Решения задач по математической статистике

КОНСТАНТИН ЗЮБИН

Задача 1

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

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

$$\alpha = \mathrm{P}_0(\tfrac{L_1(x_1)}{L_0(x_1)} > c_\alpha) = \mathrm{P}_0(x_1 < \tfrac{2\arccos(\frac{2c_\alpha}{\pi})}{\pi}) = \tfrac{2\arccos(\frac{2c_\alpha}{\pi})}{\pi}.$$

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.

$$\mathrm{P}_1(\tfrac{L_1(x_1)}{L_0(x_1)} > \tfrac{\pi}{2}\cos(\tfrac{\pi\alpha}{2})) = \mathrm{P}_1(\tfrac{\pi}{2}\cos(\tfrac{\pi x_1}{2} > \tfrac{\pi}{2}\cos(\tfrac{\pi\alpha}{2})) = \mathrm{P}_1(\tfrac{\pi x_1}{2} < \tfrac{\pi\alpha}{2}) = \mathrm{P}_1(x_1 < \alpha) = \sin(\tfrac{\pi\alpha}{2}).$$

Задача 2 (почему возникает достаточная статистика)

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

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

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

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

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

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

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

Тогда

$$\tfrac{L_1(x_1, \ldots, x_n)}{L_0(x_1, \ldots, x_n)} = s^{-n} e^{-(s^{-1}-1)\sum\limits_{i=1}^n x_i}.$$

Отсюда

$$\alpha = \mathrm{P}_0(\tfrac{L_1(X_1, \ldots, X_n)}{L_0(X_1, \ldots, X_n)} > c_\alpha) = \mathrm{P}_0(-(s^{-1}-1)\sum_{i=1}^n X_i > \ln c_\alpha + n\ln s).$$

Если $s < 1$, то

$$\alpha = \mathrm{P}_0(\overline{X} < \tfrac{\ln c_\alpha - n\ln s}{1-s^{-1}}).$$

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:

$$\beta = \mathrm{P}_1(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)} > c_\alpha) = \mathrm{P}_1(-(s^{-1}-1)\sum_{i=1}^n X_i > \ln c_\alpha - n\ln s) = \mathrm{P}_1(\overline{X} < \tfrac{\ln c_\alpha - n\ln s}{1-s^{-1}}) =$$

$$= \mathrm{P}_1(\overline{X} < \gamma_{n,\frac{1}{n},\alpha}) = \int\limits_0^{\gamma_{n,\frac{1}{n},\alpha}} \tfrac{n^n x^{n-1}}{\Gamma(n)} e^{nx} dx.$$

Если $s < 1$, то

$$\alpha = \mathrm{P}_0(\overline{X} > \tfrac{\ln c_\alpha - n\ln s}{1-s^{-1}}).$$

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:

$$\beta = \mathrm{P}_1(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)} > c_\alpha) = \mathrm{P}_1(-(s^{-1}-1)\sum_{i=1}^n X_i > \ln c_\alpha - n\ln s) = \mathrm{P}_1(\overline{X} < \tfrac{\ln c_\alpha - n\ln s}{1-s^{-1}}) =$$

$$= \mathrm{P}_1(\overline{X} > \gamma_{n,\frac{1}{n},\alpha}) = 1 - \int\limits_0^{\gamma_{n,\frac{1}{n},1-\alpha}} \tfrac{n^n x^{n-1}}{\Gamma(n)} e^{nx} dx.$$

Вычислим асимптотические квантили. Имеем $\mathrm{E}_0(\overline{X}) = 1$, $\mathrm{D}_0(\overline{X}) = \frac{1}{n}$. Тогда $\sqrt{n}(\overline{X}-1) \underset{n\to+\infty}{\xrightarrow{d}} \eta \sim \mathcal{N}(0,1)$. Поэтому

$$\mathrm{P}_0(\overline{X} < 1 + \tfrac{z_\alpha}{\sqrt{n}}) \to \alpha.$$

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

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

$$L_0(x_1, \ldots, x_n) = \prod_{i=1}^n \tfrac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}};$$

$$L_1(x_1, \ldots, x_n) = \prod_{i=1}^n \tfrac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-s)^2}{2}}.$$

Тогда

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

Отсюда

$$\beta = \mathrm{P}_0(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)} > c_\alpha) = \mathrm{P}_0(2s\overline{X} - s^2 > \tfrac{2\ln c_\alpha}{n}) = \mathrm{P}_0(2s\overline{X} > s^2 + \tfrac{2\ln c_\alpha}{n}).$$

Если $s > 0$, то

$$\alpha = \mathrm{P}_0(\overline{X} > \tfrac{s}{2} + \tfrac{\ln c_\alpha}{sn}).$$

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 $c_\alpha = e^{\sqrt{n}sz_{1-\alpha}-\frac{ns^2}{2}}$.

Вычислим мощность критерия:

$$\beta(s)=\mathrm{P}_1(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)}>c_\alpha)=\mathrm{P}_1(2s\overline{X}-s^2>\tfrac{2\ln c_\alpha}{n})=\mathrm{P}_1(2s\overline{X}>s^2+\tfrac{2\ln c_\alpha}{n})=$$
$$=\mathrm{P}_1(\overline{X}>\tfrac{s}{2}+\tfrac{\ln c_\alpha}{sn})=1-\Phi(\sqrt{n}(\tfrac{z_{1-\alpha}}{\sqrt{n}}-s))=1-\Phi(z_{1-\alpha}-\sqrt{n}s).$$

Если $s<0$, то

$$\alpha=\mathrm{P}_0(\overline{X}<\tfrac{s}{2}+\tfrac{\ln c_\alpha}{sn}).$$

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 $c_\alpha = e^{\sqrt{n}sz_\alpha-\frac{ns^2}{2}}$.

Вычислим мощность критерия:

$$\beta=\mathrm{P}_1(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)}>c_\alpha)=\mathrm{P}_1(2s\overline{X}-s^2>\tfrac{2\ln c_\alpha}{n})=\mathrm{P}_1(2s\overline{X}>s^2+\tfrac{2\ln c_\alpha}{n})=$$
$$=\mathrm{P}_1(\overline{X}<\tfrac{s}{2}+\tfrac{\ln c_\alpha}{sn})=\Phi(\sqrt{n}(\tfrac{z_\alpha}{\sqrt{n}}-s))=\Phi(z_\alpha-\sqrt{n}s).$$

Вычислим асимптотические квантили. Имеем $\mathrm{E}_0(\overline{X})=0$, $\mathrm{D}_0(\overline{X})=\frac{1}{n}$. Тогда $\sqrt{n}\overline{X}\underset{n\to+\infty}{\xrightarrow{d}}\eta\sim\mathcal{N}(0,1)$. Поэтому

$$\mathrm{P}_0(\overline{X}<\tfrac{z_\alpha}{\sqrt{n}})\to\alpha.$$

Задача 2аб ($\mathrm{Poiss}(1),\mathrm{Poiss}(s)$)

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

$$L_0(x_1,\ldots,x_n)=\mathrm{P}_\theta(X_1=x_1,\ldots,X_n=x_n)=e^{-n}\tfrac{1}{x_1!\ldots x_n!};$$
$$L_1(s,x_1,\ldots,x_n)=e^{-ns}\tfrac{1}{x_1!\ldots x_n!}s^{(x_1+x_2+\ldots+x_n)}.$$

Тогда

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

Отсюда

$$\alpha=\mathrm{P}_0(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)}>c_\alpha)=\mathrm{P}_0(n(1-s)+n\overline{X}\ln s<c_\alpha)=\mathrm{P}_0(n\overline{X}\ln s<c_\alpha+n(s-1)).$$

Если $s>1$, то

$$\alpha=\mathrm{P}_0(n\overline{X}<\tfrac{c_\alpha+n(s-1)}{\ln s}).$$

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

Вычислим мощность критерия:

$$\beta=\mathrm{P}_1(\tfrac{L_1(X_1,\ldots,X_n)}{L_0(X_1,\ldots,X_n)}>c_\alpha)=\mathrm{P}_1(n(1-s)+n\overline{X}\ln s<c_\alpha)=\mathrm{P}_1(n\overline{X}\ln s<c_\alpha+n(s-1))=$$
$$=\mathrm{P}_1(n\overline{X}<\tfrac{c_\alpha+n(s-1)}{\ln s})=1-F_{ns}(p_{s,\alpha}),$$

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

Если $s<1$, то

$$\alpha=\mathrm{P}_0(n\overline{X}>\tfrac{c_\alpha+n(s-1)}{\ln s}).$$

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

Вычислим мощность критерия:

$$\beta = \mathrm{P}_1(\frac{L_1(X_1,\dots,X_n)}{L_0(X_1,\dots,X_n)} > c_\alpha) = \mathrm{P}_1(n(1-s) + n\overline{X}\ln s < c_\alpha) = \mathrm{P}_1(n\overline{X}\ln s < c_\alpha + n(s-1)) =$$
$$= \mathrm{P}_1(n\overline{X} > \frac{c_\alpha + n(s-1)}{\ln s}) = 1 - F_{ns}(p_{s,1-\alpha}),$$

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

Вычислим асимптотические квантили. Имеем $\mathrm{E}_0(n\overline{X}) = n$, $\mathrm{D}_0(n\overline{X}) = n$. Тогда $\sqrt{n}(\overline{X} - n) \xrightarrow[n\to+\infty]{d} \eta \sim \mathcal{N}(0,1)$. Поэтому

$$\mathrm{P}_0(\overline{X} < n + \frac{z_\alpha}{\sqrt{n}}) \to \alpha.$$

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 $g(\theta_1, X_1,\dots,X_n) = \frac{\overline{X}-\theta_1}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}}$. Тогда при фиксированном θ_1 случайная величина $g\sqrt{n}$ имеет распределение t_{n-1}. Тогда

$$\mathrm{P}_0(q_{n-1,\frac{\alpha}{2}} < g\sqrt{n} < q_{n-1,1-\frac{\alpha}{2}}) = \alpha$$

и

$$\mathrm{P}_0(\frac{q_{n-1,\frac{\alpha}{2}}}{\sqrt{n}} < g < \frac{q_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}}) = \alpha.$$

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 $h(X_1,\dots,X_n) = \sum\limits_{i=1}^{n}(X_i - \overline{X})^2$. Тогда при фиксированном θ_2 случайная величина $\frac{h}{(n-1)\theta_2}$ имеет распределение χ^2_{n-1}. В предположении $\theta_2 = 1$ имеем

$$\mathrm{P}_0(\frac{h}{n-1} > c_{n-1,\alpha}) = 1 - \alpha.$$

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