# Параметрическое оценивание

## Метод максимального правдоподобия и его свойства. Дельта-метод. Случай векторного параметра.

ПМИ ФКН ВШЭ, 22 сентября 2018 г.

Денис Деркач[1]

[1] ФКН ВШЭ

# Оглавление

# Постановка задачи параметрического оценивания

# Статистическая модель

**Определение**

Статистической моделью будем называть набор $\{P_\theta, \theta \in \Theta\}$ для заданной статистической структуры, где $\Theta$ — пространство параметров.

Статистические модели бывают:

1. параметрические;
2. непараметрические;
3. смешанные.

# Параметрическая статистическая модель

**Определение**

Общий вид **параметрической** статистической модели:

$$\mathfrak{F} = \left\{ P_\theta, \theta \in \Theta \subseteq \mathbb{R}^k \right\},$$

где $\Theta$ — пространство параметров,
$\theta = (\theta_1, \ldots, \theta_k)$ — вектор параметров, $k \in \mathbb{N}$.

NB: обычно в статистике параметрическими моделями называют те, которые, в отличие от непараметрических, могут быть описаны конечным набором параметров. В некотором смысле разделение нестрогое.

# Постановка задачи

**Задача параметрического оценивания:** необходимо оценить значение $T(\theta)$, где $T$ — некоторая функция параметра $\theta$.

$$\begin{aligned} T: \quad \Theta &\to \mathcal{Y}, \\ \theta &\mapsto T(\theta). \end{aligned}$$

То есть, необходимо построить **оценку** $\hat{T}$ по имеющейся выборке $X$:

$$\hat{T}: \mathcal{X} \to \hat{\mathcal{Y}}.$$

NB: $\mathcal{Y}$ и $\hat{\mathcal{Y}}$ не обязательно должны совпадать.
NB2: Оценки могут быть детерминированными или рандомизированными.

# Лосс-функция

Для проверки качества оценки необходимо ввести функцию (штраф, лосс, дистанция):

$$\begin{aligned} \ell : \mathcal{Y} \times \hat{\mathcal{Y}} &\to \mathbb{R}, \\ T \times \hat{T} &\mapsto \ell(T, \hat{T}). \end{aligned}$$

Значения этой функции сами являются случайной величиной.

# Пример постановки задачи

**Пример**

Пусть задана выборка $X_1, ..., X_n$ с распределением $\mathcal{N}(\mu, \sigma^2)$. Необходимо оценить значение $\mu$.
В этом случае, $\theta = (\mu, \sigma)$, при этом используется функция $T(\theta) = \mu$, а $\sigma$ — мешающий параметр. Как определить $\hat{T}$? Какие возможны $l$?

## Пример

Допустим, что измерения $X \sim N(\mu, \sigma^2)$ — интегральная характеристика теста по исследованию крови. Необходимо: вычислить $\tau$ — долю наблюдений, для которых характеристика превосходит 1.

## Пример

В этом случае также, $\theta = (\mu, \sigma)$ — параметр из пространства параметров $\Theta = \{(\mu, \sigma) : \mu \in, \sigma > 0\}$.

$$\tau = (X > 1) = 1 - (X < 1) = 1 - \left(\frac{X - \mu}{\sigma} < \frac{1 - \mu}{\sigma}\right) =$$
$$= 1 - \left(Z < \frac{1 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{1 - \mu}{\sigma}\right).$$

$\tau = T(\mu, \sigma) = 1 - \Phi((1 - \mu)/\sigma)$ — параметр, который необходимо оценить, $Z$ — случайная величина со стандартным случайным распределением, а $\Phi$ — нормальное интегральное распределение.

## Пример

Оценить среднее время жизни изотопа.

## Пример

Гамма-распределение обычно используется для моделирования времени жизни. Пусть $X \sim Gamma(\alpha, \beta)$, т. е.

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \text{ где } \alpha, \beta, x > 0.$$

$\Gamma(\alpha)$ — гамма-функция, а $\theta = (\alpha, \beta)$ — параметр.

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

Для того, чтобы оценить среднее время жизни, то необходимо использовать:

$$T(\theta) = \alpha\beta,$$

$$\hat{T}(\theta) = (X).$$

# Некоторые определения

# Метод максимального правдоподобия

## Определение

Пусть задана выборка $X_1, \ldots, X_n \sim F$, при этом у распределения имеется плотность $f(x;\theta)$.

Функция правдоподобия задается формулой:

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta).$$

Логарифмическая функция правдоподобия имеет вид:

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta).$$

Будем рассматривать правдоподобие как функцию параметра $\mathcal{L}_n : \Theta \to [0,\infty)$.

Оценка максимального правдоподобия (ОМП) определяется как такое значение $\widehat{\theta}_n$ параметра $\theta$, которое максимизирует $\mathcal{L}_n(\theta)$.

# Свойства функции правдоподобия:

**Определение**

1. ОМП состоятельная, то есть $\widehat{\theta}_n \to \theta_*$, где $\theta_*$ — реальное значение параметра $\theta$;
2. ОМП не зависит от параметризации, то есть $\widehat{\theta}_n$ — ОМП для $\theta$, тогда $g(\widehat{\theta}_n)$ — ОМП для $g(\theta)$;
3. ОМП асимптотически нормальна: $(\widehat{\theta} - \theta_*)/\widehat{se} \rightsquigarrow \mathcal{N}(0, 1)$;
4. ОМП асимптотически оптимальна или эффективна (при достаточно большом объеме выборки ОМП имеет меньшую дисперсию).

# Свойства функции правдоподобия:

Замечание: вышеприведенные свойства ОМП имеют место, если функция $f(x;\theta)$ достаточно регулярная. В «слишком» сложных случаях ОМП «теряет» эти свойства.

# Информация Фишера:

Пусть $s(X;\theta) = \frac{\partial \log f(X;\theta)}{\partial \theta}$.
Тогда информация Фишера равна:

$$I_n(\theta) = \text{Var}\left(\sum_{i=1}^{n} s(X_i;\theta)\right) = \sum_{i=1}^{n} \text{Var}(s(X_i;\theta)).$$

**Теорема**

Имеет место равенство : $I_n(\theta) = nI(\theta)$, при этом

$$I(\theta) = -\left(\frac{\partial^2 \log f(X;\theta)}{\partial \theta^2}\right) = -\int \left(\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right) f(x;\theta)dx.$$

# Достаточные статистики

**Определение**

Пусть статистика $T(X^n)$ — функция выборки. Будем писать $x^n \leftrightarrow y^n$, если $f(x^n;\theta) = cf(y^n;\theta)$ для некоторой константы $c$, которая может зависеть от $x^n$ и $y^n$, но не от $\theta$.
Статистика $T(x^n)$ называется достаточной, если из того, что $T(x^n) \leftrightarrow T(y^n)$ следует, что $x^n \leftrightarrow y^n$.

Иными словами, статистика $T(X^n)$ достаточная, если мы можем подсчитать функцию правдоподобия, зная только значение $T(X^n)$.

# Экспоненциальное семейство распределений

## Определение

$\{f(x;\theta) : \theta \in \Theta\}$ — однопараметрическое экспоненциальное семейство распределений, если найдутся функции $\eta(\theta),\ B(\theta),\ T(x)$ и $h(x)$, такие, что

$$f(x;\theta) = h(x)e^{\eta(\theta)T(x) - B(\theta)}.$$

### Пример

Пусть $X \sim Poisson(\theta)$, тогда $f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}e^{x\ \log\theta - \theta}$ — принадлежит экспоненциальному распределению при $\eta(\theta) = \log\theta$, $B(\theta) = \theta$ и $T(x) = x, h(x) = 1/x!$.

## Пример

Пусть $X \sim Binomial(n, \theta)$, тогда

$$f(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} =$$

$$= \binom{n}{x} exp\left\{x\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\}.$$

$$\eta(\theta) = \log\left(\frac{\theta}{1-\theta}\right),\ B(\theta) = -n\log(\theta).$$

$$T(x) = x,\ h(x) = \binom{n}{x}.$$

# log-partition function

Экспоненциальное семейство можно записать в виде

$$f(x;\eta) = h(x)e^{\eta T(x) - A(\eta)}, \text{ где}$$

$$A(\eta) = \log \int h(x)e^{\eta T(x)}dx.$$

$A(\eta)$ — выпуклая функция.

**Теорема**

Пусть плотность случайной величины $X$ принадлежит экспоненциальному семейству. Тогда

$$\mathrm{E}\left(T(X)\right) = A'(\eta),\ \mathrm{V}\left(T(X)\right) = A''(\eta).$$

# Случай большой выборки

Пусть $X_1, \dots, X_n$ — $i.i.d.$ выборка из экспоненциального семейства. Тогда рапределение выборки $f(x^n;\theta)$ также будет принадлежать экспоненциальному семейству

$$f(x^n;\theta) = h_n(x^n) e^{\eta(\theta)T_n(x^n) - B_n(\theta)};$$

$$h_n(x^n) = \prod_i h(x_i), T_n(x^n) = \sum_i T(x_i);\ B_n(\theta) = nB(\theta).$$

Таким образом, статистика $\sum_i T(X_i)$ будет достаточной.

## Пример

Пусть $X_1, \ldots, X_n \sim Unif(0, \theta)$. Тогда

$$f(x^n; \theta) = \frac{1}{\theta^n} I(x_{(n)} \le \theta),$$

где $x_{(n)} = \max\{x_1, ..., x_n\}$, I - характеристическая функция. Таким образом, $T(X^n) = \max\{X_1, \ldots, X_n\}$ — достаточная статистика. Так как $T(X^n) \neq \sum_i T(X_i)$, то равномерное распределение не принадлежит экспоненциальному семейству распределений.

Если $\theta = (\theta_1, \ldots, \theta_k)$ — вектор параметров, то плотность $f(x;\theta)$ принадлежит экспоненциальному семейству, если

$$f(x;\theta) = h(x)\exp\left\{\sum_{j=1}^{k}\eta_j(\theta)T_j(x) - B(\theta)\right\}.$$

Аналогично, функция $T = (T_1, \ldots, T_k)$ является достаточной статистикой. Простая выборка объема $n$ также будет иметь плотность из экспоненциального распределения, причем достаточная статистика будет иметь вид $(\sum_i T_1(X_i), \ldots, \sum_i T_k(X_i))$.

## Пример

Рассмотрим нормальную плотность с двухмерным параметром $\theta = (\mu, \sigma)$. В таком случае

$$f(x;\theta) = exp\left\{\frac{\mu}{\sigma^2}x - \frac{x^2}{2\sigma^2} - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\right\}$$

(далее на следующем слайде)

## Пример

(продолжение, начало на предыдущем слайде) Это есть экспоненциальное семейство с:

$$\eta_1(\theta) = \frac{\mu}{\sigma^2},\ T_1(x) = x,$$

$$\eta_2(\theta) = -\frac{1}{2\sigma^2},\ T_2(x) = x^2,$$

$$B(\theta) = \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right),\ h(x) = 1.$$

Статистика $(\sum_i X_i, \sum_i X_i^2)$ будет достаточной.

# Дельта-метод

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim$ Bernoulli$(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim \text{Bernoulli}(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim \text{Bernoulli}(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$
- Каковы шансы на успех?

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim$ Bernoulli$(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$
- Каковы шансы на успех?
- $\frac{p}{1-p}$

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim$ Bernoulli$(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$
- Каковы шансы на успех?
- $\frac{p}{1-p}$
- Сравнить шансы на успех в случае двух распределений Bernoulli$(p)$ и Bernoulli$(r)$?

# Мотивирующий пример

Пусть есть выборка $X_1, ..., X_N \sim$ Bernoulli$(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$
- Каковы шансы на успех?
- $\frac{p}{1-p}$
- Сравнить шансы на успех в случае двух распределений Bernoulli$(p)$ и Bernoulli$(r)$?
- $\frac{p}{1-p} / \frac{r}{1-r}$

# Пример

Пусть есть выборка $X_1, ..., X_N \sim \text{Bernoulli}(p)$. Какие вопросы о параметрах мы обычно хотим задавать?

- Чему равна вероятность успеха?
- $p$
- Каковы шансы на успех?
- $\frac{p}{1-p}$
- Сравнить шансы на успех в случае двух распределений $\text{Bernoulli}(p)$ и $\text{Bernoulli}(r)$.
- $\frac{p}{1-p} / \frac{r}{1-r}$

Обычно мы используем $\hat{p} = \sum_i X_i / N$ для оценки $p$. Кажется, что для других величин мы можем использовать похожие оценки: $\frac{\hat{p}}{1-\hat{p}}$.

Как при этом оценить дисперсию?

# Ряд Тейлора

Пусть $\mathbf{T} = (\mathbf{T_1}, .., \mathbf{T_k})$ – случайные величины со средними $\theta = (\theta_\mathbf{1}, ..., \theta_\mathbf{k})$. Пусть задана дифференцируемая функция $g(\mathbf{T})$ (оценка какого-то параметра). Найти дисперсию этой оценки.

Будем называть $g_i'(\theta) = \frac{\partial}{\partial \mathbf{t_i}}\mathbf{g}(\mathbf{t})\Big|_{\mathbf{t_1}=\theta_\mathbf{1};...;\mathbf{t_k}=\theta_\mathbf{k}}$.

Разложим $g(t)$ в ряд Тейлора:

$$g(t) \approx g(\theta) + \sum_{\mathbf{i=1}}^{\mathbf{k}} \mathbf{g_i'}(\theta)(\mathbf{t_i} - \theta_\mathbf{i})$$

Из этого следует:

$$E_\theta g(T) = g(\theta).$$

# Ряд Тейлора

Аналогично дисперсия:

$$\mathsf{Var}_\theta g(T) \approx E_\theta\left([g(\mathbf{T}) - \mathbf{g}(\theta)]^{\mathbf{2}}\right) \approx E_\theta\left(\left(\sum_{i=1}^{k} g_i'(\theta)(T_i - \theta_i)\right)^2\right) =$$

$$= \sum_{i=1}^{k}[g_i'(\theta]^{\mathbf{2}}\mathsf{Var}_\theta\mathbf{T_i} + \mathbf{2}\sum_{\mathbf{i>j}}\mathbf{g_i'}(\theta)\mathbf{g_j'}(\theta)\mathsf{Cov}_\theta(\mathbf{T_i},\mathbf{T_j}).$$

Замечание: здесь мы не использовали почти никакой информации о функции $g(T)$.

Вернёмся к мотивирующему примеру, нас интересовала дисперсия оценки $\frac{\hat{p}}{1-\hat{p}}$. Здесь $g(p) = \frac{p}{1-p}$ Используя предыдущие выкладки несложно подсчитать, что:

$$\mathsf{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) \approx [g'(p)]^2 \mathsf{Var}(\hat{p}) = \left[\frac{1}{(1-p)^2}\right]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}.$$

## Пример

$X$ - случайная величина, с ненулевым матожиданием $\mu$.
Необходимо оценить матожидание и дисперсию для оценки:
$g(\mu) = 1/\mu$.
Используя:

$$E_\mu\left(g(X)\right) \approx g(\mu),$$

$$\mathsf{Var}_\mu(g(X)) \approx [g'(\mu)]^2 \mathsf{Var}_\mu(X).$$

Получаем:

$$E_\mu\left(\frac{1}{X}\right) \approx \frac{1}{\mu},$$

$$\mathsf{Var}_\mu\left(\frac{1}{X}\right) \approx \frac{1}{\mu^4}\mathsf{Var}_\mu(X).$$

(Продолжение следует)

**Теорема (Теорема Слуцкого)**

Если $X_n \to X$ по распределению и $Y_n \to a$ по вероятности, причём $a = \text{const}$, тогда:

- $Y_n X_n \to aX$ по распределению,
- $X_n + Y_n \to X + a$ по распределению.

## Теорема (Дельта-метод)

Пусть $Y_n$ — последовательность случайных величин для которых $\sqrt{n}[Y_n - \theta] \to \mathcal{N}(0, \sigma^2)$ по распределению. Тогда для дифференцииуемой в $\theta$ функции $g(.)$ с ненулевой производной, $\sqrt{n}[g(Y_n) - g(\theta)] \to \mathcal{N}(0, \sigma^2[g'(\theta)]^2)$ по распределению.

Доказательство: Ряд Тейлора для $g(Y_n)$ около $Y_n = \theta$:

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + o(Y_n - \theta)$$

Третье слагаемое стремится к $0$ по вероятности. Тогда мы сможем применить теорему Слуцкого:

$$[g(Y_n) - g(\theta)] = g'(\theta)(Y_n - \theta),$$

мы получили согласно условиям необходимую сходимость.

NB: В случае нулевой первой производной и ненулевой второй, оценка сходится к $\sigma^2 \frac{g''(\theta)}{2} \chi_1^2$ по распределению.

## Пример

Продолжим предыдущий пример. Пусть есть выборка со средним $\bar{X}$, тогда:

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \to \mathcal{N}\left(0, \left(\frac{1}{\mu}\right)^4 \mathrm{Var}_\mu X_1\right) \text{ по распределению}$$

Если мы не знаем матожидание и дисперсию, можно ввести их оценку:

$$\widehat{\mathrm{Var}}\left(\frac{1}{\bar{X}}\right) \approx \left(\frac{1}{\bar{X}}\right)^4 S^2,$$

то есть

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right)}{\left(\frac{1}{\bar{X}}\right)^2 S} \to \mathcal{N}(0, 1)$$

## Пример

Второй раз применив теорему Слуцкого, получим:

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right)}{\sigma/\mu^2} \to \mathcal{N}(0,1)$$

.

Вспомним центральную предельную теорему!

# Другая формулировка

**Теорема**

Если $\tau = g(\theta)$, где $g$ — дифференцируема и $g'(\theta) \neq 0$, тогда

$$\frac{(\widehat{\tau}_n - \tau)}{\widehat{se}(\widehat{\tau})} \rightsquigarrow \mathcal{N}(0,1),$$

где $\widehat{\tau}_n = g(\widehat{\theta}_n)$ и $\widehat{se}(\widehat{\tau}_n) = |g'(\widehat{\theta})|\widehat{se}(\widehat{\theta}_n)$.
Таким образом, если

$$C_n = (\widehat{\tau}_n - z_{\alpha/2}\widehat{se}(\widehat{\tau}_n), \widehat{\tau}_n + z_{\alpha/2}\widehat{se}(\widehat{\tau}_n)),$$

тогда $(\tau \in C_n) \to 1 - \alpha$ и $n \to \infty$.

## Пример

Пусть $X_1, \ldots, X_n \sim Bernoulli(p)$, $\psi = g(p) = \log(p/(1-p))$.
Информация Фишера равна

$$I(p) = \frac{1}{p(1-p)}.$$

Оценка стандартной ошибки

$$\widehat{se} = \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}.$$

ОМП величины $\psi$

$$\widehat{\psi} = \log \frac{\widehat{p}_n}{1-\widehat{p}_n}.$$

(далее на следующем слайде)

## Пример

(продолжение, начало на предыдущем слайде)

Т.к. $g'(p) = 1/(p(1-p))$, то в соответствии с дельта-методом

$$\widehat{se}(\widehat{\psi}_n) = |g'(\widehat{p}_n)|\widehat{se}(\widehat{p}_n) = \frac{1}{\sqrt{n\widehat{p}_n(1-\widehat{p}_n)}}.$$

Таким образом, границы приближенного 95% доверительного интервала равны

$$\widehat{\psi}_n \pm \frac{2}{\sqrt{n\widehat{p}_n(1-\widehat{p}_n)}}.$$

## Пример

Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Допустим, что $\mu$ известно, а $\sigma$ неизвестно. Необходимо оценить $\psi = \log \sigma$. Логарифм функции правдоподобия

$$\ell(\sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2,$$

значит

$$\widehat{\sigma}_n = \sqrt{\frac{\sum_i (X_i - \mu)^2}{n}}$$

(далее на следующем слайде)

## Пример

(продолжение, начало на предыдущем слайде)

Для подсчета стандартной ошибки необходимо знать информацию Фишера.

$$\log f(X;\sigma) = -\log\sigma - \frac{(X-\mu)^2}{2\sigma^2}$$

$$\frac{\partial^2(\log f(X;\sigma))}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}$$

$$I(\sigma) = -\frac{1}{\sigma^2} + \frac{3\sigma^2}{\sigma^4} = \frac{2}{\sigma^2}$$

(далее на следующем слайде)

## Пример

(продолжение, начало на предыдущем слайде)

$$\widehat{se} = \frac{\widehat{\sigma}_n}{\sqrt{2n}}.$$

Пусть $\psi = g(\sigma) = \log \sigma$, тогда $\widehat{\psi} = \log \widehat{\sigma}_n$. Так как $g' = 1/\sigma$, то

$$\widehat{se}(\widehat{\psi}_n) = \frac{1}{\widehat{\sigma}_n} \frac{\widehat{\sigma}_n}{\sqrt{2n}} = \frac{1}{\sqrt{2n}}.$$

Границы приближенного 95%-ого доверительного интервала равны

$$\widehat{\psi}_n \pm 2/\sqrt{2n}.$$

# Многопараметрический дельта-метод

Постановка задачи параметрического оценивания

Некоторые определения

Экспоненциальное семейство распределений

Дельта-метод

Многопараметрический дельта-метод

### Теорема

Пусть $\mathbf{X_1}, .., \mathbf{X_n}$ выборка случайных векторов, причём $EX_{ij} = \mu_i$ и $\mathrm{Cov}(\mathrm{X_{ik}}, \mathrm{X_{jk}}) = \sigma_{\mathrm{ij}}$. Для функции $g$ с непрывными первыми производными и значения $\mu$, для которого $\tau^2 = \sum\sum \sigma_{ij} \frac{\partial g(\mu)}{\partial \mu_i} \frac{\partial g(\mu)}{\partial \mu_j} > 0$ выполняется:

$$\sqrt{n}[g(\mathbf{\bar{X}_1}, .., \mathbf{\bar{X}_s}) - \mathbf{g}(\mu)] \to \mathcal{N}(\mathbf{0}, \tau^{\mathbf{2}})$$

## Пример

Пусть есть две случайные величины, $X$ и $Y$, с ненулевыми средними. Оценим дисперсию для функции $g(\mu_x, \mu_y) = \frac{\mu_x}{\mu_y}$.

$$\frac{\partial}{\partial \mu_X} g(\mu_X, \mu_Y) = \frac{1}{\mu_Y}$$

$$\frac{\partial}{\partial \mu_Y} g(\mu_X, \mu_Y) = -\frac{\mu_X}{\mu_Y^2}$$

## Пример

Используя разложение Тейлора, получим:

$$E\left(\frac{X}{Y}\right) \approx \frac{\mu_X}{\mu_Y}$$

$$\mathrm{Var}\left(\frac{\mathrm{X}}{\mathrm{Y}}\right) \approx \frac{1}{\mu_{\mathrm{Y}}^2}\mathrm{VarX} + \frac{\mu_{\mathrm{X}}^2}{\mu_{\mathrm{Y}}^4}\mathrm{VarY} - 2\frac{\mu_{\mathrm{X}}}{\mu_{\mathrm{Y}}^3}\mathrm{Cov}(\mathrm{X},\mathrm{Y}) =$$

$$= \frac{\mu_X^2}{\mu_Y^2}\left(\frac{\mathrm{VarX}}{\mu_X^2} + \frac{\mathrm{VarY}}{\mu_Y^2} - 2\frac{\mathrm{Cov}(\mathrm{X},\mathrm{Y})}{\mu_X\mu_Y}\right)$$

Используя довольно простые выкладки, мы получили оценки с хорошей точностью.

Пусть $\tau = g(\theta_1, ..., \theta_k)$ - функция параметра,

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial \theta_1} \\ \vdots \\ \frac{\partial g}{\partial \theta_k} \end{pmatrix}.$$

### Теорема

Допустим, что $\nabla g(\widehat{\theta}) \neq 0$. Пусть $\widehat{\tau} = g(\widehat{\theta})$, тогда

$$\frac{(\widehat{\tau} - \tau)}{\widehat{se}(\widehat{\tau})} \rightsquigarrow \mathcal{N}(0, 1),$$

где $\widehat{se}(\widehat{\tau}) = \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)}$, $\widehat{J}_n = J_n(\widehat{\theta}_n)$, $\widehat{\nabla} g = \nabla g(\theta = \widehat{\theta})$.

## Пример

Пусть $X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$, $\tau = g(\mu, \sigma) = \sigma/\mu$.

Информационная матрица Фишера

$$I_n(\mu, \sigma) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix},$$

$$J_n = I_n^{-1}(\mu, \sigma) = \frac{1}{n} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix}, \ \nabla g = \begin{pmatrix} -\frac{\sigma}{\mu^2} \\ \frac{1}{\mu} \end{pmatrix},$$

$$\widehat{se}(\widehat{\tau}) = \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)} = \frac{1}{\sqrt{n}} \sqrt{\frac{1}{\widehat{\mu}^4} + \frac{\widehat{\sigma}^2}{2\widehat{\mu}^2}}.$$