

Решения задач по математической статистике

КОНСТАНТИН ЗЮБИН

Задача 1

Пусть нулевая гипотеза H_0 состоит в том, что $X_1 \sim R[0, 1]$, а альтернатива H_1 заключается в том, что X_1 имеет функцию распределения $F_{X_1}(x) = \sin(\frac{\pi x}{2})I(0 < x < 1) + I(x \geq 1)$.

Построим наиболее мощный критерий и вычислим мощность.

Теперь $\frac{L_1(x_1)}{L_0(x_1)} = \frac{\pi}{2} \cos(\frac{\pi x_1}{2})$ и неравенство $\frac{\pi}{2} \cos(\frac{\pi x_1}{2}) > c_\alpha$ выполнено тогда и только тогда, когда $x_1 < \frac{2 \arccos(\frac{2c_\alpha}{\pi})}{\pi}$. Поэтому

$$\alpha = P_0\left(\frac{L_1(x_1)}{L_0(x_1)} > c_\alpha\right) = P_0\left(x_1 < \frac{2 \arccos(\frac{2c_\alpha}{\pi})}{\pi}\right) = \frac{2 \arccos(\frac{2c_\alpha}{\pi})}{\pi}.$$

Отсюда $c_\alpha = \frac{\pi}{2} \cos(\frac{\pi \alpha}{2})$. Таким образом, если $\frac{L_1(x_1)}{L_0(x_1)} > \frac{\pi}{2} \cos(\frac{\pi \alpha}{2})$, то нулевая гипотеза отклоняется.

Вычислим мощность критерия.

$$P_1\left(\frac{L_1(x_1)}{L_0(x_1)} > \frac{\pi}{2} \cos(\frac{\pi \alpha}{2})\right) = P_1\left(\frac{\pi}{2} \cos(\frac{\pi x_1}{2}) > \frac{\pi}{2} \cos(\frac{\pi \alpha}{2})\right) = P_1\left(\frac{\pi x_1}{2} < \frac{\pi \alpha}{2}\right) = P_1(x_1 < \alpha) = \sin\left(\frac{\pi \alpha}{2}\right).$$

Задача 2 (почему возникает достаточная статистика)

По критерию факторизации для достаточной статистики T и некоторых функций $g(\theta, t)$ и $h(x_1, \dots, x_n)$ имеется равенство $L(\theta, x_1, \dots, x_n) = g(\theta, t)h(x_1, \dots, x_n)$. При решении задач используется критерий Неймана-Пирсона, в построении которого участвует функция

$$\frac{L(s, X_1, \dots, X_n)}{L(s_0, X_1, \dots, X_n)} = \frac{g(s, T)h(X_1, \dots, X_n)}{g(s_0, T)h(X_1, \dots, X_n)} = \frac{g(s, T)}{g(s_0, T)}.$$

Тогда неравенство $\frac{g(s, T)}{g(s_0, T)} > c_\alpha$ равносильно некоторому неравенству для T .

Задача 2а6 (Exp(1), Exp($\frac{1}{s}$))

Пусть в соответствии с нулевой гипотезой случайные величины X_1, \dots, X_n имеют распределение Exp(1). Альтернативной является гипотеза о распределении Exp($\frac{1}{s}$). Если $X_1, \dots, X_n \sim \text{Exp}(\frac{1}{s})$, то $\bar{X} \sim \Gamma(n, \frac{s}{n})$. Вычислим функции правдоподобия в предположении нулевой гипотезы и альтернативы.

$$L_0(x_1, \dots, x_n) = \prod_{i=1}^n e^{-x_i} I(X_{(1)} > 0);$$
$$L_1(x_1, \dots, x_n) = \prod_{i=1}^n s^{-1} e^{-s^{-1} x_i} I(X_{(1)} > 0).$$

Тогда

$$\frac{L_1(x_1, \dots, x_n)}{L_0(x_1, \dots, x_n)} = s^{-n} e^{-(s^{-1}-1) \sum_{i=1}^n x_i}.$$

Отсюда

$$\alpha = P_0\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_0\left(-(s^{-1} - 1) \sum_{i=1}^n X_i > \ln c_\alpha + n \ln s\right).$$

Если $s < 1$, то

$$\alpha = P_0(\bar{X} < \frac{\ln c_\alpha - n \ln s}{1-s^{-1}}).$$

Поэтому $\frac{\ln c_\alpha - n \ln s}{1-s^{-1}} = \gamma_{n, \frac{1}{n}, \alpha}$, где $\gamma_{n, \frac{1}{n}, \alpha}$ — квантиль уровня α для $\Gamma(n, \frac{1}{n})$ -распределения. Тогда $c_\alpha = s^n e^{(1-s^{-1})\gamma_{n, \frac{1}{n}, \alpha}}$.

Вычислим мощность критерия:

$$\begin{aligned} \beta &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1\left(-(s^{-1} - 1) \sum_{i=1}^n X_i > \ln c_\alpha - n \ln s\right) = P_1\left(\bar{X} < \frac{\ln c_\alpha - n \ln s}{1-s^{-1}}\right) = \\ &= P_1(\bar{X} < \gamma_{n, \frac{1}{n}, \alpha}) = \int_0^{\gamma_{n, \frac{1}{n}, \alpha}} \frac{n^n x^{n-1}}{\Gamma(n)} e^{nx} dx. \end{aligned}$$

Если $s < 1$, то

$$\alpha = P_0(\bar{X} > \frac{\ln c_\alpha - n \ln s}{1-s^{-1}}).$$

Поэтому $\frac{\ln c_\alpha - n \ln s}{1-s^{-1}} = \gamma_{n, \frac{1}{n}, 1-\alpha}$, где $\gamma_{n, \frac{1}{n}, 1-\alpha}$ — квантиль уровня $1 - \alpha$ для $\Gamma(n, \frac{1}{n})$ -распределения. Тогда $c_\alpha = s^n e^{(1-s^{-1})\gamma_{n, \frac{1}{n}, 1-\alpha}}$.

Вычислим мощность критерия:

$$\begin{aligned} \beta &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1\left(-(s^{-1} - 1) \sum_{i=1}^n X_i > \ln c_\alpha - n \ln s\right) = P_1\left(\bar{X} < \frac{\ln c_\alpha - n \ln s}{1-s^{-1}}\right) = \\ &= P_1(\bar{X} > \gamma_{n, \frac{1}{n}, \alpha}) = 1 - \int_0^{\gamma_{n, \frac{1}{n}, 1-\alpha}} \frac{n^n x^{n-1}}{\Gamma(n)} e^{nx} dx. \end{aligned}$$

Вычислим асимптотические квантили. Имеем $E_0(\bar{X}) = 1$, $D_0(\bar{X}) = \frac{1}{n}$. Тогда $\sqrt{n}(\bar{X} - 1) \xrightarrow[n \rightarrow +\infty]{d} \eta \sim \mathcal{N}(0, 1)$. Поэтому

$$P_0(\bar{X} < 1 + \frac{z_\alpha}{\sqrt{n}}) \rightarrow \alpha.$$

Задача 2аб ($\mathcal{N}(0, 1), \mathcal{N}(s, 1)$)

Пусть в соответствии с нулевой гипотезой случайные величины X_1, \dots, X_n имеют распределение $\mathcal{N}(0, 1)$. Альтернативной является гипотеза о распределении $\mathcal{N}(s, 1)$. Если $X_1, \dots, X_n \sim \mathcal{N}(s, 1)$, то $\bar{X} \sim \mathcal{N}(s, \frac{1}{n})$. Вычислим функции правдоподобия в предположении нулевой гипотезы и альтернативы.

$$\begin{aligned} L_0(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}; \\ L_1(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - s)^2}{2}}. \end{aligned}$$

Тогда

$$\frac{L_1(x_1, \dots, x_n)}{L_0(x_1, \dots, x_n)} = e^{\frac{2s \sum_{i=1}^n x_i - ns^2}{2}}.$$

Отсюда

$$\beta = P_0\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_0(2s\bar{X} - s^2 > \frac{2 \ln c_\alpha}{n}) = P_0(2s\bar{X} > s^2 + \frac{2 \ln c_\alpha}{n}).$$

Если $s > 0$, то

$$\alpha = P_0(\bar{X} > \frac{s}{2} + \frac{\ln c_\alpha}{sn}).$$

Поэтому $\frac{s}{2} + \frac{\ln c_\alpha}{sn} = \frac{z_{1-\alpha}}{\sqrt{n}}$, где z_α — квантиль уровня α для $\mathcal{N}(0, 1)$ -распределения. Тогда $c_\alpha = e^{\sqrt{ns}z_{1-\alpha} - \frac{ns^2}{2}}$.

Вычислим мощность критерия:

$$\begin{aligned}\beta(s) &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1(2s\bar{X} - s^2 > \frac{2\ln c_\alpha}{n}) = P_1(2s\bar{X} > s^2 + \frac{2\ln c_\alpha}{n}) = \\ &= P_1(\bar{X} > \frac{s}{2} + \frac{\ln c_\alpha}{sn}) = 1 - \Phi(\sqrt{n}(\frac{z_{1-\alpha}}{\sqrt{n}} - s)) = 1 - \Phi(z_{1-\alpha} - \sqrt{ns}).\end{aligned}$$

Если $s < 0$, то

$$\alpha = P_0(\bar{X} < \frac{s}{2} + \frac{\ln c_\alpha}{sn}).$$

Поэтому $\frac{s}{2} + \frac{\ln c_\alpha}{sn} = \frac{z_\alpha}{\sqrt{n}}$, где z_α — квантиль уровня α для $\mathcal{N}(0, 1)$ -распределения. Тогда $c_\alpha = e^{\sqrt{ns}z_\alpha - \frac{ns^2}{2}}$.

Вычислим мощность критерия:

$$\begin{aligned}\beta &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1(2s\bar{X} - s^2 > \frac{2\ln c_\alpha}{n}) = P_1(2s\bar{X} > s^2 + \frac{2\ln c_\alpha}{n}) = \\ &= P_1(\bar{X} < \frac{s}{2} + \frac{\ln c_\alpha}{sn}) = \Phi(\sqrt{n}(\frac{z_\alpha}{\sqrt{n}} - s)) = \Phi(z_\alpha - \sqrt{ns}).\end{aligned}$$

Вычислим асимптотические квантили. Имеем $E_0(\bar{X}) = 0$, $D_0(\bar{X}) = \frac{1}{n}$. Тогда $\sqrt{n}\bar{X} \xrightarrow[n \rightarrow +\infty]{d} \eta \sim \mathcal{N}(0, 1)$. Поэтому

$$P_0(\bar{X} < \frac{z_\alpha}{\sqrt{n}}) \rightarrow \alpha.$$

Задача 2а6 (Poiss(1), Poiss(s))

Пусть в соответствии с нулевой гипотезой случайные величины X_1, \dots, X_n имеют распределение Poiss(1). Если $X_1, \dots, X_n \sim \text{Poiss}(s)$, то $n\bar{X} \sim \text{Poiss}(ns)$. Альтернативной является гипотеза о распределении Poiss(s). Вычислим функции правдоподобия в предположении нулевой гипотезы и альтернативы.

$$\begin{aligned}L_0(x_1, \dots, x_n) &= P_\theta(X_1 = x_1, \dots, X_n = x_n) = e^{-n} \frac{1}{x_1! \dots x_n!}; \\ L_1(s, x_1, \dots, x_n) &= e^{-ns} \frac{1}{x_1! \dots x_n!} s^{(x_1 + x_2 + \dots + x_n)}.\end{aligned}$$

Тогда

$$\frac{L_1(x_1, \dots, x_n)}{L_0(x_1, \dots, x_n)} = e^{n(1-s)} s^{x_1 + x_2 + \dots + x_n}.$$

Отсюда

$$\alpha = P_0\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_0(n(1-s) + n\bar{X} \ln s < c_\alpha) = P_0(n\bar{X} \ln s < c_\alpha + n(s-1)).$$

Если $s > 1$, то

$$\alpha = P_0(n\bar{X} < \frac{c_\alpha + n(s-1)}{\ln s}).$$

Поэтому $\frac{c_\alpha + n(s-1)}{\ln s} = p_{s,\alpha}$, где $p_{s,\alpha}$ — квантиль уровня α для распределения Poiss(n). Тогда $c_\alpha = p_{s,\alpha} \ln s - n(s-1)$.

Вычислим мощность критерия:

$$\begin{aligned}\beta &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1(n(1-s) + n\bar{X} \ln s < c_\alpha) = P_1(n\bar{X} \ln s < c_\alpha + n(s-1)) = \\ &= P_1(n\bar{X} < \frac{c_\alpha + n(s-1)}{\ln s}) = 1 - F_{ns}(p_{s,\alpha}),\end{aligned}$$

где F_{ns} — функция распределения для случайной величины с распределением Poiss(ns).

Если $s < 1$, то

$$\alpha = P_0(n\bar{X} > \frac{c_\alpha + n(s-1)}{\ln s}).$$

Поэтому $\frac{c_\alpha + n(s-1)}{\ln s} = p_{n,1-\alpha}$, где $p_{n,\alpha}$ — квантиль уровня α для распределения Poiss(n). Тогда $c_\alpha = p_{s,1-\alpha} \ln s - n(s-1)$.

Вычислим мощность критерия:

$$\begin{aligned}\beta &= P_1\left(\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} > c_\alpha\right) = P_1(n(1-s) + n\bar{X} \ln s < c_\alpha) = P_1(n\bar{X} \ln s < c_\alpha + n(s-1)) = \\ &= P_1(n\bar{X} > \frac{c_\alpha + n(s-1)}{\ln s}) = 1 - F_{ns}(p_{s, 1-\alpha}),\end{aligned}$$

где F_{ns} — функция распределения для случайной величины с распределением $\text{Poiss}(ns)$.

Вычислим асимптотические квантили. Имеем $E_0(n\bar{X}) = n$, $D_0(n\bar{X}) = n$. Тогда $\sqrt{n}(\bar{X} - n) \xrightarrow[n \rightarrow +\infty]{d} \eta \sim \mathcal{N}(0, 1)$. Поэтому

$$P_0(\bar{X} < n + \frac{z_\alpha}{\sqrt{n}}) \rightarrow \alpha.$$

Задача 3аб

Пусть $X_1, \dots, X_n \sim \mathcal{N}(\theta_1, \theta_2)$.

Пусть нулевая гипотеза состоит в том, что $\theta_1 = 0$, а альтернатива заключается в том, что $\theta_1 \neq 0$. Построим критерий на основе $g(\theta_1, X_1, \dots, X_n) = \frac{\bar{X} - \theta_1}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}}$. Тогда при фиксиро-

ванном θ_1 случайная величина $g\sqrt{n}$ имеет распределение t_{n-1} . Тогда

$$P_0(q_{n-1, \frac{\alpha}{2}} < g\sqrt{n} < q_{n-1, 1-\frac{\alpha}{2}}) = \alpha$$

и

$$P_0\left(\frac{q_{n-1, \frac{\alpha}{2}}}{\sqrt{n}} < g < \frac{q_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}}\right) = \alpha.$$

Критерий уровня $1 - \alpha$ будет состоять в отклонении нулевой гипотезы при $|g(0, X_1, \dots, X_n)| > \frac{t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}}$.

Пусть теперь нулевая гипотеза состоит в том, что $\theta_2 = 1$, а альтернатива утверждает обратное. Построим критерий на основе $h(X_1, \dots, X_n) = \sum_{i=1}^n (X_i - \bar{X})^2$. Тогда при фиксированном θ_2 случайная величина $\frac{h}{(n-1)\theta_2}$ имеет распределение χ_{n-1}^2 . В предположении $\theta_2 = 1$ имеем

$$P_0\left(\frac{h}{n-1} > c_{n-1, \alpha}\right) = 1 - \alpha.$$

Критерий уровня α будет состоять в отклонении нулевой гипотезы при $h(0, X_1, \dots, X_n) > (n-1)c_{n-1, \alpha}$.