. *Пусть X, Y — независимы. Тогда $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$*

▼

По определению функции распределения

$$F_{X,Y}(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(x, y)\,\mathrm{d}x\mathrm{d}y$$

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

$$F_X(x) \cdot F_Y(y) = \int\limits_{-\infty}^{x} p_X(x)\,\mathrm{d}x \cdot \int\limits_{-\infty}^{y} p_Y(y)\,\mathrm{d}y = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p_X(x) \cdot p_Y(y)\,\mathrm{d}x\mathrm{d}y$$

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

$$\int\limits_{B} (p(x,y) - p_X(x) \cdot p_Y(y))\,\mathrm{d}x\mathrm{d}y = 0$$

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)**.** $g\colon \mathbb{R}^n \to \mathbb{R}^m$.

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 $p_y(y) = p_X(g(y)) \left|\frac{\mathrm{d}g}{\mathrm{d}y}\right|$.*

▼

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)

$$\int\limits_{D} p_Y(y)\,\mathrm{d}y = \int\limits_{B} p_X(x)\mathrm{d}x = \int\limits_{D} p_X(g(y)) \left|\frac{\mathrm{d}g}{\mathrm{d}y}\right|$$

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.

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

▲

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

$$p_{X_1+X_2}(y) = \int\limits_{\mathbb{R}} p_1(y-t) p_2(t)\,\mathrm{d}t$$

▼

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

$$F_Y(y) = \iint\limits_{x_1+x_2<y} p(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 = \int\limits_{-\infty}^{\infty} \mathrm{d}x_1 \int\limits_{-\infty}^{y-x_1} p(x_1,x_2)\,\mathrm{d}x_2 = \int\limits_{-\infty}^{\infty} \mathrm{d}x_1 \int\limits_{-\infty}^{y} p(x_1, u-x_1)\,\mathrm{d}u$$

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

$$p_Y(y) = \int\limits_{-\infty}^{\infty} p(t, y-t)\,\mathrm{d}t$$

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

$$p_Y(y) = \int\limits_{-\infty}^{\infty} p_1(y-t) \cdot p_2(t)\,\mathrm{d}t$$

▲

Теперь можно перейти и к содержанию билета

Утверждение 3. $\mathrm{M}\left(\sum_i X_i\right) = \sum_i \mathrm{M}\,X_i$. *Да и вообще оно линейно.*

[1] слишком много раз пользовались на физике, так что оставим на 4 семестр

▼

Пусть $f(X,Y)=X+Y$

$$\begin{aligned}
\mathrm{M}(X+Y) &= \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)p(x,y)\,\mathrm{d}x\mathrm{d}y = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xp(x,y)\,\mathrm{d}x\mathrm{d}y + \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} yp(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int\limits_{-\infty}^{+\infty} x\left(\int\limits_{-\infty}^{+\infty} p(x,y)\mathrm{d}y\right)\,\mathrm{d}x + \int\limits_{-\infty}^{+\infty} y\left(\int\limits_{-\infty}^{+\infty} p(x,y)\mathrm{d}x\right)\,\mathrm{d}y \\
&= \mathrm{M}\,X + \mathrm{M}\,Y
\end{aligned}$$

Часть про константу слишком очевидна, не будем её доказывать.

▲

Утверждение 4 (Дисперсия суммы)**.** $\mathrm{D}\left(\sum_i X_i\right) = \sum_i \mathrm{D}\,X_i$

▼

Сначала заметим, что $\mathrm{D}\,X = \mathrm{M}(X-\mathrm{M}\,X)^2$, $\mathrm{M}(X-\mathrm{M}\,X) = \mathrm{M}\,X - \mathrm{M}\,X = 0$

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

▲

Утверждение 5. *Если* X, Y *— независимы, то* $\mathrm{M}\,XY = \mathrm{M}\,X\,\mathrm{M}\,Y$

§ 12 Матожидание функции случайной величины

Определение 1 (ᐤ)**.** Пусть $f(X)$ — функция от случаной величины. Тогда $\mathrm{M}\,f(X) = \int\limits_{-\infty}^{+\infty} f(x)p(x)\,\mathrm{d}x$. 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* $\mathrm{M}\,f_1(X_1)f_2(X_2) = \mathrm{M}\,f_1(X_1)\,\mathrm{M}\,f_2(X_2)$

§ 13 Неравенство Шварца

Утверждение 1. $(\mathrm{M}\,XY)^2 \leqslant \mathrm{M}\,X^2\,\mathrm{M}\,Y^2$

▼

$\mathrm{M}(X+tY)^2 = t^2\,\mathrm{M}\,Y^2 + 2t\,\mathrm{M}\,XY + \mathrm{M}\,X^2 \geqslant 0$ из свойств матожидания. Ну там и подынтегральная функция положительна. Тогда квадратное уравнение в правой части может иметть не более одного корня.

$$(2\,\mathrm{M}\,XY)^2 - 4\,\mathrm{M}\,X^2\,\mathrm{M}\,Y^2 \leqslant 0 \Leftrightarrow (\mathrm{M}\,XY)^2 \leqslant \mathrm{M}\,X^2\,\mathrm{M}\,Y^2$$

▲

§ 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 2.11.3

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

▲

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

$$p_{X_1+\dots+X_n} = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \big(\Phi(t)\big)^n \,\mathrm{d}t$$

§ 15 Суммирование большого числа случайных величин

`:set aflame⚒⍨`

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

☐ Пусть $\mathrm{M}\, X_k = a, \mathrm{D}\, X_k = \sigma^2$. Рассмотрим характеристическую функцию $\Phi(t) = \mathrm{M}\, e^{itX_k}$. 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

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

А характеристическая функция дифференцируема дважды из существования дисперсии.

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

А при $n \to \infty$

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

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

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

$$F_{Z_n}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\,\mathrm{d}s$$

Дальше — вариация из Агекяна. Используя утверждение 2.10.2 про замену переменной как раз получаем нормальное распределение. Только тут нужно поменять в процессе σ

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

$$F_{S_n}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma\sqrt{n}} e^{-\frac{u^2}{2\sigma^2 n}}\,\mathrm{d}u$$

Вернёмся обратно к ненулевому a

$$S_n = S_n - na$$

$$F_{S_n} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma\sqrt{n}}\int\limits_{-\infty}^{x} e^{-\frac{(u-na)^2}{2\sigma^2 n}}\,\mathrm{d}u$$

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

■

§ 16 Центральная предельная теорема

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

$$S_n = \sum_i x_i$$

$$a_k = \mathrm{M}\,X_k \qquad \sigma_k^2 = \mathrm{D}\,x_k \qquad \gamma_k = \mathrm{M}\,|X_k - a_k|^3$$

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

Тогда

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

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 $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ и независимы. Тогда говорят, что случайная величина $\chi_n^2 = \sum\limits_{k=1}^{n} X_k^2$ имеет распределение χ^2 с n степенями свободы.

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

▼

Характеристическая функция χ может быть найдена из 1. Найдём сначала характеристическую функцию X_k^2. Для этого было бы недурно найти плотность соответвующего распределения

$$P(y < X^2 < y + \mathrm{d}y) = P(\sqrt{y} < X < \sqrt{y+\mathrm{d}y}) + P(-\sqrt{y} > X > -\sqrt{y+\mathrm{d}y})$$
$$= \frac{2}{\sqrt{2\pi}} \int\limits_{\sqrt{y}}^{\sqrt{y+\mathrm{d}y}} e^{-u^2/2}\,\mathrm{d}u = \sqrt{\frac{2}{\pi}}\frac{e^{-y/2}}{2\sqrt{y}}$$

А теперь можно и фурье-образ найти

$$\int\limits_{-\infty}^{+\infty} \frac{e^{-y/2}}{2\sqrt{y}}\mathrm{d}y = \int\limits_{0}^{+\infty} \frac{e^{-y/2}}{2\sqrt{y}}\mathrm{d}y = \int\limits_{0}^{+\infty} \exp\left(\frac{-\eta^2(1-2it)}{2}\right)\mathrm{d}\eta = (1-2it)^{-1/2}\sqrt{\frac{\pi}{2}}$$

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

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

Тогда

$$p_\chi(z) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} (1-2it)^{-n/2}\, e^{-itz}\mathrm{d}t$$

Дальше немного жесть

$$\int\limits_{-\infty}^{+\infty} (1-2it)^{-n/2}\, e^{-itz}\mathrm{d}t = \int\limits_{-\infty}^{+\infty} e^{-l}\left(1 - 2\,\frac{l}{z}\right)^{-n/2}\frac{1}{iz}\,\mathrm{d}l = 2\cdot\frac{z^{n/2-1}e^{-z/2}}{2^{n/2}}\int\limits_{0}^{\infty} s^{-n/2}e^{-s}\mathrm{d}s$$
$$= 2\frac{z^{n/2-1}e^{-z/2}}{2^{n/2}}\cdot\Gamma\left(1-\frac{n}{2}\right)$$

Из правила отражения для Γ-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* $X_1, \ldots, X_n$ *— независимые случайные величины с дискретным распределением:*

$$X_k\colon \begin{array}{c|c|c} 1 & \cdots & r \\ \hline p_1 & \cdots & p_r \end{array}$$

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

$$\sum_{k=1}^{r}\left(\frac{\nu_k - np_k}{\sqrt{np_k}}\right)^2 \xrightarrow[n\to\infty]{d} \chi^2_{r-1}$$

§ 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: $P_\alpha(T_1 < \hat{\theta} < T_2) = \alpha$
2. Проверка гипотез

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

   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:

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

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

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

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

Генеральная совокупность		Выборка	
Матожидание	$\mathrm{M}\,X$	$\overline{X} = \frac{1}{n}\sum_k X_k$	Выборочное среднее
Дисперсия	$\mathrm{D}\,X$	$S^2 = \frac{1}{n}\sum_k (X_k - \overline{X})$	Выборочная дисперсия
Момент порядка l	$\mathrm{M}\,X^k$	$m_l = \frac{1}{n}\sum_k X_k^l$	
Ковариация	$\mathrm{M}(X - \mathrm{M}\,X)(Y - \mathrm{M}\,Y)$	$\frac{1}{n}\sum_k (X_k - \overline{X})(Y_k - \overline{Y})$	
Ассиметрия (γ_3)	$\mathrm{M}(X - \mathrm{M}\,X)^3/\sigma^3$	$\frac{1}{n}\sum (X - \overline{X})^3/S^3$	
Эксцесс (γ_4)	$\frac{\mathrm{M}(X - \mathrm{M}\,X)^4}{\sigma^4} - 3$	$\frac{1}{n}\sum_k (X - \overline{X})^4/S^4 - 3$	
Функция распределения	$P(X < x)$	$\frac{1}{n}\sum_k I(X_k < y)$	эмпирическая
Квантиль порядка $p \in (0;1)$	$\sup\{x \mid F(x) \leqslant p\}$	$X_{([pn]+1)}$	член вариационного ряда

⚒Свойства оценок

1. Несмещённость: $\mathrm{M}\,\hat{\theta} = \theta$
2. Асимпотическая несмещённость: $\mathrm{M}\,\hat{\theta} \xrightarrow[n\to\infty]{} \theta$

Метод моментов

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

$$\hat{\theta} = h^{-1}(m),\ m \text{ — эмпирический момент}$$

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 $L(\theta) := \prod\limits_{k=1}^{n} p(X_k,\theta)$ — функция правдоподобия. Ещё берут её логарифм.

Определение 3. Сам метод состоит в следующем:

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

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

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

`<+гора примеров+>`

§* Эффективные оценки

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

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

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

1. *Существует $C \subset \mathbb{R}\colon \forall\, \theta \in \Theta P(X_1 \in C) = 1$ и $\forall\, y \in C \sqrt{p(X, \theta)} \in C^1_\theta(\Theta)$*
2. *$I(\theta) \in C_\theta(\Theta)$, $I \geqslant 0$*

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

то

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

Замечание 1. Всякая регулярность нужна, чтобы можно было законно запихивать производную по параметру по интеграл.

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

§ 20 Лемма Фишера

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

Утверждение 1. *Плотность распределения χ^2_n 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 $X_1, \ldots, X_n$ — независимые случайные величины с распределением $\mathcal{N}(0, 1)$. Тогда $\dfrac{X_1}{\sqrt{\frac{1}{n}\sum_i X_i^2}}$ имеет распределение Стьюдента с n степенями свободы.

Утверждение 2. *Плотность распределения T_n ищется по формуле*

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

Определение 3. Пусть $C\colon V_1 \to V_1$. Тогда C — ортгональный, если $CC^T = E$

Следствие 1. $\det C = 1$

Следствие 2. $\|Cx\| = \|x\|$

Утверждение 3. *Оператор ортогональный $\Leftrightarrow$ строки его матрицы (как векторы линейного пространсва 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 пребразовании плотности при замене переменных 2.10.2 и следствие 2. Независимость получится просто из того, что вышло нормальное распределение. А σ-у можно сначала засунуть в случайные величины, а потом вытащить обратно.
▲

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

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

2. $\overline{X}, S^2$ *независимы* [1]

3. $\dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

4. $\sqrt{n-1}\,\dfrac{\overline{X} - \theta}{S} \sim t_{n-1}$

□

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

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

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

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

2. Пусть

   $$C = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \multicolumn{3}{c}{\dotfill} \\ \multicolumn{3}{c}{\dotfill} \end{pmatrix}$$

   Первая строка как вектор имеет норму 1. А значит можно вспомнить процесс Грамма-Шмидта и набрать ортонормированный базис. Тогда C будет

[1] Здесь смещённая дисперсия

ортогональным. Пусть $Y = CX$. Из горы утверждений выше мы получим:

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

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

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

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

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

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

4. Из всего предыдущего

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

■

§ 21 Доверительные 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 γ, если $P(T_1 < \theta < T_2) = \gamma$

Дальше всюду ведутся рассуждения про доверительные интервалы (уровня γ) параметров нормального распределения.

Утверждение 1. *Доверительный интервал для θ при известном σ равен* $\left(\overline{X} - \sigma\,\dfrac{z_{(1+\gamma)/2}}{\sqrt{n}}; \overline{X} + \sigma\,\dfrac{z_{(1+\gamma)/2}}{\sqrt{n}}\right)$

▼
Интервал ищем явно симметричный, так что посчитаем

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

Поскольку величина внутри подчиняется стандартному нормальному распределению,

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

По дороге сделали замену переменной, не пугайтесь. А дальше $z = F_n\left(\frac{1+\gamma}{2}\right)$, 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 θ при неизвестном σ равен* $\left(\overline{X} - S\,\dfrac{t_{n-1,(1+\gamma)/2}}{\sqrt{n}}; \overline{X} + S\,\dfrac{t_{n-1,(1+\gamma)/2}}{\sqrt{n}}\right)$

▼

аналогично 2.21.4, только пользуемся 4 пунктом леммы Фишера 2.20.5.

▲

Утверждение 3. *Доверительный интервал для σ^2 при неизвестном θ равен* $\left(\dfrac{nS^2}{v^2}; \dfrac{nS^2}{u}\right)$. *Чиселки u, v определяются с помощью χ^2.*

Утверждение 4. *Доверительный интервал для σ^2 при неизвестном θ нормально не выражается. Проще численно.*

1. $\dfrac{\sum_{i=1}^{n}(X_i - \theta)^2}{\sigma^2} \sim \chi^2_n$
2. $\dfrac{n\,(\overline{X} - \theta)^2}{\sigma^2} \sim \chi^2_1$

§ 22 Проверка гипотез по параметрам нормального распределения

Будем рассматривать здесь простую гипотезу:

$$H_0\colon \theta = \theta_0$$
$$H_1\colon \theta \neq \theta_0$$

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

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

   из 1 пункта леммы Фишера (2.20.5).

   Рассмотрим α — уровень значимости — какое-нибудь маленькое число. Часто берут 0.05.

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

   Таким образом можно найти границы критической области.

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

2. σ^2 неизвестна. Здесь всё то же самое, только с распределением Стьюдента.

А здесь такую

$$H_0\colon \theta_1 = \theta_2$$
$$H_1\colon \theta_1 \neq \theta_2$$

Будем считать, что X_i, Y_i независимы, и нормально распеделены:

$$X_1, \ldots, X_{n_1} \sim \mathcal{N}(\theta_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \mathcal{N}(\theta_2, \sigma_2^2)$$

1. σ_1^2, σ_2^2 — известны.
$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}}} \sim \mathcal{N}(0,1)$$

2. $\sigma_1^2 = \sigma_2^2$, но не известны
$$\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\overline{X} - \overline{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sim S_{n_1+n_2-2}$$
Это как раз тот самый `t-тест`

3. σ_1^2, σ_2^2 неизвестны. Тут вообще ничего не понятно. (Проблемы Беренса-Фишера)

§ 23 Линейная регрессия

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

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

Здесь x — матрица $n \times m$, $\theta \in \mathbb{R}^m$, $\varepsilon \in \mathbb{R}^n$, $y \in \mathbb{R}^n$

Замечание 1. Можно с тем же успехом написать $Y = y(X) + \varepsilon$, если $\mathrm{M}\,\varepsilon = 0$

Определение 3. Y называется откликом, X — регрессоры (предикторы), ε — шум, θ — параметры.

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

А если расписывать, то получается следущее: если все ε_i независимы и нормально распределены, то функция максимального правдоподобия будет выглядеть так:
$$L(\theta_1, \ldots, \theta_m) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j - \sum_{k=1}^{m} x_{jk}\theta_k\right)^2\right)$$
Как видно, условие максимума такой функции совпадает с минимумом суммы квадратов.

При этом нужны условия Гаусса-Маркова:

1. $X^T X$ обратима

2. $\mathrm{M}\,\varepsilon_i = 0$, $\mathrm{D}\,\varepsilon_i = \sigma^2$, $\operatorname{cov}\varepsilon_i\varepsilon_j = 0$

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

▼
Пусть

$$Q(\theta) = (Y - X\theta)^T(Y - X\theta) = Y^TY - \theta^TX^TY - Y^TX\theta + \theta^TX^TX\theta \in \mathbb{R}$$

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

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

Тогда можно и продифференцировать

$$\frac{\partial Q}{\partial \theta_j} = 2X_{si}X_{sj}\theta_j - 2X_{ji}y_j = 0 \Leftrightarrow 2X^TX\theta - 2X^TY = 0 \Leftrightarrow \hat{\theta} = (X^TX)^{-1}X^TY$$
$$\frac{\partial^2 Q}{\partial \theta_j \partial \theta_i} = X_{si}X_{sj}\delta_{ij}$$

Как видно, там и правда мимимум. 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

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

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

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

Теорема 1. *Пусть*

1. X^TX *обратима*
2. $\mathrm{M}\,\varepsilon_i = 0$, $\mathrm{D}\,\varepsilon_i = \sigma^2$, $\operatorname{cov}\varepsilon_i\varepsilon_j = 0$

Тогда

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

□

1. $\hat{\theta} = (X^TX)^{-1}X^T(X\theta + \varepsilon) = \theta + (X^TX)^{-1}X^T\varepsilon$. Так как $\mathrm{M}\,\varepsilon = 0$, ε не зависит от X, последнее слагаемое обращается в ноль. Ну $\mathrm{M}\,\hat{\theta} = \mathrm{M}\,\theta = \theta$, ибо θ — константа.

2. Пусть $\widetilde{\theta} = HY$. Такая оценка тоже будет несмещённой, если $HX = E$. Ну для несмещённости,

$$\mathrm{M}\, HY = \mathrm{M}\, H(X\theta + \varepsilon) = \mathrm{M}\, HX\theta = \theta$$

а это как раз единичность этой матрицы. Тогда

$$\operatorname{cov} HY = \mathrm{M}(HY-\theta)(HY-\theta)^T = \mathrm{M}(\underbrace{(HX-E)}_{0}\theta + H\varepsilon)\,(\underbrace{(HX-E)}_{0}\theta + H\varepsilon)^T = \mathrm{M}\, H\varepsilon\varepsilon^T H^T =$$

Для изначальной оценки $H_0 = (X^TX)^{-1}X^T$, так что $H_0H_0^T = (X^TX)^{-1}X^T X\,(X^TX)^{-1} = (X^TX)^{-1}$.

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

$$CX = HX - E = 0$$
$$HH^T = (C + (X^TX)^{-1}X^T)\,(C^T + X\,(X^TX)^{-1}) = CC^T + (X^TX)^{-1}$$

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

■

§ 25 Оценка дисперсии погрешностей

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

□

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

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

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

потому что всё, что не на главной диагонали обратится в ноль в силу независимости ε_i.

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

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

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

§ 26 Критерий согласия Пирсона

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

$$H_0\colon F = F_0 \;\; H_1\colon F \neq F_0$$

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 $P(|g| \geqslant C) = \gamma$, $g \sim G$. Тогда критерий введём так: будум отвергать H_0, если $|\rho(X)| \geqslant C$.

Определение 2 (Критерий согласия Пирсона)**.** Разобъем всю область значений X_i на интервалы I_i, с заданными вроятностями — 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 χ^2, и критическая область выбирается из этого распределения.

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

§ 27 Непараметрические критерии

Условия на $X_1, \ldots, X_n$ те же

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

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

Тогда $P(\sqrt{n} D_n^+ < z) \to \begin{cases} \sum\limits_{k\in\mathbb{Z}} (-1)^k e^{-2k^2 z^2}, & z \geqslant 0 \\ 0, & z < 0 \end{cases}$

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

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

Распределение в которому всё сошлось — распределение Колмогорова.

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

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

Утверждение 4 (Критерий Крамера–фон-Мизеса)**.** $\rho = \omega^2 = \int\limits_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 \, \mathrm{d}F_0(x)$

Утверждение 5 (Критерий Андерсона-Дарлинга)**.** $\rho = \int\limits_{-\infty}^{+\infty} \frac{(F_n(x) - F_0(x))^2}{F_n(x)(1 - F_0(x))} \, \mathrm{d}F_0(x)$

Здесь ещё была теорема почему они непараметрические, но я уже немного параметризован.

Глава 3: Случайные процессы

§ 28 Процессы с независимыми приращениями

Определение 1. Случайный процесс — измеримое отображение $X\colon \Omega \to L(T)$, $L(T)$ — 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

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

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:

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

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

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

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

Следствие 1.

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

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

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

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

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

$$\langle X(s), X(t) \rangle = \operatorname{cov}(X(s), X(t)) \tag{3.2}$$

Теперь сделаем из него гильбертово пространство, пополнив по метрике, соотвествующей скалярному произведению Дальше 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 ξ_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. $p_{ij}^{(n)} = P(\xi_{k+n} = j \mid \xi_k = k)$ (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 ξ_i независимы, то они образуют цепь Маркова

Пример 2. Если ξ_i независимы, то $\eta_n \colon \eta_n = f(\eta_{n-1}, \xi_n)$ 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 , ξ_k принимает значения $\pm e_i$ — ортонормированный базис $\mathbb{Z}^d$

Случайные блуждания тоже можно считать цепью Маркова

Определение 6. $f_{ii}^{(n)} = P(\xi_{n+1} = i \mid \xi_n \neq i, \ldots, \xi_1 = i)$. Состояние i возратное, если $\sum_{n=1}^{\infty} f_{ii}^n = 1$ и невозратное 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

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

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 возрата)**.** i *— возратное* $\Leftrightarrow \sum\limits_{n=0}^{\infty} p_{ii}^{(n)} = +\infty$

▼
Из формулы выше

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

а $\sum\limits_{n=1}^{\infty} f_{ii}^{(n)} = F(1) = 1$ в случае возратного состояния.
▲

Определение 7.

Глава A: Обозначения

Ω — пространство элементарных исходов

ω — элемент пространства элементарных исходов

$\mathcal{A}$ — алгебра множеств

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

P — вероятность

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

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

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

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

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

$\square \cdots \blacksquare$ — начало и конец доказательства теоремы

$\blacktriangledown \cdots \blacktriangle$ — начало и конец доказательства более мелкого утверждения

$\ddot{\sim}$ — кривоватая формулировка

`:set aflame` — набирающему зело не нравится билет

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

Глава B: Стандартные распределения

Распределение	**Плотность**
Бернулли	$P(x=1)=p, P(x=0)=0$
Пуассона	$P(x=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}$, $k \in \mathbb{N} \cup 0$
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.