§ 1 Случайные векторы

Определение 1 (Случайный вектор)**.** Случайным вектором назовём произвольное хорошее отображение $X\colon \Omega \to \mathbb{R}^n$. См. примечание про измеримость в **??**. 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. $\lim\limits_{x_i \to -\infty} F(x) = 0$
3. $\lim\limits_{x \to +\infty} F(x) = 1$
4. $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$ *(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.

Определение 3. Векторная случайная величина называется дискретной, если

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

то есть

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

Определение 4. Векторная случайная величина называется непрерывной, если

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

Замечание 1. Для функций распределения:

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

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

▼

По определению функции распределения

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

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.

▲

§ 2 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 $P(Y^i = b^i_j) = P(X^i = a^i_j)$.*

▼

$$P(Y^i = b^i_j) = P(f(X^i) = f(a^i_j)) = P(\omega \mid f(X^i(\omega)) = f(a^i_j))$$

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.

▲

§ 3 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*

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

▼

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

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

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

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.*

▼

Пусть $f(X,Y) = X + Y$

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

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

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

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

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* $\mathrm{M}\,XY = \mathrm{M}\,X\,\mathrm{M}\,Y$

§ 4 Матожидание функции случайной величины

Определение 1 (⍨)**.** 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* $\mathrm{M}\, f_1(X_1) f_2(X_2) = \mathrm{M}\, f_1(X_1)\,\mathrm{M}\, f_2(X_2)$

§ 5 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.

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

▲

§ 6 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 0.6.1

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

▲

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

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

§ 7 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.

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

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

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

А при $n \to \infty$

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

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 0.2.2 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

$$S_n \leftarrow S_n - na$$

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

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

■

§ 8 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:[?], а здесь [?] так другую теорему обозвали

```
:set aflame
```

§ 9 Обобщённая теорема Муавра-Лапласа

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

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

▼

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 $X_1,\dots,X_n$ — 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

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

§ 10 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 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:

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

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)	$\frac{\mathrm{M}(X - \mathrm{M}\,X)^4}{\sigma^4} - 3$		
Функция распределения		$\frac{1}{n}\sum_k I(X_k < y)$	эмпирическая
Квантиль порядка $p \in (0;1)$	$\sup\{x \mid F(x) \leqslant p\}$	$X_{([pn]+1)}$	член вариационного ряда

⚒Свойства оценок

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.

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

В случае чего там векторы, но особо не страшно.

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

§ 11 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 $L(\theta) := \prod\limits_{k=1}^{n} f_\theta(X_k)$ — 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 0.11.1. 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.

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

Утверждение 2. *Плотность распределения T_n ищется по формуле*

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

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 0.2.2 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 $X_1, \dots, X_n$ 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 $\sum_i Z_i = \Sigma$.

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

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

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

[1]Здесь смещённая дисперсия

2. Пусть

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

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:

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

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

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

А дальше надо честно посчитать $\operatorname{cov}\left(\frac{Y_1}{\sqrt{n}}, \sum_{i=2}^n Y_i\right)$. 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. $\dfrac{nS^2}{\sigma^2} = \displaystyle\sum_{i=2}^n \left(\frac{Y_i}{\sigma}\right) = \chi^2_{n-1}$

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

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

■

§ 13 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 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* $\left(\overline{X} - \sigma\,\frac{z_{(1+\gamma)/2}}{\sqrt{n}}; \overline{X} + \sigma\,\frac{z_{(1+\gamma)/2}}{\sqrt{n}}\right)$

▼

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 $z = F_n(\frac{1+\gamma}{2}$, 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 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 0.13.4, только пользуемся 4 пунктом леммы Фишера 0.12.5.

▲

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. $\dfrac{n\,(\overline{X}-\theta)^2}{\sigma^2} \sim \chi^2_1$

§ 14 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 (0.12.5).

   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)

§ 15 Линейная регрессия

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.

Определение 4 (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 5 (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 $n \times m$, $\theta \in R^m$, $\epsilon \in \mathbb{R}^n$, $y \in \mathbb{R}^n$

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

Определение 6. Y называется откликом, X — регрессоры (предикторы), ε — шум, θ — параметры.

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

При этом нужны условия Гаусса-Маркова:

1. $X^T X$ обратима
2. $\mathrm{M}\,\varepsilon_i = 0$ (нет систематической ошибки), $\mathrm{M}\,\varepsilon_i^2 = \sigma^2$, $\mathrm{M}\,\varepsilon_i\varepsilon_j = 0$. Последние 2 равносильны $\operatorname{cov}\varepsilon = \sigma^2 E_n$