Математическая статистика. ДЗ 4.

Прохоров Юрий, 776

Задача 1

Пусть T — тестовая статистика для проверки некоторой гипотезы H_0. Доказать, что если распределение $T\big|_{H_0}$ (распределение T при условии истинности H_0) непрерывно, то p-значение имеет равномерное распределение $\mathcal{U}[0,1]$.

Считается, что критическая область имеет вид $\Omega_\alpha^{\text{кр}} = (t_\alpha, +\infty)$.

Решение:

Пусть функция распределения T при истинности H_0 есть $F(x)$. Эта функция непрерывна из условия непрерывности распределения.

Обозначим случайную величину p-value как p. По определению, имеем

$$p = \inf\left\{\alpha \;\middle|\; T \in \Omega_\alpha^{\text{кр.}}\right\} = \int\limits_T^{+\infty} f(t)\,dt = 1 - F(T)$$

Несложно показать, что $F(T) \sim \mathcal{U}[0,1]$ (например, см. ДЗ 2, задачу 1). Тогда

$$p = 1 - F(T) \sim \mathcal{U}[0,1],$$

что и требовалось доказать.

Задача 2 (задача 24 из раздела доп. задачи)

Карта города разбита на $n = 576$ областей. Для каждой области измерено, сколько в нее попало стрелковых снарядов. В таблице приведена эта статистика:

# попаданий, k	0	1	2	3	4	5	6	7
# областей, ν_k	229	211	93	35	7	0	0	1

Исследовать гипотезу о низкой точности стрельбы.

В силу большого количества участков вероятность попадания снаряда в отдельную область мала, значит, при справедливости гипотезы о низкой точности стрельбы можно воспользоваться законом редких событий, согласно которому число попаданий в любую из областей есть (приближенно) пуассоновская с.в. с некоторым общим для всех участков параметром θ. Попадания в разные области независимы.

Решение:

В данной задаче проверяется сложная гипотеза:

$$H_0: \text{ имеет распределение Poiss}(\theta)$$

Проверяем условия применимости критерия χ^2-Пирсона для сложной гипотезы:

- $n \geq 50$ — выполнено ($n = 576$),
- $\nu_k \geq 5$ — не выполнено, поэтому сделаем следующую перегруппировку:

# попаданий, k	0	1	2	3	≥ 4
# областей, ν_k	229	211	93	35	8

- $r - s - 1 \geq 1$ — выполнено (для новой группировки $r = 5$, $s = 1$)

Выберем $r = 5$ областей $\Delta_0, \ldots, \Delta_4$, на которые мы делим множество значений случайной величины $\mathbb{Z}_+$. В нашем случае:

$$\Delta_k = \{k\}, \quad k = \overline{0,3}, \qquad \Delta_4 = \{4, 5, \ldots\}$$

Тогда теоретические вероятности (X — исследуемая с.в.):

$$p_k(\theta) = \mathbb{P}\{X \in \Delta_k\} = \begin{cases} \dfrac{e^{-\theta}\theta^k}{k!} & , \quad k = \overline{0,3} \\ 1 - \displaystyle\sum_{m=0}^{3} \dfrac{e^{-\theta}\theta^m}{m!} & , \quad k = 4 \end{cases}$$

Оценка параметка $\widehat{\theta}$, согласно теореме Фишера-Крамера, находится из уравнения

$$\sum_{k=0}^{4} \frac{\nu_k}{p_k} \frac{dp_k}{d\theta} = 0$$

Вычислим производные:

$$\frac{dp_k}{d\theta} = \begin{cases} p_{k-1} - p_k & , \quad k = \overline{0,3} \\ p_3 & , \quad k = 4 \end{cases}, \qquad p_{-1} = 0$$

Тогда уравнение принимает вид:

$$\sum_{k=0}^{3} \nu_k \left(\frac{k}{\theta} - 1\right) + \nu_4 \frac{e^{-\theta}\frac{\theta^3}{3!}}{1 - \sum_{j=0}^{3} \frac{e^{-\theta}\theta^j}{j!}} = 0$$

Точное решение этого уравнения есть $\widehat{\theta} \approx 0.93001$.

В качестве приближенного решения можно взять оценку выборочного среднего:

$$\widehat{\theta} = \frac{\sum_{k=0}^{4} k\nu_k}{\sum_{k=0}^{4} \nu_k} \approx 0.92708$$

Для этого значения параметра считаем статистику χ^2-Пирсона

$$T(\mathbf{X}; \widehat{\theta}) = \sum_{k=0}^{4} \frac{\left(\nu_k - np_k(\widehat{\theta})\right)^2}{np_k(\widehat{\theta})} \approx 1.03$$

По теореме Фишера-Крамера, $T_n \xrightarrow[n\to\infty]{d} \chi^2(r - s - 1) = \chi^2(3)$. По таблице находим, что p-value $= 0.79$, то есть это означает, что гипотеза H_0 принимается на уровне 0.79.