

## std::variant

## Контрольная работа №2. Решения

0. Напишите фамилии хотя бы двух гроссмейстеров, вышедших на данный момент в Турнир Претендентов 2026.

**Решение.** Любые два из перечисленных:

- Фабиано Каруана;
- Жавохир Синдаров;
- Аниш Гири;
- Вэй И;
- Маттиас Блюбаум;
- Андрей Есипенко;
- Рамешбабу Прагнанандха.

1. Докажите, что в модели линейной регрессии

$$X_i = a + b/\sqrt{i} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

МНК-оценка  $\hat{b}$  является состоятельной оценкой  $b$ .

**Решение.** Обозначим  $z_i = 1/\sqrt{i}$ . Ковариационная матрица МНК-оценки равна  $\sigma^2(Z^T Z)^{-1}$ . Матрица признаков  $Z$  состоит из двух столбцов, единичного и столбца  $\mathbf{z}$ . Тогда

$$Z^T Z = \begin{pmatrix} n & \sum z_i \\ \sum z_i & \sum z_i^2 \end{pmatrix}, \quad (Z^T Z)^{-1} = \frac{1}{\det Z^T Z} \begin{pmatrix} \cdots & \cdots \\ \cdots & n \end{pmatrix},$$

откуда  $D\hat{b} = \frac{\sigma^2 n}{n \sum z_i^2 - (\sum z_i)^2}$ . Так как  $\sum z_i^2 \sim \ln n$ ,  $\sum z_i \sim \sqrt{n}$ , то  $D\hat{b} \rightarrow 0$ , что вкупе с несмещённостью даёт требуемое.

2. Пусть  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$  — три независимые выборки размера  $n$ , причём  $X_{ij} \sim \mathcal{N}(\theta_i, 1)$ ,  $i \in \{1, 2, 3\}$ ,  $j \in \{1, \dots, n\}$ . Придумайте состоятельный критерий для проверки

$$H_0: \theta_1 \leq \theta_2 \leq \theta_3$$

против общей альтернативы  $H_1$  уровня значимости  $\alpha$  (не в асимптотическом смысле).

**Решение.** Рассмотрим критерий

$$R = \{\sqrt{n/2} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \geq z_{1-\alpha/2}\} \cup \{\sqrt{n/2} \cdot (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_3) \geq z_{1-\alpha/2}\}.$$

Для первого элемента объединения при верности  $H_0$ :

$$\begin{aligned} P_{\theta}(\sqrt{n}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > z_{1-\alpha/2}) &= P_{\theta}(\sqrt{n/2} \cdot (\mathbf{x}_1 - \theta_1 - \mathbf{x}_2 - \theta_2) - \underbrace{\sqrt{n/2} \cdot (\theta_2 - \theta_1)}_{\geq 0} > z_{1-\alpha/2}) \leq \\ &\leq P_{\theta}(\underbrace{\sqrt{n/2} \cdot (\mathbf{x}_1 - \theta_1 - \mathbf{x}_2 - \theta_2)}_{\sim \mathcal{N}(0,1)} > z_{1-\alpha/2}) = \alpha/2. \end{aligned}$$

Аналогично  $P_{\theta}(\sqrt{n/2} \cdot (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_3) \geq z_{1-\alpha/2}) \leq \alpha/2$ , поэтому  $R$  обладает требуемым уровнем значимости. При неверности  $H_0$  (без потери общности  $\theta_1 > \theta_2$ ) имеем

$$\sqrt{n/2} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \underbrace{\sqrt{n/2} \cdot (\mathbf{x}_1 - \theta_1 - \mathbf{x}_2 - \theta_2)}_{\sim \mathcal{N}(0,1)} + \underbrace{\sqrt{n/2} \cdot (\theta_1 - \theta_2)}_{\rightarrow +\infty} \xrightarrow{P_{\theta}} +\infty,$$

поэтому вероятность попасть в критерий стремится к 1.

**3.** Для выборки  $X_1, \dots, X_n$  посчитайте статистику- $\omega^2$  критерия Крамера-фон Мизеса-Смирнова (а.к.а с весовой функцией  $\psi \equiv 1$ ) для  $P_0 = U\{1, 2, 3\}$  и представьте её предельное распределение при верности нулевой гипотезы в виде  $\|\xi\|^2$ , где  $\xi \sim \mathcal{N}(0, \Sigma)$ . Будет ли критерий, основанный на этой статистике, состоятельным против общем альтернативы?

**Решение.** Пусть  $N_1 = \sum_{i=1}^n I(X_i \leq 1)$ ,  $N_2 = \sum_{i=1}^n I(X_i \in (1, 2])$  и  $N_3 = \sum_{i=1}^n I(X_i \in (2, 3])$ . Тогда  $\hat{F}_n(1) = N_1/n$ ,  $\hat{F}_n(2) = (N_1 + N_2)/n$ ,  $\hat{F}_n(3) = (N_1 + N_2 + N_3)/n$ . Остальные точки имеют  $P_0$ -меру нуль. Значит,

$$n\omega^2 = \frac{n}{3} \left[ \left( \frac{N_1}{n} - \frac{1}{3} \right)^2 + \left( \frac{N_1 + N_2}{n} - \frac{2}{3} \right)^2 + \left( \frac{N_1 + N_2 + N_3}{n} - 1 \right)^2 \right].$$

В случае верности  $H_0$ , очевидно,  $N_1 + N_2 + N_3 = n$ , поэтому статистика принимает вид

$$n\omega^2 = \frac{n}{3} \left[ \left( \frac{N_1}{n} - \frac{1}{3} \right)^2 + \left( \frac{N_3}{n} - \frac{1}{3} \right)^2 \right],$$

что есть треть квадрата длины вектора с левой стороны многомерной ЦПТ:

$$\sqrt{n} \left( \begin{pmatrix} N_1/n \\ N_3/n \end{pmatrix} - \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

где  $\Sigma = \text{cov}(I(X_1 = 1), I(X_3 = 1)) = \begin{pmatrix} 2/9 & -1/9 \\ -1/9 & 2/9 \end{pmatrix}$ .

Возьмём выборку из любого другого распределения с  $F(1) = 1/3$ ,  $F(2) = 2/3$ ,  $F(3) = 1$ . Тогда значения  $\hat{F}_n$  в точках 1, 2 и 3 будут иметь то же распределение, что и в случае  $P_0$ , поэтому и статистика- $\omega^2$  будет иметь то же распределение. Значит, критерий не состоятелен.

**4.** По результатам принятия пересдач по теории вероятностей было поставлено  $A$  отлов,  $B$  хоров и  $C$  удов (в силу специфики принимающих заранее известно, что неудов быть не может). Постройте критерий для проверки на уровне значимости  $\alpha$  гипотезы, что вероятности получить уд/хор/отл образуют арифметическую прогрессию (именно в таком порядке).

**Решение.** Пусть  $p_C$ ,  $p_B$  и  $p_A$  — вероятности получить уд, хор и отл. Они образуют арифм. прогрессию тогда и только тогда, когда  $p_B = (p_A + p_C)/2$ . Так как они в сумме дают 1, то  $p_B = 1/3$ , поэтому все эти вероятности выражаются через один параметр, разность прогрессии:

$$p_C = \frac{1}{3} - \theta, \quad p_B = \frac{1}{3}, \quad p_A = \frac{1}{3} + \theta; \quad \theta \in (-1/3, 1/3).$$

Таким образом,  $\dim \Theta = 3 - 1 = 2$ ,  $\dim \Theta_0 = 1$ . Найдём статистику КОП. Супремум правдоподобия на  $\Theta$  известен, он достигается при  $\hat{p}_A = \frac{A}{A+B+C}$ ,  $\hat{p}_B = \frac{B}{A+B+C}$  и  $\hat{p}_C = \frac{C}{A+B+C}$ , найдём ОМП на  $\Theta_0$ :

$$\log \rho_\theta(\mathbf{x}) = A \log \left( \frac{1}{3} + \theta \right) + B \log \frac{1}{3} + C \log \left( \frac{1}{3} - \theta \right) \rightarrow \max$$

$$\frac{d}{d\theta} \log \rho(\mathbf{x}) = \frac{A}{\frac{1}{3} + \theta} - \frac{C}{\frac{1}{3} - \theta} = 0 \implies \hat{\theta} = \frac{A - C}{3(A + C)}.$$

Таким образом, статистика КОП имеет вид

$$\begin{aligned} T(A, B, C) &= 2(\log \rho_{\hat{p}_A, \hat{p}_B, \hat{p}_C}(\mathbf{x}) - \log \rho_{\hat{\theta}}(\mathbf{x})) = 2A \log \frac{A}{(A + B + C)(1/3 + \hat{\theta})} + \\ &+ 2B \log \frac{B}{(A + B + C)/3} + 2C \log \frac{C}{(A + B + C)(1/3 - \hat{\theta})} = \\ &= 2(A + C) \log \frac{3(A + C)}{2(A + B + C)} + 2B \log \frac{3B}{A + B + C} \xrightarrow{d} \chi_{\Delta}^2, \end{aligned}$$

где  $\Delta = \dim \Theta - \dim \Theta_0 = 1$ . Итого, критерий имеет вид

$$R_{\alpha} = \{(a, b, c) : T(a, b, c) > \chi_{1, 1-\alpha}^2\}.$$

**5.** Целевые величины  $X_i$ ,  $i \in \{1, \dots, n\}$ , хорошо описываются параболой от признака  $z_i$ , поэтому исследователь решил ввести следующую гауссовскую линейную модель:

$$X_i = \alpha_0 + \alpha_1 z_i + \alpha_2 z_i^2 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Он описал гипотезу о форме параболы в виде

$H_0$ : парабола проходит через нуль, а касательная в 1 имеет наклон 1.

Постройте точный критерий уровня значимости  $\alpha = 0.05$  для проверки  $H_0$  и посчитайте его статистику для векторов  $\mathbf{X} = (-9, -5, 2, 7, 5)^T$  и  $\mathbf{z} = (-2, -1, 0, 1, 2)^T$ .

**Решение.** Формально гипотеза записывается как  $H_0: \alpha_0 = 0, \alpha_1 + 2\alpha_2 = 1$ . Критерий Фишера для её проверки имеет вид

$$\left\{ \mathbf{x} : \frac{(RSS_0 - RSS)/m}{RSS/(n - k)} > f_{m, n-k, 1-\alpha} \right\},$$

где  $n = 5$  — количество наблюдений,  $k = 3$  — кол-во параметров в исходной модели,  $m = 2$  — кол-во линейных ограничений на параметры. Осталось найти остаточные суммы квадратов.

Сначала найдём матрицу из скалярных произведений признаков, соответствующим параметрам  $(\alpha_0, \alpha_1, \alpha_2)^T$ . В данном случае она легко обращается, так как она блочная:

$$Z^T Z = \begin{pmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{pmatrix}, \quad (Z^T Z)^{-1} = \begin{pmatrix} 17/35 & 0 & -1/7 \\ 0 & 1/10 & 0 \\ -1/7 & 0 & 1/14 \end{pmatrix},$$

$$\hat{\alpha} = (Z^T Z)^{-1} \cdot Z^T X = (Z^T Z)^{-1} \begin{pmatrix} 0 \\ 40 \\ -14 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix}.$$

Таким образом,

$$RSS = \sum_{i=1}^5 (X_i - \hat{\alpha}_0 - z_i \hat{\alpha}_1 - z_i^2 \hat{\alpha}_2)^2 = 10.$$

В рамках нулевой гипотезы  $X_i - z_i = \alpha_2(z_i^2 - 2z_i) + \varepsilon$ , то есть матрица признаков состоит из единственного столбца  $\mathbf{z}_0 = (8, 3, 0, -1, 0)^T$ . Тогда  $(\mathbf{z}_0^T \mathbf{z}_0)^{-1} = 1/74$ , и

$$\hat{\alpha}_{2,0} = (\mathbf{z}_0^T \mathbf{z}_0)^{-1} \mathbf{z}_0^T (X - \mathbf{z}_i) = -1$$

В таком случае уже  $RSS_0 = 40$ . Итоговая статистика равна

$$\frac{(40 - 10)/2}{10/2} = 3.$$