

# Математическая статистика. Теория

Александр Сергеев

## 1 Введение

Есть генеральная совокупность. Надо выбрать часть генеральной совокупности – выборку. По выборке хотим сделать вывод о всей совокупности

В исследовании есть следующие этапы

1. Сбор данных
2. Препроцессинг (чистка)
3. Построение модели и анализ
4. Интерпретация

### Определение

Репрезентативность – свойство выборки, означающее, что по выборке можно судить по всей совокупности

## 2 Простейшая модель выборки. Эмпирическая функция распределения

### Простейшая модель выборки

$X_1, \dots, X_n$  – i.i.d.

$F$  – функция распределения (теоретическая, мы ее не знаем)

$x_1, \dots, x_n$  – реализация выборки

Глобальная цель – оценить из реализации  $x_1, \dots, x_n$  теоретическую функцию  $F$

**Определение (эмпирическая функция выборки)**

$$\mu_n(t) = \sum_{i=1}^n \mathbb{1}(X_i \leq t)$$

$$F_n(t) = \frac{\mu_n(t)}{n} - \text{эмпирическая функция распределения}$$

**Замечание**

Заметим, что  $\mathbb{1}(X_i \leq t) \sim \text{Bin}(F(t))$

Тогда  $\mu_n(t) \sim \text{Bin}(n, F(t))$

$$P(F_n(x) = \frac{k}{n}) = P(\mu_n(x) = k) = \binom{n}{k} F_n^k(x) (1 - F_n(x))^{n-k}$$

Отсюда  $E(\mu_n(x)) = nF_n(x)$

$E(F_n(t)) = F(t)$  – несмещенность

$$\text{Var } F_n(t) = \frac{F(t)(1 - F(t))}{n}$$

По ЦПТ  $\frac{\mu_n(t) - nF(t)}{\sqrt{F(t)(1 - F(t))n}} = \sqrt{n} \frac{F_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \xrightarrow{d} N(0, 1)$  – асимптотическая нормальность

**Теорема Гливенко – Кантелли**

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

**Теорема Колмагорова**

$$D_n = \sup_x |F_n(x) - F(x)|, F \in C(\mathbb{R}), t \geq 0 \Rightarrow P(\sqrt{n}D_n \leq t) \rightarrow K(t) =$$

$$\sum_{j=-\infty}^{\infty} e^{-2j^2 t^2} - \text{функция распределения Колмагорова}$$

**Теорема Смирнова**

$X_1, \dots, X_n, Y_1, \dots, Y_n$  – независимы

Обе распределены на  $F \in C(\mathbb{R})$

$$D_{m,n} = \sum_x |F_n(x) - F_m(x)|$$

$$\text{Тогда } P(\sqrt{\frac{nm}{n+m}} D_{m,n} \leq t) \xrightarrow{m,n \rightarrow \infty} K(t)$$

### 3 Выборочные моменты

$\alpha_k = EX_1^k$  –  $k$ -ый теоретический момент

$\beta_k = E(X_1 - EX_1)^k$  –  $k$ -ый теоретический момент

$$\overline{g(X)} := \frac{1}{n} \sum_{k=1}^n g(X_k), g(\bullet) : \mathbb{R} \rightarrow \mathbb{R}$$

$$\hat{\alpha}_k = \overline{X^k} = \frac{1}{n} \sum_{j=1}^n X_j^k \text{ – } k\text{-ый выборочный момент}$$

$E\hat{\alpha}_k = \alpha_k$  – несмещенность

$$\text{Var } \hat{\alpha}_k = \frac{1}{n} \text{Var}(X_1^k) = \frac{1}{n} ((EX_1^{2k}) - (EX_1^k)^2)$$

$$\text{По ЦПТ } \sqrt{n} \frac{\hat{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \approx N(0, 1)$$

$$\sqrt{n} \frac{\hat{\alpha}_k - \alpha_k}{\sqrt{\hat{\alpha}_{2k} - \hat{\alpha}_k^2}} = \sqrt{n} \frac{\hat{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \underbrace{\sqrt{\frac{\alpha_{2k} - \alpha_k^2}{\hat{\alpha}_{2k} - \hat{\alpha}_k^2}}}_{\rightarrow 1} \approx N(0, 1)$$

**Пояснение**

$$\hat{\alpha}_k \xrightarrow{P} \alpha_k \text{ (ЗБЧ)}$$

$$\hat{\beta}_k = \overline{(X - \bar{X})^k} = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^k \text{ – } k\text{-ый центральный выборочный мо-}$$

мент

$$\hat{\beta}_2 = S_*^2 \text{ – выборочная дисперсия}$$

$S_*$  – выборочное отклонение

**Замечание**

Выборочные моменты – моменты, посчитанные относительно эмпирического распределения

Тогда для них действуют утверждения, свойственные обычным моментам

$$S_*^2 = \overline{X^2} - (\bar{X})^2$$

$$\hat{\beta}_k = \text{Poly}(\hat{\alpha}_k, \dots, \hat{\alpha}_1)$$

$$\text{Т.к. } \hat{\alpha}_k \xrightarrow{P} \alpha_k, \text{ то } \hat{\beta}_k \xrightarrow{P} \beta_k$$

**Отступление**

Пусть  $\xi_n$  – последовательность случайных векторов и  $\sqrt{n}(\xi_n - \mu) \xrightarrow{d} N(0, \Sigma)$

$\mu$  – какой-то вектор (необязательно матожидание)

$$1. \xi_n \xrightarrow{P} \mu$$

$$\text{Т.к. } (\xi_n - \mu) \frac{\sqrt{n}}{\sqrt{n}} \xrightarrow{d} 0$$

$$2. \text{ Пусть } \phi : \mathbb{R}^m \rightarrow \mathbb{R}, \phi \in C^1(\mathbb{R})$$

$$\phi(\xi_n) \approx \phi(\mu) + \nabla \phi(\mu)(\xi_n - \mu)$$

$$\phi(\xi_n) - \phi(\mu) \approx \nabla \phi(\mu)(\xi_n - \mu)$$

$$\text{Var}(\phi(\xi_n) - \phi(\mu)) \approx \text{Var}(\nabla \phi(\mu)(\xi_n - \mu)) = \nabla \phi(\mu) \text{Var}(\xi_n) (\nabla \phi(\mu))^T$$

$$\text{Тогда } \sqrt{n}(\phi(\xi_n) - \phi(\mu)) \approx \sqrt{n} \nabla \phi(\mu)(\xi_n - \mu) \rightarrow N(0, \nabla \phi(\mu) \Sigma (\nabla \phi(\mu))^T)$$

### Теорема

Пусть  $\xi_n = (\bar{X}, \dots, \bar{X}^k)$

(Многомерная ЦПТ  $\Rightarrow \sqrt{n}(\xi_n - \alpha) \xrightarrow{d} N(0, \sigma), \alpha = (\alpha_1, \dots, \alpha_k), \Sigma = \text{Var}(X_1, \dots, X_1^k)$ )

$\phi : \mathbb{R}^k \rightarrow \mathbb{R}, \phi \in C^1(\mathbb{R})$

$$\sigma^2 = \nabla \phi(\alpha) \Sigma (\nabla \phi(\alpha))^T > 0$$

Тогда  $\sqrt{n} \frac{\phi(\xi_n) - \phi(\alpha)}{\sigma} \xrightarrow{d} N(0, 1)$

Кроме того,  $\sigma = \sigma(\alpha) \in C^1(\mathbb{R}) \Rightarrow \sqrt{n} \frac{\phi(\xi_n) - \phi(\alpha)}{\sigma(\xi_n)} \xrightarrow{d} N(0, 1)$

### Упражнение

$$\sqrt{n} \frac{S_*^2 - \sigma^2}{\sqrt{\hat{\beta}_4 - S_*^4}} \approx N(0, 1)$$

$$ES_*^2 = \frac{n-1}{n} \sigma^2$$

$S^2 := \frac{n}{n-1} S_*^2 = \frac{1}{n-1} \sum_{i=1}^n (X - \bar{X})^2$  – исправленная (несмещенная) дисперсия

### Коэффициент асимметрии

$$\frac{E(X - EX)^3}{\sigma^3}$$

Тогда  $\frac{\hat{\beta}_3}{S_*^3}$  – выборочный коэффициент асимметрии

### Коэффициент эксцесса

$$\frac{E(X - EX)^4}{\sigma^4} - 3 \mapsto \frac{\hat{\beta}_4}{\sigma_*^4} - 3$$

### Ковариация

$$\text{Cov}(X, Y) = EXY - EXEY \mapsto S_{*XY} = \frac{1}{n} \sum_j (X_j - \bar{X})(Y_j - \bar{Y}) = \frac{1}{n} \sum_j X_j Y_j - \bar{X}\bar{Y}$$

#### Корреляция

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X \text{Var } Y}} \mapsto \rho_n = \frac{S_{*XY}}{S_{*X} S_{*Y}}$$

## 4 Порядковые статистики

#### Определение

Вариационный ряд – выборка  $X_{(1)} \leq \dots \leq X_{(m)}$

#### Определение

$X_{(k)}$  –  $k$ -ая порядковая статистика

#### Напоминание

Квантиль порядка  $\alpha$  –  $q_\alpha : P(X \geq q_\alpha) \geq 1 - \alpha, P(X \leq q_\alpha) \geq \alpha$

Если  $F$  – строго монотонная, то  $F(q_\alpha) = \alpha \Leftrightarrow q_\alpha = F^{-1}(\alpha)$

#### Определение

Выборочный квантиль порядка 0 =  $\min X_i$

Выборочный квантиль порядка 1 =  $\max X_i$

Выборочный квантиль порядка  $\alpha \in (0, 1)$ :

$$\exists 0 \leq k \leq n - 1 : \frac{k}{n} \leq \alpha < \frac{k+1}{n}$$

Тогда  $X_{(k)}$  – искомый квантиль

$$\alpha = \frac{1}{4} - \text{первый (нижний) квартиль}$$

$$\alpha = \frac{1}{2} - \text{второй квартиль, выборочная медиана}$$

$$\alpha = \frac{3}{4} - \text{третий (верхний) квартиль}$$

$$\alpha = 1 - \text{четвертый квартиль / максимум}$$

$$n = 2m \Rightarrow \text{med}(X) = \frac{X_{(m)} + X_{(m+1)}}{2}$$

$$n = 2m + 1 \Rightarrow \text{med}(X) = X_{(m+1)}$$

$$P(X_{(k)} \leq t) = P(\mu_n(t) \geq k) = B(F(t), k, n - k + 1)$$

Пусть  $p(t)$  – теоретическая плотность т.е.  $p = F'$

$$P(X_{(k)} \leq t)'_t = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} F^{k-1}(t)(1-F(t))^{n-k} p(t) - \text{плотность } k\text{-ой}$$

порядковой статистики

$$g(x_1, x_2) = \frac{n!}{(k-1)!(r-k-1)!(n-r)!} F^{k-1}(x_1)(F(x_2) - F(x_1))^{r-k-1}(1 - F(x_2))^{n-r} p(x_1)p(x_2) - \text{совместная плотность вектора } (x_{(k)}, x_{(r)}), k < r$$

$$g(x_1, \dots, x_n) = n!p(x_1) \dots p(x_n) - \text{плотность для вектора всех статистик } (X_{(1)}, \dots, X_{(n)})$$

**Определение**

Средний член вариационного ряда  $-X_{(k(n))}, \frac{k(n)}{n} \rightarrow \text{const} \in (0, 1)$

Крайний член вариационного ряда  $-X_{(r)}, r$  — ограничено по  $n$  или  $X_{n+1-s}, s$  — ограничено

**Теорема (об асимптотике среднего члена вариационного ряда)**

Пусть  $0 < \alpha < 1, p$  — теоретическая плотность,  $q_\alpha$  — теоретическая квантиль порядка  $\alpha$

$p \in C^1(U(q_\alpha))$

Тогда  $\sqrt{n}p(q_\alpha) \frac{X_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} N(0, 1)$

**Доказательство**

Комментарии к доказательству в лекции 3, 0:55

1.  $k := \lfloor n\alpha \rfloor$ . Выпишем плотность  $X_{(k)}$

2. Напишем плотность преобразования над  $X_{(k)}$

**Теорема (об асимптотике крайнего члена вариационного ряда)**

Пусть  $r, s$  — фиксированные,  $p$  — плотность

Тогда  $nF(X_{(r)}) \xrightarrow{d} \Gamma(r, 1), nF(X_{(n+1-s)}) \xrightarrow{d} \Gamma(s, 1)$  — независимые

## 5 Точечное оценивание параметров

### 5.1 Постановка задачи точечного оценивания параметра

Пусть  $X_1, \dots, X_n \sim F_\theta, \theta \in \textcircled{\mathbb{H}} \subset \mathbb{R}^d$  — параметр

В классической постановке  $\theta$  — фиксированный неизвестный вектор

Цель: оценить  $\theta$  в виде функции  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  от выборки

**Замечание**

1. Функции от выборки принято называть статистиками

2. Байесовская постановка:  $\theta$  – случайная величина из известного априорного распределения

**Определение (Состоятельность)**

$\hat{\theta}$  – состоятельная оценка  $\theta \Leftrightarrow \hat{\theta} \xrightarrow{P} \theta \Leftrightarrow P(\|\hat{\theta} - \theta\| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$

**Определение (несмещенность)**

$bias(\hat{\theta}) = E\hat{\theta} - \theta$  – смещение

$bias(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta} \forall n$  – несмещенная

$bias(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta}$  при  $n \rightarrow \infty$  – асимптотическая несмещенная

**Определение (асимптотическая нормальность)**

$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{n} N(0, \Sigma_\theta)$

**Определение (эффективность/оптимальность)**

$\hat{\theta}_1$  – эффективнее  $\hat{\theta}_2 \Leftrightarrow MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$ , где  $MSE(\hat{\theta}) = E\|\hat{\theta} - \theta\|^2 = E(\hat{\theta} - \theta)^T(\hat{\theta} - \theta)$

$MSE(\hat{\theta}) = \text{tr}(\text{Var } \hat{\theta}) + \|bias(\hat{\theta})\|^2$

**Доказательство**

$E(\hat{\theta} - \theta)^T(\hat{\theta} - \theta) = E(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^T(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta) = E \underbrace{(\hat{\theta} - E\hat{\theta})^T(\hat{\theta} - E\hat{\theta})}_{\text{Var } \hat{\theta}} + \|bias(\hat{\theta})\|^2$

**Метод моментов**

Рассмотрим  $g_1, \dots, g_d : \exists E g_i(X_1) = m_j(\theta_1, \dots, \theta_d) \xrightarrow{\text{выборочные аналоги}} \overline{g_i(X_1)} =$

$m_i(\hat{\theta}_1, \dots, \hat{\theta}_d)$  – получили уравнения от  $\theta$

Решая уравнения, получаем оценки

Часто берут  $g_i(x) = x^i$  – отсюда метод моментов (но можно брать и другие функции)

1. Асимптотическая нормальность  $\Rightarrow$  состоятельность

**Доказательство**

$$\hat{\theta} - \theta = \frac{1}{\sqrt{n}} \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{P} 0$$

2. Асимптотическая нормальность  $\Rightarrow bias(\hat{\theta}) \rightarrow 0$

**Доказательство**

Пусть  $d = 2$

$$P(|\theta - \hat{\theta}| > \varepsilon) = P\left(\frac{\sqrt{n}|\theta - E\hat{\theta}|}{\sigma} > \frac{\sqrt{n}\varepsilon}{\sigma}\right) = 1 - P\left(\frac{\sqrt{n}|\theta - E\hat{\theta}|}{\sigma} < \frac{\sqrt{n}\varepsilon}{\sigma}\right) \approx 2(1 - \Phi(\frac{\sqrt{n}\varepsilon}{\sigma})) = 2(1 - \Phi(\frac{\sqrt{n}\varepsilon}{\sigma})) \rightarrow 0$$

3. Состоятельность  $\Rightarrow bias(\hat{\theta}) \rightarrow 0$

**Доказательство**

Следует из УЗБЧ

$$\overline{X} \xrightarrow{a.s.} \mu \Rightarrow E\overline{X} \rightarrow \mu$$

4. Пусть  $d = 1, bias\hat{\theta} \rightarrow 0, Var\tilde{\theta} \rightarrow 0 \Rightarrow \hat{\theta}$  – состоятельная

### Замечание

1. Если  $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$  – состоятельная оценка для  $(\dots, \overline{Eg_i(X)}m_i, \dots)$ ,  $\alpha_1, \dots, \alpha_d$  – непрерывные от  $\overline{g_1(X)}, \dots, \overline{g_d(X)}$ , то они состоятельные
2. Если  $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$  – асимптотически нормальные и  $g_1, \dots, g_d$  – гладкие, то каждая оценка асимптотически нормальная

## 6 Метод максимального правдоподобия

$$pmf : p(x, \theta) = p(x|\theta)$$

$$pdf : p(x, \theta) = p(x|\theta)$$

Все это будем называть плотностью

$$X_1, \dots, X_n \sim p(X|\theta)$$

$$L(X|\theta) = \prod_i p(X_i|\theta) \text{ – функция правдоподобия}$$

$$\hat{\theta}_* = \operatorname{argmax} L(X|\theta_i)$$

Предположим, что  $\theta \in B$  – откp.,  $\theta_1 \neq \theta_2 \Rightarrow L(X, \theta_1) \neq L(X, \theta_2)$  Алгоритм

1. Рассмотрим  $\ln L(X, \theta)$
2. Приравняем производную к нулю
3. Найдем максимум

### Пример

$$Poly(1, p), p = (p_1, \dots, p_m)$$

$\nu_1, \dots, \nu_m$  – количество наблюдений типа  $1, \dots, m$

$$L(X, p) = p_1^{\nu_1} \cdot \dots \cdot p_m^{\nu_m}$$



$$\ln L(X, p) = \sum_{j=1}^{m-1} \nu_j \ln p_j + \nu_m \ln(1 - p_1 - \dots - p_{m-1})$$

$$\frac{\partial \ln L}{\partial \ln p_j} = \frac{\nu_j}{p_j} - \frac{\nu_m}{1 - p_1 - \dots - p_{m-1}}$$

Суммируем уравнения:  $\nu_j(1 - \hat{p}_1 - \dots - \hat{p}_{n-1}) = \hat{p}_j \nu_m$

$$\hat{p}_m(n - \nu_m) = \nu_m(1 - \hat{p}_m)$$

$$\hat{p}_m = \frac{\nu_m}{n}$$

$$\text{Аналогично } \hat{p}_j = \frac{\nu_j}{n}$$

**Определение (информация Фишера)**

Для  $d = 1$

$$L(X, \theta) = \prod p(X_j, \theta)$$

$$\ln L(X, \theta) = \sum \ln p(X_j, \theta)$$

$$V(X, \theta) = \frac{\partial \ln L}{\partial \theta} = \sum \frac{\partial \ln p}{\partial \theta} - \text{вклад выборки}$$

Пусть  $\theta \in \textcircled{\mathbb{H}}$  – открыто

$$\theta_1 \neq \theta_2 \Rightarrow p(X, \theta_1) \neq p(X, \theta_2)$$

Регулярность:

$$1. \quad \frac{\partial}{\partial \theta} \int_X T(X) L(X, \theta) = \int_X \frac{\partial}{\partial \theta} L(X, \theta) T(X) \, dx$$

**Необходимое условие**

$\text{supp } P_x$  нне зависит от  $\theta$

$$U[0, \theta] : \int_0^\theta \frac{1}{\theta} \, dt = 1$$

$$\left( \int_0^\theta \frac{1}{\theta} \, dt \right)' = \left( \frac{1}{\theta} \int_0^\theta \, dt \right)' = -\frac{1}{\theta^2} \int_0^\theta \, dt + \frac{1}{\theta} \neq \int_0^\theta \left( \frac{1}{\theta} \right)' \, dt$$

$$2. \quad EV^2(X, \theta) < \infty$$

$$\int_X L(X, \theta) \, dX = 1$$

$$\int_X \frac{\partial L}{\partial \theta} \, dX = \int_X \frac{\frac{L}{\theta}}{L} L \, dX = \int_X V(X, \theta) L(X, \theta) \, dX = EV(X, \theta) = 0$$

$I(\theta) = \text{Var} V(X, \theta) = EV^2(X, \theta)$  – информация Фишера всей выборки

$$V(X, \theta) = \sum_j \frac{\partial \ln \hat{p}(X_j, \theta)}{\partial \theta} = \text{Var } V(X, \theta) = n \underbrace{\text{Var} \frac{\partial \ln p(X_j, \theta)}{\partial \theta}}_{i(\theta) - \text{информация Фишера набора}}$$

$$\begin{aligned}
i(\theta) &= E\left(\frac{\partial \ln p(X_j, \theta)}{\partial \theta}\right)^2 = -E \frac{\partial^2 \ln p}{\partial \theta^2} \\
\frac{\partial}{\partial \theta} \int_{\mathbb{R}} \frac{\partial \ln p(x, \theta)}{\partial \theta} p(x, \theta) dx &= \int_{\mathbb{R}} \frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2} p(x, \theta) dx + \int_{\mathbb{R}} \frac{\partial L}{\partial \theta} \frac{\partial p}{\partial \theta} dx = E \frac{\partial^2 \ln p}{\partial \theta^2} + \\
\underbrace{E\left(\frac{\partial \ln p}{\partial \theta}\right)^2}_{i(\theta)} &= 0
\end{aligned}$$

Для произвольного  $d$

$$\begin{aligned}
i(\theta) &= -\left(E \frac{\partial^2 \ln p(X_1, \theta)}{\partial \theta_i \partial \theta_j}\right)_{1 \leq i, j \leq d} - \text{информационная матрица для 1 набора} \\
I(\theta) &= n \cdot i(\theta)
\end{aligned}$$

Рассмотрим  $N(\theta_1, \theta_2)$

$$\begin{aligned}
p(x, \theta_1, \theta_2) &= \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2}\right) \\
\ln p(x, \theta_1, \theta_2) &= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \theta_2 - \frac{(x - \theta_1)^2}{2\theta_2} \\
\frac{\partial \ln p}{\partial \theta_1} &= \frac{(x - \theta_1)}{\theta_2} \\
\frac{\partial \ln p}{\partial \theta_2} &= -\frac{1}{2\theta_2} + \frac{(x - \theta_1)^2}{2\theta_2^2} \\
\frac{\partial^2 f}{\partial \theta_1^2} &= -\frac{1}{\theta_2} \\
\frac{\partial^2 f}{\partial \theta_2^2} &= \frac{1}{2\theta_2^2} - \frac{(x - \theta_1)^2}{\theta_2^3} \\
\frac{\partial^2 f}{\partial \theta_1 \partial \theta_2} &= -\frac{(x - \theta_1)}{\theta_2^2} \\
i(\theta) &= \begin{pmatrix} \frac{1}{\theta_2} & 0 \\ 0 & \frac{1}{2\theta_2^2} \end{pmatrix}
\end{aligned}$$

По неравенству Рао-Крамера  $\text{Var } \hat{\theta}_1 \geq \frac{\theta_2}{n}$

**Теорема (неравенство Рао-Крамера)**

Пусть модель регулярна,  $d = 1$

$\tau(\theta)$  – оцениваемая функция,  $\tau \in C^1$ ,  $\tau(\theta) \equiv \theta$

$\hat{\tau}(\theta)$  – оценка несмещенная, т.е.  $E(\hat{\tau}(\theta)) = \tau(\theta)$

Тогда  $\text{Var } \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{n \cdot i(\theta)}$

**Доказательство**

$$\tau'(\theta) = \int \hat{\tau}(\theta) \frac{\partial L}{\partial \theta} dx = \frac{\int \tau(\theta) V(X, \theta) L(X, \theta) dx}{E \hat{\tau}(\theta) V(x, \theta)} - EV(X, \theta) E \hat{\tau}(\theta) = \text{Cov}(V(X, \theta), \hat{\tau}(\theta))$$

$$\text{Cov}^2(V(X, \theta), \hat{\tau}(\theta)) \leq \text{Var } V(X, \theta) \text{Var } \hat{\tau}(\theta)$$

**Замечание**

1.  $E \hat{\tau}(\theta) - \tau(\theta) = \text{bias}(\theta) \neq 0$   
 $E \hat{\tau}(\theta) = \tau(\theta) + \text{bias}(\theta)$   
 $\text{Var } \hat{\tau}(\theta) \geq \frac{(\tau'(\theta) + \text{bias}'(\theta))^2}{ni(\theta)}$   
 $MSE(\tau(\theta)) \geq \frac{(\tau'(\theta) + \text{bias}'(\theta))^2}{ni(\theta)} + \text{bias}^2(\theta)$
2.  $\text{Cov}^2(V(X, \theta), \hat{\tau}(\theta)) = \text{Var } V(X, \theta) \text{Var } \hat{\tau}(\theta)$ , если  $\hat{\tau}(\theta) = \alpha(\theta)V(X, \theta) + \tau(\theta)$

**Теорема (неравенство Рао-Крамера)**

Пусть модель регулярна,  $d > 1$

$\tau(\theta) : \mathbb{R}^d \rightarrow \mathbb{R}, \tau \in C^1, \tau(\theta) \equiv \theta$

$\hat{\tau}(\theta)$  – оценка несмещенная, т.е.  $E(\hat{\tau}(\theta)) = \tau(\theta)$

Тогда  $\text{Var } \hat{\tau}(\theta) \geq \frac{\nabla \tau(\theta) i^{-1}(\theta) \nabla^T \tau(\theta)}{n}$

**Свойства оценки максимального правдоподобия** Если существует несмещенная оптимальная оценка в регулярном случае, то она совпадает с оценкой максимального правдоподобия

## 6.1 Состоятельность оценки максимального правдоподобия

Пусть  $\theta_0$  – реальный параметр,  $\theta \neq \theta_0$

Тогда  $P_{\theta_0}(L(X, \theta_0) > L(X, \theta)) \rightarrow 1$

**Доказательство**

$$\frac{L(X, \theta)}{L(X, \theta_0)} < 1$$

$$\frac{1}{n} \sum \ln \frac{p(X_j, \theta)}{p(X_j, \theta_0)} < 0$$

$$\text{По ЗБЧ } E_{\theta_0} \ln \frac{p(X_j, \theta)}{p(X_j, \theta_0)} \leq E_{\theta_0} \left( \frac{p(X_j, \theta)}{p(X_j, \theta_0)} - 1 \right) = \int_X p(X, \theta) dX - \int p(X, \theta_0) dX =$$

0

Пусть  $S_n = \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 - \alpha)\} \cap \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 + \alpha)\}$

$P_{\theta_0}(S_n) \rightarrow 1$

$A_n = \{X : |\hat{\theta} - \theta_0| < \alpha\}$

$B_n = \{X : \left. \frac{\partial \ln L(X, \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0\}$

$S_n \subset A_n B_n \subset A_n$

Отсюда  $P(A_n) \rightarrow 1$  – т.о. оценка состоятельная

## 6.2 Принцип инвариантности правдоподия

$\theta \in \mathbb{H} \xleftrightarrow[\phi]{\text{биекция}} \gamma \in \Gamma$

$\gamma = \Phi(\theta)$

$\theta = \phi^{-1}$

$\sup_{\theta} L(X, \phi(\theta)) = \sup_{\gamma} L(X, \theta)$

### Пример

Дано  $Exp(\lambda)$

Тогда  $\lambda e^{-\lambda x} \mapsto \frac{1}{\bar{X}}$

$\frac{1}{\lambda} e^{-\frac{x}{\lambda}} \mapsto \bar{X}$

### Теорема (асимптотическая нормальность оценки максимально-го правдоподобия)

Пусть модель регулярна

$\left| \frac{\partial^3 \ln F(X, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M(X), EM(X) < \infty$

$\theta_*$  – оценка максимального правдоподия

$\nabla \ln F(X, \theta) = 0$  – имеет единственное решение

Тогда

1.  $\sqrt{n}(\theta_* - \theta) \rightarrow N(0, i^{-1}(\theta))$

2. Если  $\tau(\theta)$  – оцениваемая и  $\in C^1$ , то

$$\sqrt{n}(\tau(\theta_*) - \tau(\theta)) \rightarrow N(0, \theta^2)$$

$$\sigma^2 = \nabla \tau(\theta) i^{-1}(\theta) \nabla \tau(\theta)$$

3. Если  $\sigma^2$  непрерывная от  $\sigma$ , то  $\sqrt{n} \frac{\tau(\sigma_*) - \tau(\theta)}{\sigma(\theta_*)} \rightarrow N(0, 1)$

### Доказательство 1

$$V(X, \theta) = \frac{\partial \ln L(X, \theta)}{\partial \theta}$$

$\theta_0$  – реальный параметр

$$V(X, \theta) = V(X, \theta_0) + V'_\theta(X, \theta_0)(\theta - \theta_0) + V''_\theta(X, \tilde{\theta}) \frac{(\theta - \theta_0)^2}{2}, \tilde{\theta} \in (\theta, \theta_0)$$

Выполним подстановку  $\theta = \theta_*$

$$0 = V(X, \theta_0) + V'_\theta(X, \theta_0)(\theta_* - \theta_0) + V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

$$V'_\theta(X, \theta_0)(\theta_* - \theta_0) = -V(X, \theta_0) - V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

$$\sqrt{n} V'_\theta(X, \theta_0)(\theta_* - \theta_0) = -\sqrt{n} V(X, \theta_0) - \sqrt{n} V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

Заметим, что  $V(X, \theta_0)$  на самом деле представляет сумму независимых одинаково распределенных величин с матожиданием 0 и дисперсией, равной информации Фишера

Тогда по ЦПТ  $-\sqrt{n} V(X, \theta_0) = A_n = N(0, i(\theta))$

$$\sqrt{n} V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2} = n^{\frac{3}{2}} \underbrace{\frac{V''_\theta(X, \tilde{\theta})}{n}}_{\text{ограниченно по ЗБЧ}} \frac{(\theta_* - \theta_0)^2}{2}$$

ограниченно по ЗБЧ

$$V'_\theta(X, \theta_0) = n \frac{V'_\theta(X, \theta_0)}{n} \rightarrow -ni(\theta) \text{ по ЗБЧ}$$

$$\sqrt{n} V'_\theta(X, \theta_0)(\theta_* - \theta_0) \rightarrow N(0, i(\theta))$$

TO BE CONTINUED

Какая-то лажа, смотри <https://t.me/c/2069367863/1/726>

### Определение

\*Вспомним неравенство Рао-Крамера\*

Показатель эффективности  $-\frac{(\tau'(\theta))^2}{n \cdot i(\theta) \text{Var } \hat{\tau}(\theta)} \in [0, 1]$  для регулярных моделей

Если ПЭ = 1, то оценка эффективная

### Определение

Пусть  $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, \underbrace{\frac{\sigma^2}{n}}_{\text{какое-то число}})$

Показатель асимптотической эффективности  $-\frac{1}{i(\theta)\sigma^2}$

### 6.3 Экспоненциальное семейство распределений

Пусть модель регулярна

Если  $p(x, \theta) = \exp(A(\theta)B(x) + C(\theta) + D(x))$ , то распределение относится к регулярному распределению

Примеры:  $N, \Gamma, Pois, Bin, NB$

**Свойства**

$$\ln p(x, \theta) = A(\theta)B(x) + C(\theta) + D(x)$$

$$\frac{\partial \ln p(x, \theta)}{\partial \theta} = A'(\theta)B(x) + C'(\theta)$$

$$V(X, \theta) = A'(\theta) \sum B(X_i) + nC'(\theta)$$

$$V(X, \theta) = n(A'(\theta)\overline{B(X)} + C'(\theta))$$

$$\frac{V(X, \theta)}{n} - C'(\theta) = A'(\theta)\overline{B(X)}$$

$$\overline{B(X)} = \frac{V(X, \theta)}{nA'(\theta)} - \frac{C'(\theta)}{A'(\theta)}$$

Тогда  $\overline{B(X)}$  – оптимальная оценка для  $-\frac{C'(\theta)}{A'(\theta)}$  по критерию оптимальности

### 6.4 Байесовская постановка

$$X_1, \dots, X_n \in F_\theta$$

$\theta$  – неизвестный параметр

$\theta \sim \Pi(\theta)$  – априорное распределение

$l(\hat{\theta}, \theta)$  – функция потерь

$R(\hat{\theta}, \theta) = E_{F_\theta} l(\hat{\theta}, \theta)$  – риск

$r(\hat{\theta}) = E_{\pi(\theta)} R(\hat{\theta}, \theta)$  – байесовский риск

$$\hat{\theta}_B = \operatorname{argmin}_{\hat{\theta}} r(\hat{\theta})$$

$$r(\hat{\theta}) = E_{(\pi(\theta), F_\theta)} l(\hat{\theta}, \theta)$$

**Теорема Байеса для плотностей**

$$p(\theta|X) = \frac{L(X|\theta)\pi(\theta)}{\int L(X|\theta)\pi(\theta) d\theta}$$

**Утверждение**

$$\hat{\theta}_B = \operatorname{argmin}_{\hat{\theta}} E[l(\hat{\theta}, \theta)|X]$$

**Доказательство**

Пусть  $\theta_* = \operatorname{argmin} \dots$

$$r(\theta_*) = EE[l(\theta_*, \theta)|X] \leq EE[l(\hat{\theta}, \theta)|X] = r(\hat{\theta})$$

**Замечание**

$$l(\hat{\theta}, \theta) = (\theta - \hat{\theta})^2 \Rightarrow \hat{\theta}_B = E[\theta|X]$$

## 6.5 Минимаксная оценка

$$m(\hat{\theta}) = \sup_{\theta} R(\hat{\theta}, \theta)$$

$$\hat{\theta}_{wc} = \operatorname{argmin}_{\theta} m(\hat{\theta}) - \text{минимаксная оценка}$$

**Утверждение**

$$r(\hat{\theta}) \leq \mu(\hat{\theta}) - \text{по определению}$$

**Утверждение**

Если  $\exists \pi(\theta)$  – априорное распределение :  $R(\hat{\theta}_B, \theta) \equiv \text{const}$

Тогда  $\hat{\theta}_{wc} = \hat{\theta}_B$

**Доказательство**

Пусть  $\exists \hat{\theta} : m(\hat{\theta}) < m(\theta)$

Тогда  $r(\hat{\theta}) \leq m(\hat{\theta}) < m(\hat{\theta}_B) = r(\hat{\theta}_B)$  – противоречие с определением  $\hat{\theta}_B$

## 6.6 Интервальное оценивание

**Определение (доверительный интервал)**

$$X_1, \dots, X_n \sim F_{\theta}, \theta \in \hat{\mathbb{H}} \subset \mathbb{R}$$

$1 - \alpha = \gamma \in (0, 1)$  – уровень доверия

Рассмотрим  $(T_l(X), T_r(X))$  – доверительный интервал уровня  $\gamma = 1 - \alpha$ ,  
если  $P(\theta \in (T_l(X), T_r(X))) \geq \gamma$

**Классическая схема построения доверительных интервалов**

Пусть  $T(X, \theta) \sim G$  – не зависит от  $\theta$

Рассмотрим  $P(q_1 < T(X, \theta) < q_2) = 1 - \alpha$  – доверительный интервал

$$\text{Потребуем, чтобы } P(T(X, \theta) \leq q_1) = P(T(X, \theta) \geq q_2) = \frac{\alpha}{2}$$

Тогда  $q_1 = q_{\frac{\alpha}{2}}, q_2 = q_{1-\frac{\alpha}{2}}, q_{\bullet}$  – квантили

## 6.7 Доверительные интервалы параметров нормального закона

**Лемма о независимости линейной и квадратической статистик**

$X_1, \dots, X_n \sim N(\theta, \sigma^2) - i.i.d.$

$T = AX; X = (X_1, \dots, X_n)^T, A \in M_{m \times n}(\mathbb{R})$  – линейная статистика

$Q = X^T B X, B \in M_n(\mathbb{R}), B = B^T$  – квадратичная статистика

$AB = 0$

Тогда  $T, Q$  – независимые

**Доказательство**

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m, 0, \dots, 0), \lambda_i \neq 0$  – потенциально нули в конце

$\Lambda - U^T B U$  – в силу симметричности

$U = (u_1, \dots, u_n)$  – собственные вектора, образуют ортонормированный базис

$$B = U \Lambda U^T = \sum_{j=1}^m \lambda_j u_j u_j^T$$

$$Q = \sum_{j=1}^m \lambda_j (X^T u_j)(u_j^T X) = \sum_j \lambda_j (u_j^T X)^2$$

$$A \left( \sum_{j=1}^m \lambda_j u_j u_j^T \right) = 0$$

$$\sum_{j=1}^m \lambda_j A u_j u_j^T = 0$$

Возьмем  $k \in [1, m]$

Домножим справа на  $u_k$

$$\sum_{j=1}^m \lambda_j A u_j \underbrace{u_j^T u_k}_{\mathbb{1}(j=k)} = 0$$

$$\lambda_k A u_k = 0$$

Отсюда  $A u_k = 0$

Рассмотрим вектор  $\begin{pmatrix} u^T \\ A \end{pmatrix} X$  – гауссовский вектор

Проверим, что  $A_i X$  и  $u_k^T X$  – независимые  $\forall i, k$

$$\text{Cov}(A_i X, u_k^T X) = A_i \underbrace{\text{Cov}(X, X^T)}_{\neq 0} u_k = A_i \underbrace{\text{Var } X}_{\sigma^2 \cdot E} u_k = \sigma^2 A_i u_k = 0$$

Т.е. статистики независимые

**Лемма о независимости двух независимых статистик**



$$Q_1 = X^T B_1 X$$

$$Q_2 = X^T B_2 X$$

$$B_1 B_2 = B_2 B_1 = 0$$

Тогда  $Q_1, Q_2$  – независимые

**Доказательство**

Аналогично

**Определение**

$$X_1, \dots, X_n \sim N(0, 1)$$

Тогда  $\sum_{k=1}^n X_k^2 \sim \chi^2(n)$  – распределение хи-квадрат с  $n$  степенями свободы

**Лемма о распределении квадратичной статистики**

$$X_1, \dots, X_n \sim N(0, 1)$$

$$Q = X^T B X, B = B^2$$

$$r := \text{rg}(B)$$

Тогда  $Q \sim \chi^2(r), r = \text{tr}(B)$

**Доказательство**

Заметим, что собственные числа либо 0, либо 1:  $\lambda_{\underbrace{u}_{\text{с.в.}}} = Bu = B^2u =$

$$B\lambda u = \lambda^2 u$$

$$B = \sum_{k=1}^n u_k u_k^T$$

$$Q = \sum_{k=1}^n (u_k^T X)^2$$

$$u_k^T X \sim N(\underbrace{u_k^T E X}_0, \underbrace{u_k^T E u_k}_1)$$

$$\text{Cov}(u_k^T X, u_j^T X) = 0$$

Тогда  $Q \sim \chi(r)$

$$B = U \Lambda U^T$$

$$\text{rg } B = \text{rg } \Lambda = \text{tr } \Lambda$$

$$B_{jj} = \lambda_j u_j u_j^T = \lambda_j$$

Тогда  $\text{tr } B = \text{tr } \Lambda = \text{rg } B$

**Теорема Фишера**

Пусть  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

Тогда

$$1. \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

**Доказательство** очевидно

$$2. \frac{nS_*^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

**Доказательство**

$$Y_j = \frac{X_j - \mu}{\sigma}$$

$$\bar{Y} = \frac{1}{\sigma}(\bar{X} - \mu)$$

$$S_*^2(Y) = \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2 = \frac{S_*^2(X)}{\sigma^2}$$

$$\bar{Y} = \frac{\sum Y_j}{n} = \underbrace{\begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}}_b \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = bY$$

$$nS_*^2(Y) = (Y - bY)^T(Y - bY) = Y^T(E - B)^T(E - B)Y, B = \begin{pmatrix} b \\ \vdots \\ b \end{pmatrix}$$

$$(I - B)^T(I - B) = I - B$$

По предыдущей лемме  $Y^T(E - B)^T(E - B)Y \sim \chi^2(\text{tr}(I - B))$

$$\text{tr}(I - B) = \sum_{k=1}^n \left(1 - \frac{1}{n}\right) = n - 1$$

$$3. S^2, \bar{X} - \text{независимые } S_*^2, \bar{X} - \text{независимые}$$

**Доказательство**

$$b(I - B) = b - b = 0$$

Тогда по лемме 1

**Определение**

$X_0, \dots, X_n - \text{i.i.d } N(0, 1)$

Тогда  $\frac{X_0}{\sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}} \sim T(n)$  – распределение Стьюдента

**Определение**

$$\chi_n^2 \sim \chi^2(n)$$

$$\chi_m^2 \sim \chi^2(m) - \text{независимые}$$

$$\frac{\chi_n^2}{\chi_m^2} \sim F(n, m) - \text{распределение Фишера}$$

## 6.8 Асимптотические доверительные интервалы

$$\lim_{n \rightarrow \infty} P(\theta \in (l_n, r_n)) \geq 1 - \alpha$$

$$T(X, \theta) \xrightarrow[d]{} G, G \text{ не зависит от } \theta$$

## 7 Проверка статистических гипотез

Основное предположение (по умолчанию)

Альтернативное предположение (хотим доказать)

### Определение

Пусть есть выборка в широком смысле  $X_1, \dots, X_n$

Будем считать, что  $(X_1, \dots, X_n) \sim F$

$H_0 := (F \in \mathcal{F}_0)$  – нулевая гипотеза (основная гипотеза)

$H_1 := (F \in \mathcal{F}_1)$  – альтернатива

$\alpha \in (0, 1)$  – уровень значимости

Проводим стат.тест/критерий  $\delta(X, \alpha)$ :

$$\delta(X, \alpha) = \begin{cases} \underbrace{\text{accept } H_0}_{\text{данные не противоречат нулевой гипотезе}} \\ \underbrace{\text{reject } H_0 \text{ with respect to } H_1}_{\text{данные противоречат нулевой и свидетельствуют альтернативной}} \end{cases}$$

Тест не подтверждает и не опровергает гипотезу, но позволяет делать некоторые выводы о гипотезе

Рассмотрим  $T(X)$  – статистику критерия

$$T(X) \sim G \text{ или } T(X) \xrightarrow[n \rightarrow \infty]{d} G \text{ при условии } H_0$$

$$P(T(X) \in T_0(\alpha) | H_0) = 1 - \alpha$$

$$P(T(X) \in T_1(\alpha) | H_0) = \alpha \text{ (при сходимости } \approx)$$

Если  $T(x) \in T_1(\alpha)$  – reject  $H_0$

Иначе – accept  $H_0$

$$\text{supp } G = \underbrace{T_0(\alpha)}_{\text{область принятия}} \sqcup \underbrace{T_1(\alpha)}_{\text{область опровержения}}$$

Виды тестов в зависимости от критической области:

### 1. Left-sided

$$T_0(\alpha) = [q_\alpha, +\infty), T_1(\alpha) = (-\infty, q_\alpha)$$

$$p_l = P(\Gamma \leq T(x) | H_0)$$

2. Right-sided

$$T_0(\alpha) = (-\infty, q_{1-\alpha}], T_1(\alpha) = (q_{1-\alpha}, +\infty) \quad p_r = P(\Gamma > T(x)|H_0)$$

3. Two-sided

$$T_0(\alpha) = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}], T_1 = \overline{T_0} \quad p = 2 \min(p_l, p_r)$$

Тогда мы получили еще одно условие:

Если  $p < \alpha$  – reject  $H_0$

Иначе ассепт  $H_0$

$p$  (p-value) – это максимальный уровень, при котором мы принимаем  $H_0$

Виды ошибок:

1. first type error / false positive

$$P(T(X) \in T_1(\alpha)|H_0) = \alpha$$

2. second type error / false negative

$$P(T(X) \in T_0(\alpha)|H_1) = \beta$$

$1 - \beta$  – мощность критерия

### Пример (spam classifier)

$H_0$  – не спам

$H_1$  – спам

Письма не фильтруются,  $\alpha = 0 \Rightarrow \beta$  – большое, т.к. спама много

Все письма – спам,  $\beta = 0 \Rightarrow \alpha$  – большое

## 7.1 Статистические критерии и доверительные интервалы

Вспомним задачу построения доверительного интервала

$$X_1, \dots, X_n \sim F_\theta, T(X, \theta) \rightarrow G, P(q_{\frac{\alpha}{2}} \leq T(X, \theta) \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Рассмотрим следующую стат. гипотезу

$$H_0 : \theta = \theta_0$$

$$P(T(X, \theta_0) \in T_\theta | \theta = \theta_0) = 1 - \alpha$$

$$H_1 : \theta \neq \theta_0 \text{ или } \theta > \theta_0 \text{ или } \theta < \theta_0$$

## 7.2 Критерий Колмагорова

$$X_1, \dots, X_n \sim F$$

$H_0 : F = F_0, F_0$  – непрерывная

$$H_1 : F \neq F_0$$

**Теорема Колмагорова (напоминание)**

$$P(\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq t) \rightarrow K(t)$$

Статистика критерия:  $D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, F_n$  – э.ф.р.

Если  $D_n > q_{1-\alpha}$  – reject  $H_0$

Иначе ассепт  $H_0$

## 7.3 Критерий Смирнова

$X_1, \dots, X_n, Y_1, \dots, Y_m$  – независимые

$H_0 : F_X = F_Y (= F_0), F_0$  – непрерывная

$$H_1 : \neg H_0$$

$$D_{m,n} = \sqrt{\frac{nm}{n+m}} \sup_x |F_n(x) - F_m(x)|$$

$$T_1(\alpha) = (q_{1-\alpha}, +\infty)$$

## 7.4 Критерии типа хи-квадрат

**Критерий согласия Пирсона**

Пусть есть  $N$ -элементное множество

$p = (p_1, \dots, p_N)$  – настоящий вектор вероятностей (не знаем)

$p_0 = (p_{01}, \dots, p_{0N})$  – ожидаемый, фиксированный вектор вероятностей

$\nu_k$  – кол-во элементов типа  $k$  в выборке,  $n = \sum_k \nu_k$

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

$$\text{Тогда возьмем статистику } \chi_N^2 = \sum_{k=1}^N \frac{(\nu_k - np_{0k})^2}{np_{0k}}$$

**Теорема**

$$\chi_N^2 \xrightarrow[n \rightarrow \infty]{} \chi^2(N-1) \text{ – при условии } H_0$$

**Доказательство для  $N = 2$**

$$\frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(\nu_2 - np_{02})^2}{np_{02}} = \frac{(\nu_1 - np_{01})^2}{n} \left( \frac{1}{p_{01} + \frac{1}{1-p_{01}}} \right) = \frac{(\nu_1 - np_{01})^2}{np_{01}(1-p_{01})} \rightarrow (N(0, 1))^2$$

### Замечание

Критерий состоятельный (мощность стремится к 1)

Критическая область правосторонняя (если  $H_0$  верна, то  $\chi$  будет мало)

Усложним задачу

$$H_0 : p = p_0(\theta), \theta \in \circ H \subset \mathbb{R}^d, d < N - 1$$

$$\text{Тогда } \chi_N^2 = \sum_{k=1}^N \frac{(\nu_k - np_{0k}(\theta))^2}{np_{0k}(\theta)}$$

Вместо  $\theta$  подставим оценку максимального правдоподия

### Теорема

$$p_0(\theta) > 0 \forall \theta$$

$$\frac{\partial p_0}{\partial \theta}, \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} p_0 - \text{непрерывные}$$

$$\text{rg}\left(\frac{\partial p_{0k}}{\partial \theta_j}\right)_{1 \leq k \leq N, 1 \leq j \leq d} = d$$

$$\text{Тогда } \chi_N^2 \rightarrow \chi^2(N - 1 - \underbrace{d}_{\text{количество неизвестных}})$$

### Критерий однородности

Есть  $K$  независимых выборок

Все они из  $\{1 \dots, N\}$

Пусть  $p^{(j)}$  – истинный вектор вероятностей для  $j$ -ой выборки

$H_0 : p^{(1)} = \dots = p^{(k)}$  – мы их не знаем

$H_1 : \neg H_0$

$\nu_{ij}$  – количество элементов  $j$  в выборке  $i$

$$n_i = \sum \nu_{i*}$$

$$n = n_1 + \dots + n_k$$

Пусть  $p^{(*)}$  известны

$$\chi_{n_i}^2 = \sum_j \frac{(\nu_{ij} - n_i p_j^{(i)})^2}{n_i p_j^{(i)}}, df = N - 1$$

$$\chi_n^2 = \sum \chi_{n_i}^2, df = k(N - 1)$$

Теперь  $p^{(*)}$  – неизвестные. Тогда суммарно  $(N - 1)$  неизвестных

Тогда из прошлой теоремы  $df = k(N - 1) - (N - 1) = (N - 1)(k - 1)$

Подставим вместо  $p^{(i)}$  оценку максимального правдоподия  $\hat{p}_j = \frac{\nu_{1j} + \dots + \nu_{kj}}{n}$

### Критерий независимости

$$x_1, \dots, x_n \in \{1, \dots, N\}$$

$$y_1, \dots, y_n \in \{1, \dots, M\}$$

$\nu_{ij}$  – количество пар, в которых первая компонента  $i$ , вторая –  $j$

$$p_{ij} = P(X = i, Y = j)$$

$$p_{X_i} = p(X = i)$$

$$p_{Y_j} = p(Y = j)$$

$$H_0 : p_{ij} = p_{X_i} p_{Y_j}$$

$$H_1 : \neg H_0$$

$$\chi^2 = \sum_{ij} \frac{(\nu_{ij} - p_{ij}n)^2}{np_{ij}}, df = NM - 1$$

Учитывая, что  $p_{ij} = p_{X_i} p_{Y_j}$ ,  $df = NM - 1 - N - M$

### Критерий квантилей

$$H_0 : F(q_1) = \alpha_1, \dots, F(q_N) = \alpha_N, \text{ где } 0 < \alpha_1 < \dots < \alpha_N < 1$$

$$H_1 : \neg H_0$$

Разобьем множество по  $\underbrace{\inf \text{supp } P}_{q_0} < q_1 < \dots < q_N < \underbrace{\sup \text{supp } P}_{q_{N+1}}$

$$\Delta_i = [q_{i-1}, q_i), p(\Delta_i) = q_i - q_{i-1}$$

Далее воспользуемся хи-квадрат

### Критерий знаков

$$H_0 : F(q) = \frac{1}{2}$$

$$H_1 : \neg H_0$$

В общем случае:

Есть парная выборка  $\begin{pmatrix} X_i \\ Y_i \end{pmatrix}_i$

Хотим проверить, что выборки i.i.d.  $\Leftrightarrow F(x, y) = F(x)F(y)$

Сведем к критерию знаков: рассмотрим  $U = X - Y$ . Если  $H_0$ , то  $\text{med } U = 0$

### Замечание

Рассмотрим  $\nu_1$  – количество элементов  $> \text{med}$

$$Z_n = \frac{2}{\sqrt{n}}(\nu_1 - \frac{n}{2}) \rightarrow N(0, 1)$$

С помощью данного факта можно более гибко настроить  $H_1$

### Тест на коэффициент корреляции

Есть парная выборка  $\begin{pmatrix} X_i \\ Y_i \end{pmatrix}_i \sim N(\dots, \dots)$

Тогда  $\frac{\sqrt{n-2}\rho_n}{\sqrt{1-\rho_n^2}} \sim T(n-2)$   
 $H_0 : \rho = 0$   
 $H_1 : \rho \neq 0, \rho > 0, \rho < 0$

## 7.5 Ранговые критерии

### Определение

$X_1, \dots, X_n$  – выборка

$r(X_k)$  – номер в вариационном ряде

### Замечание

Если число повторяется несколько раз, то можно взять

1. минимум
2. максимум
3. среднее арифметическое

$X_1, \dots, X_n, Y_1, \dots, Y_m$  – две независимые выборки

$R_i$  – ранг  $X_i$  в объединенной выборке

$T = R_1 + \dots + R_n$  – статистика Вилкоксона

$Z_{rs} = \mathbb{1}(X_r < Y_s)$

$U = \sum_{r=1}^n \sum_{s=1}^m Z_{rs}$  – статистика Манн-Уитни

$$T + U = nm + \frac{n(n+1)}{2}$$

Пусть хотим проверить, что распределение  $X$  = распределение  $Y$

$$EU = mnE\mathbb{1}(X < Y) = mnP(X < Y) = \frac{mn}{2} \text{ – при условии } H_0$$

$$\text{Т.е. } H_0 : P(X < Y) = \frac{1}{2}$$

$$H_1 : P \neq \frac{1}{2}, P > \frac{1}{2}, P < \frac{1}{2}$$

$$(\text{Статистика нормальная} - U \sim N(\frac{nm}{2}, \frac{nm(n+m+1)}{12}))$$

### Коэффициент корреляции Спирмена

$\begin{pmatrix} X_i \\ Y_i \end{pmatrix}_i$  – парные выборки



$R_i$  – ранг  $X_i$  в своей выборке

$S_i$  – ранг  $Y_i$  в своей выборке

$\rho$  – выборочный коэффициент корреляции между  $R_i$  и  $S_i$  – коэффициент корреляции Спирмена

$H_0 : \rho = 0$

$H_1$  – можем выбирать, т.к.  $\sqrt{n}\rho \rightarrow N(0, 1)$

**Статистический тест Кенделла**

Рассмотрим последовательность  $\begin{pmatrix} R_i \\ S_i \end{pmatrix}_i$

Упорядочим по  $R$ :  $\begin{pmatrix} i \\ T_i \end{pmatrix}_i$

Тогда коэффициент корреляции Кенделла  $\tau = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(T_i - T_j)$

Если  $H_0$ , то  $E\tau = 0$ ,  $\text{Var } \tau = \frac{2(2n+5)}{9n(n-1)}$

Тогда  $\tau \approx N(0, \frac{4}{9n})$

$H_0 : \tau = 0$

$H_1$  – можем настраивать

## 7.6 Проверка независимости наблюдений

$X_1, \dots, X_n$  – выборка из непрерывного распределения

$H_0$  – все наблюдения независимые и пришли из одного и того же закона

Т.е.  $F(x_1, \dots, x_n) = F(x_1) \dots F(x_n)$

**Критерий инверсий**

Рассмотрим вариационный ряд:  $X_{(1)} \leq \dots \leq X_{(n)}$

$\nu_i$  – количество инверсий для элемента  $X_i$

$T = \nu_1 + \dots + \nu_{n-1}$

$ET = \frac{n(n-1)}{4}$

$\text{Var } T = \frac{n(n-1)(2n+5)}{72}$

Статистика  $T$  – асимптотически нормальная

## 8 Линейные статистические модели

$$Y = Xb + \varepsilon$$

$X \in M_{n \times m}(\mathbb{R})$  – матрица переменных (детерминированные)

$X_{ij}$  – количественная переменная

$Y \in \mathbb{R}^n$  – наблюдения зависимой переменной (количественная)

$b \in \mathbb{R}^m$  – вектор коэффициентов (не знаем)

$\varepsilon \in \mathbb{R}^n$  – ошибка (случайная величина)

$$n \gg m$$

*Минимальные предположения:*

1.  $E\varepsilon = 0$
2.  $\text{Var } \varepsilon_i = \sigma^2$  – гомоскедастичность
3.  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$

Цель – «оценить»  $b$  и  $\sigma^2$

**Определение (оценка наименьших квадратов)**

$$\hat{b} = \operatorname{argmin} S^2(b) = \operatorname{argmin} (Xb - Y)^T (Xb - Y)$$

Рассмотрим  $A = X^T X \in M(m \times m)$

Будем предполагать, что  $\operatorname{rg} A = m$  – отсутствие коррелированности у переменных

**Теорема**

$$\hat{b} = A^{-1} X^T Y$$

**Доказательство**

Докажем, что  $\hat{b} = A^{-1} X^T Y$  – оценка наименьших квадратов

$$S^2(\hat{b} + \delta) = (X\hat{b} - Y + X\delta)^T (X\hat{b} - Y + X\delta) = S^2(\hat{b}) + \delta^T X^T X \delta + (X\hat{b} - Y)^T X \delta + \delta^T X^T (X\hat{b} - Y)$$

Подставим вместо  $\hat{b}$  нашу формулу

$$\text{Тогда } S^2(\hat{b} + \delta) = S^2(\hat{b}) + \delta^T X^T X \delta$$

Отсюда  $\hat{b} = A^{-1} X^T Y$  – оценка наименьших квадратов

**Свойства**

$$t = Tb, T \in M_{k \times m}, m \leq m, \operatorname{rg} T = k$$

$$\text{Тогда } \hat{t} = T\hat{b}$$

**Теорема Гаусса-Маркова**

1.  $\hat{t}$  – несмещенная оценка  $t$   
 $\text{Var } \hat{t} = \sigma^2 T A^{-1} T^T$

$$\begin{aligned}\text{Var } \hat{b} &= \sigma^2 A^{-1} \\ MSE(\hat{t}) &= \text{tr Var } \hat{t}\end{aligned}$$

2.  $\hat{t}$  – оптимальная оценка в классе линейных несмещенных оценок

### Доказательство

TODO

$$ES^2(b) = n\sigma^2$$

$$E(\hat{b} - b)^T A(\hat{b} - b) = \sigma^2 m$$

$$\hat{\sigma}^2 = \frac{S^2(\hat{b})}{n - m} - \text{несмещенная оценка для } \sigma^2$$

## 8.1 Условные оценки наименьших квадратов

### Определение

$$Tb = t_0, T \in M_{k \times m}, \text{rg } T = k$$

$$\hat{b}_{T,t_0} = \arg \min_{Tb=t_0} S^2(b) - \text{условная оценка наименьших квадратов}$$

### Теорема

$$\hat{b}_{T,t_0} = \hat{b} - A^{-1} T^T D^{-1} (T\hat{b} - t_0), D = T A^{-1} T^T$$

### Доказательство

$$\text{Можно показать, что } S^2(b) = S^2(\hat{b}_{T,t_0}) + (\hat{b}_{T,t_0} - b)^T A(\hat{b}_{T,t_0} - b)$$

### Замечание

$$S^2(b) = S^2(\hat{b}) + (\hat{b} - b)^T A(\hat{b} - b)$$

$$\text{Подставим } b = \hat{b}_{T,t_0}$$

$$S^2(\hat{b}_{T,t_0}) = S^2(\hat{b}) + (\hat{b} - \hat{b}_{T,t_0})^T A(\hat{b} - \hat{b}_{T,t_0})$$

$$S^2(\hat{b}_{T,t_0}) - S^2(\hat{b}) = (T\hat{b} - t_0)^T D^{-1} (T\hat{b} - t_0)$$

«Обычные» предположения = Минимальные предположения +  $\varepsilon \sim N(0, \sigma^2 E_n)$

$$Y = Xb + \varepsilon \sim N(Xb, \sigma^2 E_n)$$

$$L(b, \sigma^2; Y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \underbrace{(Xb - Y)^T (Xb - Y)}_{S^2(b)}\right)$$

Если  $\sigma^2$  – фиксированное, то  $\sup_b L(b, \sigma^2; Y) \Leftrightarrow \inf_b S^2(b)$

$\hat{b}$  – оценка максимального правдоподобия

Тогда  $\hat{b}$  – эффективно в классе несмещенных оценок

### Основная теорема о линейной регрессии

1.  $S^2(\hat{b}), \hat{b}$  – независимые
2.  $\hat{b} \sim N(b, \sigma^2 A^{-1})$
3.  $\frac{S^2(\hat{b})}{\sigma^2} \sim \chi^2(n - m)$   
 $S^2(b) - S^2(\hat{b}) \sim \chi^2(m)$

**Следствие (Т-тест для регрессии)**

$$\frac{\frac{\hat{b}_j - b_j}{\sigma \sqrt{A_{jj}^{-1}}}}{\sqrt{\frac{1}{n-m} \frac{S^2(\hat{b})}{\sigma^2}}} = \sqrt{n-m} \frac{\hat{b}_j - b_j}{S(\hat{b}) \sqrt{A_{jj}^{-1}}} \sim T(n-m)$$

Отсюда можем проверять гипотезы:  $H_0 : b_j = b_{j0}, H_1 : \neq, >, <$

**F-тест для регрессии**

$H_0 : Tb = t_0$  (default  $T = E_m, t_0 = 0$ , т.е. все  $b_i$  одновременно равны 0)

$H_1 : Tb \neq t_0$

$$F = \frac{n-m}{k} \frac{S^2(\hat{b}_{T,t_0}) - S^2(\hat{b})}{S^2(\hat{b})} = \frac{n-m}{k} \frac{(T\hat{b} - t_0)^T D^{-1} (T\hat{b} - t_0)}{S^2(\hat{b})} \sim F(k, n-m)$$

Рассмотрим вспомогательную задачу:

$$E(U^T B U) = \sum_{i,j} B_{ij} E(U_i, U_j) = \sum_{i,j} B_{ij} \text{Cov}(U_j, U_i) + \sum_{i,j} B_{ij} E U_i E U_j =$$

$$\text{tr}(B \text{Var } U) + (E U)^T B E U$$

$$\text{Возьмем } U = T\hat{b} - t_0, b = D^{-1}, \frac{1}{k} E(U^T B U) = \frac{1}{k} \text{tr} \underbrace{(D^{-1} \text{Var } U)}_{\sigma^2 E} + (T\hat{b} -$$

$$t_0)^T D^{-1} (T\hat{b} - t_0)$$

Тогда если  $T\hat{b} \neq t_0$ , то матожидание числителя будет больше  $\sigma^2$

Тогда F-тест правосторонний

$$R = \text{corr}(Y, \hat{Y}), \hat{Y} = X\hat{b}$$

$R^2$  – коэффициент детерминации

**Замечание**

Если  $y_i = b_0 + b_1 x_1 + \dots + b_m x_m, H_0 : b_1 = \dots = b_m = 0$ ,

$$\text{то } F = (n-m) \frac{R^2}{1-R^2}$$

**Замечание**

$$S^2(\hat{b}) = (1 - R^2) \underbrace{\sum (Y_i - \bar{Y}_i)^2}_{\text{константа в рамках датасета}}$$

## 9 Дисперсионный анализ

Неформально:  $y$  – количественная переменная,  $x$  – категориальные переменные

### 9.1 Однофакторный дисперсионный анализ

**$F$ -тест**

$Y_{ij} = \mu_j + \varepsilon_{ij}$ ,  $\mu_j$  – среднее влияние на  $j$ -ом факторе

$1 \leq j \leq J$  – уровень фактора

$1 \leq i \leq I_j$  –  $i$ -ое наблюдение для фактора с уровнем  $j$

$$I = \sum I_j$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$H_0 : \mu_1 = \dots = \mu_J$$

$$H_1 = \neg H_0$$

$$F = \frac{\frac{S_B^2}{df_B}}{\frac{S_W^2}{df_W}} \sim F(J-1, I-J)$$

$$S_B^2 = \sum_{j=1}^J I_j (\bar{Y}_{*j} - \bar{Y})^2 - \text{межгрупповая дисперсия}$$

$$df_B = J - 1$$

$$S_W^2 = \sum_{j=1}^J \sum_{i=1}^{I_j} (Y_{ij} - \bar{Y}_{*j})^2 - \text{внутригрупповая дисперсия}$$

$$df_W = I - J$$

**Замечание:**  $S^2 = S_W^2 + S_B^2$

Тест правосторонний (как и для регрессии)

ПРОДОЛЖИТЬ НА ЛЕКЦИИ 13