

§1 Случайные векторы

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

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

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.

$$2. \lim_{\exists i \ x_i \rightarrow -\infty} F(x) = 0$$

$$3. \lim_{\forall i \ x_i \rightarrow +\infty} F(x) = 1$$

$$4. \lim_{x \rightarrow x_0 - 0} F(x) = F(x_0) \text{ (по совокупности переменных). Это просто следует из непрерывности меры.}$$

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

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

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

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

Замечание 1. Для функций распределения:

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

Утверждение 3. $p(x_1, \dots, x_{n-1}) = \int_{-\infty}^{+\infty} p(x_1, \dots, x_{n-1}, x_n) dx_n$

▼

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

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

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

▲

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

$$\forall \{B_i\} \quad P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

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

▼

По определению функции распределения

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

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.

▲

§ 2 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.



§ 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

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



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

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

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

Переменные независимы, так что можно поменять местами интегралы (ещё же по области интегрируем, неважно как)
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 $f(X, Y) = X + Y$

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

Часть про константу слишком очевидна, не будем её доказывать.

▲

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 $M X Y = M X M Y$

§ 4 Матожидание функции случайной величины

Определение 1 (\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 $M f_1(X_1)f_2(X_2) = M f_1(X_1) M f_2(X_2)$

§ 5 Неравенство Шварца

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



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

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



§ 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 [0.3.3](#)

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



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

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

§7 Суммирование большого числа случайных величин

:set aflame ✖ ~

Теорема 1 (ЦПТ Линдберга-Леви-Агеяна ~). 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 Агеяна $S_n \sim N(na, n\sigma^2)$

□ 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 Агеяна. Используя утверждение 0.2.2 про замену переменной как раз получаем нормальное распределение. Только тут нужно поменять в процессе σ

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

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

Вернёмся обратно к ненулевому a

$$S_n = S_n - na$$

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

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



§ 8 Центральная предельная теорема

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

$$S_n = \sum_i x_i$$

$$a_k = M X_k$$

$$A_n = \sum_{k=1}^n a_k$$

$$\sigma_k^2 = D X_k$$

$$B_n^2 = \sum_{k=1}^n \sigma_k^2$$

$$\gamma_k = M |X_k - a_k|^3$$

$$C_n = \sum_{k=1}^n \gamma_k$$

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 жесь. 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 $\chi_n^2 = \sum_{k=1}^n X_k^2$ имеет распределение χ^2 с n степенями свободы.

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

▼

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

$$\begin{aligned} P(y < X^2 < y + dy) &= P(\sqrt{y} < X < \sqrt{y + dy}) + P(-\sqrt{y} > X > -\sqrt{y + dy}) \\ &= \frac{2}{\sqrt{2\pi}} \int_{\sqrt{y}}^{\sqrt{y+dy}} e^{-u^2/2} du = \sqrt{\frac{2}{\pi}} \frac{e^{-y/2}}{2\sqrt{y}} \end{aligned}$$

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 жесть

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

Из правила отражения для Г-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

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

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

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:

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

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.

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

✂Свойства оценок

1. Несмещённость: $M \hat{\theta} = \theta$
2. Асимптотическая несмещённость: $M \hat{\theta} \xrightarrow{n \rightarrow \infty} \theta$

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 θ как решение уравнения выше.

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

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 $L(\theta) := \prod_{k=1}^n p(X_k, \theta)$ — функция правдоподобия.

Ещё берут её логарифм.

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 0.11.1. 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 записывать производную по параметру по интеграл.

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

§ 12 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 χ^2 с n степенями свободы. Ещё так обозначается: K_n

Утверждение 1. Плотность распределения χ_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 $\frac{X_1}{\sqrt{\frac{1}{n} \sum_i X_i^2}}$ имеет распределение Стьюдента с n степенями свободы.

Утверждение 2. Плотность распределения T_n ищется по формуле

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

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

2. \bar{X}, S^2 независимы

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

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

□

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

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

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

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

2. Пусть

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

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 $\text{cov}\left(\frac{Y_1}{\sqrt{n}}, \sum_{i=2}^n Y_i^2\right)$. 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. Из всего предыдущего

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

■

§ 13 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 $\left(\bar{X} - \sigma \frac{Z_{(1+\gamma)/2}}{\sqrt{n}}; \bar{X} + \sigma \frac{Z_{(1+\gamma)/2}}{\sqrt{n}} \right)$

▼

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 $z = F_n\left(\frac{1+\gamma}{2}\right)$, 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 $\left(\bar{X} - S \frac{t_{n-1, (1+\gamma)/2}}{\sqrt{n}}; \bar{X} + S \frac{t_{n-1, (1+\gamma)/2}}{\sqrt{n}} \right)$

▼

аналогично [0.13.4](#), только пользуемся 4 пунктом леммы Фишера [0.12.5](#).

▲

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

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

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

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

§ 14 Проверка гипотез по параметрам нормального распределения

Будем рассматривать здесь простую гипотезу:

$$H_0: \theta = \theta_0$$

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

1. Пусть $X_1, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, σ^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 (0.12.5).

Рассмотрим α — 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. σ^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. σ_1^2, σ_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. σ_1^2, σ_2^2 неизвестны. Тут вообще ничего не понятно. (Проблемы Беренса-Фишера)

§ 15 Линейная регрессия

Определение 1 (Регрессия). Пусть Y, X_1, \dots, X_m — случайные векторы. Тогда если определено уравнение $y(x_1, \dots, x_m) = M(Y | X_1 = x_1, \dots, X_m = x_m)$, 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 $y(x) = M(Y | X_i = x_i \forall i)$, 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 $M\varepsilon = 0$

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 $\arg \min_{\hat{\theta}} (Y - X\hat{\theta})^T (Y - X\hat{\theta})$

А если расписывать, то получается следующее: если все ε_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. $M\varepsilon_i = 0$, $D\varepsilon_i = \sigma^2$, $\text{cov} \varepsilon_i \varepsilon_j = 0$

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

▼

Пусть

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

В координатах это переписется 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.

▲

§ 16 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. $M \varepsilon_i = 0, D \varepsilon_i = \sigma^2, \text{cov } \varepsilon_i \varepsilon_j = 0$

Тогда

1. $\hat{\theta} = (X^T X)^{-1} X^T Y$ — несмещённая оценка θ ,
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 $M \varepsilon = 0$, ε не зависит от X , последнее слагаемое обращается в ноль. Ну $M \hat{\theta} = M \theta = \theta$, ибо θ — константа.
2. Пусть $\tilde{\theta} = H Y$. Такая оценка тоже будет несмещённой, если $H X = E$. Ну для несмещённости,

$$M H Y = M H (X\theta + \varepsilon) = M H X \theta = \theta$$

а это как раз единичность этой матрицы. Тогда

$$\text{cov } H Y = M(H Y - \theta)(H Y - \theta)^T = M(\underbrace{(H X - E)}_0 \theta + H \varepsilon)(\underbrace{(H X - E)}_0 \theta + H \varepsilon)^T = M H \varepsilon \varepsilon^T H^T = \sigma^2 H H^T$$

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 $C = H - (X^T X)^{-1} X^T$. Тогда

$$C X = H X - E = 0$$

$$H H^T = (C + (X^T X)^{-1} X^T)(C^T + X (X^T X)^{-1}) = C C^T + (X^T X)^{-1}$$

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

■

§ 17 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 σ^2

□

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

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 ε_i .

Осталось доказать, что $\text{Tr}(X(X^T X)^{-1}X^T) = m$. 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 . ■

§ 18 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 $P(|g| \geq C) = \gamma$, $g \sim G$. 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 $\forall i \quad P(X_1 \in I_i) = p_i$. 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 $\rho(X) \xrightarrow[n \rightarrow \infty]{d} \chi_{r-1}^2$.

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

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

Распределение в котором 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). $\rho = \omega^2 = \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 dF_0(x)$

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

Покажем, что эти критерии не зависят от распределения F .

Лемма 6. Пусть F — функция распределения X , F непрерывна, $\exists F^{-1}$. Тогда $F(X) \sim U[0, 1]$.



Посмотрим на распределение $F(X)$. Тут правда применяют распределение к самой случайной величине

$$P(F(X) < x) = P(X < F^{-1}(x)) = F(F^{-1}(x)) = x$$

А это как раз равномерное распределение.



Теорема 7. Пусть F — функция распределения X , F_0 непрерывна и обратима. Тогда при выполнении H_0 статистики не зависят от F_0 .

□ Поскольку гипотеза выполняется, то $F = F_0$.

$$\sup_x |F_n(x) - F_0(x)| = \sup_y |F_n(F_0^{-1}(y)) - y|$$

Из непрерывности F_0

$$F_n(F_0^{-1}(y)) = \frac{1}{n} \#\{k \mid X_k < F_0^{-1}(y)\} = \frac{1}{n} \#\{k \mid F_0(X_k) < y\}$$

Раз X_k независимы, то $F_0(X_k)$ тоже независимы. Тогда по лемме $F_0(X_k)$ — выборка из равномерного распределения. Так что

$$\sup_x |F_n(x) - F_0(x)| = \sup_y \left| \frac{1}{n} \#\{k \mid U(X_k) < y\} - y \right|$$

А значит, не столь важно какая F_0 , ведь изначально предположения на неё были весьма широкие. ■