Решения задач по математической статистике

КОНСТАНТИН ЗЮБИН

Задача 1аб

Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\theta_x, \theta^2)$ и $Y_1, \ldots, Y_n \sim \mathcal{N}(\theta_y, \theta^2)$ — независимые случайные величины.
Пусть нулевая гипотеза предполагает истинность неравенства $\theta_x \geqslant \theta_y$.
Вычислим функцию правдоподобия и найдём её точку максимума.

$$L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = \prod_{i=1}^{n} \tfrac{1}{2\pi\theta^2} e^{-\frac{(X_i-\theta_x)^2+(Y_i-\theta_y)^2}{2\theta^2}};$$

$$\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = -n\ln 2\pi - 2n\ln\theta - \tfrac{1}{2\theta^2}\sum_{i=1}^{n}\left((X_i-\theta_x)^2+(Y_i-\theta_y)^2\right).$$

Частные производные равны

$$\tfrac{\partial}{\partial\theta_x}\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = \tfrac{1}{\theta^2}\left(-n\theta_x + \sum_{i=1}^{n} X_i\right);$$

$$\tfrac{\partial}{\partial\theta_y}\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = \tfrac{1}{\theta^2}\left(-n\theta_y + \sum_{i=1}^{n} Y_i\right);$$

$$\tfrac{\partial}{\partial\theta}\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = -\tfrac{2n}{\theta} + \tfrac{1}{\theta^3}\sum_{i=1}^{n}\left((X_i-\theta_x)^2+(Y_i-\theta_y)^2\right).$$

Точка максимума имеет координаты $\hat\theta_x = \overline{X}, \hat\theta_y = \overline{Y}, \hat\theta^2 = \frac{1}{2n}\sum\limits_{i=1}^{n}\left((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\right)$. Значение в точке максимума равно

$$\tfrac{e^{-n}}{2^n\pi^n\hat\theta^{2n}}.$$

Если $\overline{X} > \overline{Y}$, то точка максимума при условии гипотезы оказывается такой же, как и в общем случае. Поэтому будем рассматривать случай, когда $\overline{X} \leqslant \overline{Y}$. Тогда максимум достигается на границе $\theta_x = \theta_y =: \theta_{xy}$. Вычислим точку максимума и его значение. Имеем

$$L(\theta_0, \theta, \overrightarrow{X}, \overrightarrow{Y}) = \prod_{i=1}^{n} \tfrac{1}{2\pi\theta^2} e^{-\frac{(X_i-\theta_{xy})^2+(Y_i-\theta_{xy})^2}{2\theta^2}};$$

$$\ln L(\theta_0, \theta, \overrightarrow{X}, \overrightarrow{Y}) = -n\ln 2\pi - 2n\ln\theta - \tfrac{1}{2\theta^2}\sum_{i=1}^{n}\left((X_i-\theta_0)^2+(Y_i-\theta_0)^2\right).$$

Частные производные равны

$$\tfrac{\partial}{\partial\theta_x}\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = \tfrac{1}{2\theta^2}\left(-2n\theta_{xy} + \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i\right);$$

$$\tfrac{\partial}{\partial\theta}\ln L(\theta_x, \theta_y, \theta, \overrightarrow{X}, \overrightarrow{Y}) = -\tfrac{2n}{\theta} + \tfrac{1}{\theta^3}\sum_{i=1}^{n}\left((X_i-\theta_{xy})^2+(Y_i-\theta_{xy})^2\right).$$

Точка максимума имеет координаты $\hat{\theta}_{xy} = \frac{\overline{X}+\overline{Y}}{2}, \hat{\theta}^2_{H_0} = \frac{1}{2n}\sum\limits_{i=1}^{n}\Big((X_i - \frac{\overline{X}+\overline{Y}}{2})^2 + (Y_i - \frac{\overline{X}+\overline{Y}}{2})^2\Big)$. Значение в точке максимума равно

$$\frac{e^{-n}}{2^n\pi^n\hat{\theta}^{2n}_{H_0}}.$$

Отсюда

$$\lambda = \frac{\sup_{\theta\in\Theta} L(\theta_x,\theta_y,\theta,\overrightarrow{X},\overrightarrow{Y})}{\sup_{\theta\in\Theta_0} L(\theta_0,\theta,\overrightarrow{X},\overrightarrow{Y})} = \frac{\hat{\theta}^{2n}_{H_0}}{\hat{\theta}^{2n}}.$$

$$\sqrt[n]{\lambda} = \frac{\sum\limits_{i=1}^{n}\Big((X_i-\frac{\overline{X}+\overline{Y}}{2})^2+(Y_i-\frac{\overline{X}+\overline{Y}}{2})^2\Big)}{\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)} = \frac{\sum\limits_{i=1}^{n}\Big((X_i-\overline{X}-\frac{\overline{Y}-\overline{X}}{2})^2+(Y_i-\overline{Y}-\frac{\overline{X}-\overline{Y}}{2})^2\Big)}{\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)} = 1 + \frac{2n(\frac{\overline{X}-\overline{Y}}{2})^2}{\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}.$$

Отсюда

$$\sqrt[n]{\lambda} - 1 = \frac{n}{2}\cdot\frac{(\overline{X}-\overline{Y})^2}{\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}.$$

Имеем следующие распределения для случайных величин

$$(\overline{X}-\overline{Y}) - (\theta_x-\theta_y) \sim \mathcal{N}(0, \tfrac{2\theta^2}{n});$$

$$\frac{1}{\theta^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \xi^2_{n-1};$$

$$\frac{1}{\theta^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 + \frac{1}{\theta^2}\sum_{i=1}^{n}(Y_i-\overline{Y})^2 \sim \xi^2_{2n-2};$$

Тогда

$$\sqrt{\sqrt[n]{\lambda}-1}\cdot\sqrt{\frac{2}{n}\cdot\frac{\frac{n}{2\theta^2}}{\frac{1}{(2n-2)\theta^2}}} = \frac{\sqrt{\frac{n}{2\theta^2}}((\overline{X}-\overline{Y})-(\theta_x-\theta_y))}{\sqrt{\frac{1}{(2n-2)\theta^2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}} \sim t_{2n-2}$$

и

$$\sqrt{\sqrt[n]{\lambda}-1}\cdot\frac{1}{\sqrt{2n-2}} \sim t_{2n-2}.$$

Пусть q_α — квантиль уровня α для распределения Стьюдента t_{2n-2}. Будем отклонять гипотезу при $\frac{\sqrt{\frac{n}{2}}(\overline{X}-\overline{Y})}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}} < q_\alpha$.

Обозначим через $F_{t_{2n-2}}$ — функцию распределения Стьюдента. Тогда

$$\sup_{\theta_x\geqslant\theta_y} \mathrm{P}\left(\frac{\sqrt{2n}(\frac{\overline{X}-\overline{Y}}{2})}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}} < q_\alpha\right) =$$

$$= \sup_{\theta_x\geqslant\theta_y} \mathrm{P}\left(\frac{\sqrt{\frac{n}{2}}((\overline{X}-\overline{Y})-(\theta_x-\theta_y))}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}} < q_\alpha - \frac{\sqrt{\frac{n}{2}}(\theta_x-\theta_y)}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}}\right) =$$

$$= \sup_{\theta_x\geqslant\theta_y} F_{t_{2n-2}}\left(q_\alpha - \frac{\sqrt{\frac{n}{2}}(\theta_x-\theta_y)}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\Big((X_i-\overline{X})^2+(Y_i-\overline{Y})^2\Big)}}\right) = F_{t_{2n-2}}(q_\alpha - 0) = \alpha.$$

Аналогично, имеем выражение для мощности:

$$\beta(\theta_1, \theta_2) = F_{t_{2n-2}}\left(q_\alpha - \frac{\sqrt{\frac{n}{2}}(\theta_x - \theta_y)}{\sqrt{\frac{1}{2n-2}\sum\limits_{i=1}^{n}\left((X_i - \overline{X})^2 + (Y_i - \overline{Y})^2\right)}}\right),$$

что больше $F_{t_{2n-2}}(q_\alpha) = \alpha$ при $\theta_x < \theta_y$.

Задача 2

Рассмотрим независимые случайные величины $X_1, \ldots, X_n \sim \mathcal{N}(\mu_x, \theta_x^2)$ и $Y_1, \ldots \mathcal{N}(\mu_y, \theta_y^2)$. Параметры μ_x, μ_y будем считать известными числами.

Пусть гипотеза состоит в том, что $\theta_x \geqslant 2\theta_y$. Построим точный критерий обобщённого отношения правдоподобий. Имеем

$$L(\theta_x, \theta_y, \overrightarrow{X}, \overrightarrow{Y}) = \prod_{i=1}^{n} \frac{1}{2\pi\theta_x\theta_y} e^{-\frac{(X_i - \mu_x)^2}{2\theta_x^2} - \frac{(Y_i - \mu_y)^2}{2\theta_y^2}},$$

$$\ln L(\theta_x, \theta_y, \overrightarrow{X}, \overrightarrow{Y}) = -n \ln 2\pi - n \ln \theta_x - n \ln \theta_y - \frac{1}{2\theta_x^2}\sum_{i=1}^{n}(X_i - \mu_x)^2 - \frac{1}{2\theta_y^2}\sum_{i=1}^{n}(Y_i - \mu_y)^2.$$

Тогда

$$\frac{\partial}{\partial\theta_x} \ln L(\theta_x, \theta_y, \overrightarrow{X}, \overrightarrow{Y}) = -\frac{n}{\theta_x} + \frac{1}{\theta_x^3}\sum_{i=1}^{n}(X_i - \mu_x)^2,$$

$$\frac{\partial}{\partial\theta_y} \ln L(\theta_x, \theta_y, \overrightarrow{X}, \overrightarrow{Y}) = -\frac{n}{\theta_y} + \frac{1}{\theta_y^3}\sum_{i=1}^{n}(Y_i - \mu_y)^2.$$

Максимум достигается в точке с координатами $S_x^2 := \hat{\theta}_x^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \mu_x)^2$ и $S_y^2 := \hat{\theta}_y^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(Y_i - \mu_y)^2$.

Заметим, что

$$\frac{S_x^2}{S_y^2} = \frac{\sum\limits_{i=1}^{n}(X_i - \mu_x)^2}{\sum\limits_{i=1}^{n}(Y_i - \mu_y)^2} = \frac{\frac{1}{n\theta_x^2}\sum\limits_{i=1}^{n}(X_i - \mu_x)^2}{\frac{1}{n\theta_y^2}\sum\limits_{i=1}^{n}(Y_i - \mu_y)^2} \cdot \frac{\theta_x^2}{\theta_y^2}$$

и $\dfrac{\frac{1}{n\theta_x^2}\sum\limits_{i=1}^{n}(X_i - \mu_x)^2}{\frac{1}{n\theta_y^2}\sum\limits_{i=1}^{n}(Y_i - \mu_y)^2} \sim F_{n,n}$.

Если $S_x^2 \geqslant 4S_y^2$, то точка максимума лежит в подмножестве $\{\theta_x > 2\theta_y\}$ и поэтому отношение супремумов правдоподобий будет равно 1. Будем рассматривать случай, когда $S_x^2 < 4S_y^2$. Тогда максимум достигается на границе области $\theta_x = 2\theta_y$, $\theta_y =: \theta_{xy}$. Найдём его

$$L(\theta_{xy}, \overrightarrow{X}, \overrightarrow{Y}) = \prod_{i=1}^{n} \frac{1}{4\pi\theta_{xy}^2} e^{-\frac{(X_i - \mu_x)^2}{8\theta_{xy}^2} - \frac{(Y_i - \mu_y)^2}{2\theta_{xy}^2}},$$

$$\ln L(\theta_{xy}, \overrightarrow{X}, \overrightarrow{Y}) = -n \ln 4\pi - 2n \ln \theta_{xy} - \frac{1}{8\theta_{xy}^2}\sum_{i=1}^{n}(X_i - \mu_x)^2 - \frac{1}{2\theta_{xy}^2}\sum_{i=1}^{n}(Y_i - \mu_y)^2.$$

Тогда

$$\frac{\partial}{\partial\theta_{xy}} \ln L(\theta_x, \theta_y, \overrightarrow{X}, \overrightarrow{Y}) = -\frac{2n}{\theta_{xy}} + \frac{1}{\theta_{xy}^3}\sum_{i=1}^{n}(\tfrac{1}{4}(X_i - \mu_x)^2 + (Y_i - \mu_y)^2).$$

Точка максимума есть $\hat{\theta}^2_{xy} = \frac{1}{2n}\sum\limits_{i=1}^{n}(\frac{1}{4}(X_i-\mu_x)^2+(Y_i-\mu_y)^2) = \frac{1}{8}S_x^2+\frac{1}{2}S_y^2$.

Имеем

$$\lambda = \frac{\sup\limits_{\vec{\theta}\in\Theta} L(\theta_x,\theta_y,\vec{X},\vec{Y})}{\sup\limits_{\vec{\theta}\in\Theta_0} L(\theta_x,\theta_y,\vec{X},\vec{Y})} = \left(\frac{\hat{\theta}^2_{xy}}{S_xS_y}\right)^n = \left(\frac{\frac{1}{8}S_x^2+\frac{1}{2}S_y^2}{S_xS_y}\right)^n = \left(\frac{1}{8}\frac{S_x}{S_y}+\frac{1}{2}(\frac{S_x}{S_y})^{-1}\right)^n,$$

$$\sqrt[n]{\lambda} = \frac{1}{8}\frac{S_x}{S_y}+\frac{1}{2}(\frac{S_x}{S_y})^{-1}.$$

По предположению $0<\frac{S_x}{S_y}<2$. Рассмотрим функцию $f(u)=\frac{u}{8}+\frac{1}{2u}$. Её производная равна $f'(u)=\frac{1}{8}-\frac{1}{2u^2}$ и она меньше нуля при $0<u<2$. Следовательно, функция f монотонно убывает на $(0,2)$.

Имеем

$$\mathrm{P}\left(\frac{S_x^2}{S_y^2}<c_\alpha\right) = \mathrm{P}\left(\frac{S_x^2}{S_y^2}\cdot\frac{\theta_y^2}{\theta_x^2}<\frac{c_\alpha\theta_y^2}{\theta_x^2}\right) = F_{F_{2n,2n}}(\frac{c_\alpha\theta_y^2}{\theta_x^2}).$$

Будет отвергать гипотезу при $\frac{S_x^2}{S_y^2}<4q_\alpha=c_\alpha$, где q_α — квантиль уровня α распределения Фишера-Снедекора. Тогда можно вычислить уровень значимости

$$\sup_{\theta_x\geqslant 2\theta_y}\mathrm{P}\left(\frac{S_x^2}{S_y^2}<c_\alpha\right) = \sup_{\theta_x\geqslant 2\theta_y}F_{F_{2n,2n}}(\frac{2\theta_y^2}{\theta_x^2}\cdot q_\alpha)=\alpha$$

и мощность критерия:

$$\beta(\theta_x,\theta_y) = \mathrm{P}\left(\frac{S_x^2}{S_y^2}<c_\alpha\right) = F_{F_{2n,2n}}(\frac{4\theta_y^2}{\theta_x^2}\cdot q_\alpha)>\alpha.$$

Задача 3

Предположим, что в предыдущей задаче μ_x и μ_y являются параметрами. Построим точный критерий обобщённого правдоподобия для гипотезы $\theta_x\leqslant 2\theta_y$.

Точкой максимума для $L(\theta_x,\theta_y,\mu_x,\mu_y,\vec{X},\vec{Y})$ будет точка с координатами $\hat{\mu}_x=\overline{X}, \hat{\mu}_y=\overline{Y}$ и $S_x^2=\hat{\theta}_x^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2, S_y^2=\hat{\theta}_y^2=\frac{1}{n}\sum\limits_{i=1}^{n}(Y_i-\overline{Y})^2$.

В этом случае

$$\frac{S_x^2}{S_y^2} = \frac{\frac{1}{(n-1)\theta_x^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\frac{1}{(n-1)\theta_y^2}\sum\limits_{i=1}^{n}(Y_i-\overline{Y})^2}$$

При условии гипотезы, если $\hat{\theta}_y^2$