Решения задач по математической статистике

КОНСТАНТИН ЗЮБИН

Задача 1а (о плотности распределения Стьюдента)

Пусть $X \sim \mathcal{N}(0,1)$ и $Y \sim \chi^2_n$. Положим $\xi = \frac{X}{\sqrt{\frac{Y}{n}}}$. Вычислим условную плотности ξ при условии Y. При фиксированном $Y = y_0$ плотность ξ равна плотности случайной величины $\frac{X}{\sqrt{\frac{y_0}{n}}}$, то есть равна $\sqrt{\frac{y_0}{2\pi n}} e^{-\frac{y_0 z^2}{2n}}$. Тогда $f_{\xi|Y}(z \mid y) = \sqrt{\frac{y}{2\pi n}} e^{-\frac{yz^2}{2n}}$.

Задача 1б (о плотности распределения Стьюдента)

Вычислим условные матожидания $\mathrm{E}(\xi \mid Y)$ и $\mathrm{E}(\xi^2 \mid Y)$.

Имеем

$$\mathrm{E}(\xi \mid Y) = \int\limits_{-\infty}^{+\infty} z\sqrt{\frac{Y}{2\pi n}} e^{-\frac{Yz^2}{2n}}\, dz = 0,$$

так как подынтегральная функция нечётна по z.

Далее,

$$\mathrm{E}(\xi^2 \mid Y) = \int\limits_{-\infty}^{+\infty} z^2\sqrt{\frac{Y}{2\pi n}} e^{-\frac{Yz^2}{2n}}\, dz = 0 + \frac{Y}{n} = \frac{Y}{n},$$

как дисперсия нормально распределённой случайной величины с параметрами 0 и $\frac{Y}{n}$.

Задача 1а (о плотности распределения Стьюдента)

Вычислим плотность величины ξ. Имеем

$$f_\xi(z) = \int\limits_{-\infty}^{+\infty} f_{\xi|Y}(z \mid y) f_Y(y) dy = \int\limits_{-\infty}^{+\infty} \sqrt{\frac{y}{2\pi n}} e^{-\frac{yz^2}{2n}} \cdot \frac{y^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} e^{-\frac{y}{2}} dy =$$

$$= \int\limits_{-\infty}^{+\infty} \frac{y^{\frac{n-1}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} \cdot \sqrt{\frac{1}{2\pi n}} e^{-\frac{y}{2}(1+\frac{z^2}{n})} dy =$$

$$= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} \cdot 2^{\frac{n+1}{2}} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}} \cdot \frac{1}{\sqrt{2\pi n}} \cdot I,$$

где I — интеграл по всему $\mathbb{R}$ от плотности случайной величины с распределением $\Gamma(\frac{n+1}{2}, \frac{1}{2} + \frac{z^2}{2n})$. Тогда $I = 1$ и

$$f_\xi(z) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}.$$

Задача 2а (распределение Фишера-Снедекора)

Пусть $X \sim \xi^2_n = \Gamma(\frac{n}{2}, 2)$ и $Y \sim \xi^2_n = \Gamma(\frac{m}{2}, 2)$ независимы. Положим $\xi = \frac{X/n}{Y/m}$.

Найдём условную плотность ξ при условии Y. При $Y = y$ величина $\xi = \frac{X}{yn/m}$ имеет распределение $\Gamma(\frac{n}{2}, \frac{2m}{yn})$ и плотность $\frac{z^{\frac{n}{2}-1}(yn)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})(2m)^{\frac{n}{2}}} e^{-\frac{zyn}{2m}} I(z > 0)$.

Задача 2б (распределение Фишера-Снедекора)

Найдём плотность ξ. Имеем

$$f_\xi(z) = \int\limits_{-\infty}^{+\infty} f_{\xi|Y}(z \mid y) f_Y(y) dy = \int\limits_{-\infty}^{+\infty} \frac{z^{\frac{n}{2}-1}(yn)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})(2m)^{\frac{n}{2}}} e^{-\frac{zyn}{2m}} \cdot \frac{y^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}} e^{-\frac{y}{2}} I(z>0) I(y>0) dy =$$

$$\frac{z^{\frac{n}{2}-1} n^{\frac{n}{2}}}{B(\frac{n}{2},\frac{m}{2})(2m)^{\frac{n}{2}}} \cdot 2^{\frac{n+m}{2}} \cdot \left(1+\tfrac{zn}{m}\right)^{-\frac{n+m}{2}} I(z>0) \cdot \int\limits_{-\infty}^{+\infty} \frac{y^{\frac{(n+m)}{2}-1}}{\Gamma(\frac{n+m}{2})2^{\frac{n+m}{2}} \cdot (1+\frac{zn}{m})^{-\frac{n+m}{2}}} e^{-\frac{y}{2}(1+\frac{zn}{m})} I(y>0) dy =$$

$$= 2^{\frac{m}{2}} z^{\frac{n}{2}-1} \cdot \frac{n^{\frac{n}{2}} m^{\frac{m}{2}}}{B(\frac{n}{2},\frac{m}{2})} \cdot (m+zn)^{-\frac{n+m}{2}} I(z>0).$$

Интеграл в предпоследней строчке равен 1, как интеграл от плотности случайной величины по $\mathbb{R}$.

Задача 2в (распределение Фишера-Снедекора)

Найдём $\mathrm{E}\,\xi$ и $\mathrm{E}\,\xi^2$. Предварительно для случайной величины T с распределением $\Gamma(\alpha,\beta)$ вычислим $\mathrm{E}\,T^s$. Имеем

$$\mathrm{E}\,T^s = \int\limits_{-\infty}^{+\infty} \frac{t^{s+\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{x}{\beta}} = \frac{\Gamma(s+\alpha)}{\Gamma(\alpha)} \cdot \beta^s \int\limits_{-\infty}^{+\infty} \frac{t^{s+\alpha-1}}{\Gamma(s+\alpha)\beta^{s+\alpha}} e^{-\frac{x}{\beta}} = \frac{\Gamma(s+\alpha)}{\Gamma(\alpha)} \cdot \beta^s.$$

Теперь

$$\mathrm{E}\,\xi = \mathrm{E}\,\frac{mX}{nY} = \frac{m}{n}\,\mathrm{E}\,X \cdot \mathrm{E}\,\frac{1}{Y} = \frac{m}{n} \cdot \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{m}{2}-1)}{\Gamma(\frac{m}{2})} = \frac{mn}{n(m-2)} = \frac{m}{m-2}.$$

и матожидание будет конечным при $m > 2$.

Далее,

$$\mathrm{E}\,\xi^2 = \mathrm{E}\,\frac{m^2X^2}{n^2Y^2} = \frac{m^2}{n^2}\,\mathrm{E}\,X^2 \cdot \mathrm{E}\,\frac{1}{Y^2} = \frac{m^2}{n^2} \cdot \frac{\Gamma(\frac{n}{2}+2)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{m}{2}-2)}{\Gamma(\frac{m}{2})} = \frac{m^2 n(n+2)}{n^2(m-4)(m-2)} = \frac{m^2(n+2)}{n(m-4)(m-2)}.$$

и матожидание будет конечным при $m > 4$.

Задача 3

Пусть $X_n \sim \mathcal{N}(0,1)$ и $Y_m \sim \mathcal{N}(0,1)$ — последовательности попарно независимых (между последовательностями тоже) случайных величин. Положим $\tilde{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$, $\tilde{Y}_m = \frac{1}{m}\sum\limits_{i=1}^{m} Y_i^2$ и $\xi_m^{(n)} = \frac{\tilde{X}_n}{\tilde{Y}_n} \sim F_{n,m}$. Тогда $\mathrm{E}\,Y_i^2 = 1$ и по закону больших чисел $\tilde{Y}_m \xrightarrow[m\to+\infty]{\mathrm{P}} 1$. По лемме Слуцкого $\xi_m^{(n)} = \tilde{X}_n \cdot \frac{1}{\tilde{Y}_m} \xrightarrow[m\to+\infty]{d} \tilde{X}_n$.

Пусть $Z \sim \mathcal{N}(0,1)$ — ещё одна случайная величина, независимая с Y_m. Положим $\eta_m = \frac{Z}{\sqrt{\tilde{Y}_m}} \sim t_m$. Выше было показано, что $\tilde{Y} \xrightarrow[m\to+\infty]{\mathrm{P}} 1$. Тогда $\sqrt{\tilde{Y}_m} \xrightarrow[m\to+\infty]{\mathrm{P}} 1$ и снова по лемме Слуцкого

$$\eta_m = Z \cdot \frac{1}{\sqrt{\tilde{Y}_m}} \xrightarrow[m\to+\infty]{d} Z.$$

Задача 4

Пусть $X_1,\dots,X_n \sim \mathcal{N}(\theta_x,\sigma^2)$ и $Y_1,\dots,Y_m \sim \mathcal{N}(\theta_y,\sigma^2)$. Построим точный доверительный интервал для $\theta_x - \theta_y$ на основе оценки

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

Положим $S^2 = \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2+\sum\limits_{i=1}^{n}(-Y_i+\overline{Y})^2}{n+m-2}$. Случайная величина $(\overline{X}-\overline{Y})-(\theta_x-\theta_y)$ имеет распределение $\mathcal{N}(0,\sigma^2(\frac{1}{n}+\frac{1}{m}))$. Тогда случайная величина $\frac{(\overline{X}-\overline{Y})-(\theta_x-\theta_y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$ имеет распределение Стьюдента t_{n+m-2}. Имеем

$$\mathrm{P}\left(t_{\frac{\alpha}{2}} < \frac{(\overline{X}-\overline{Y})-(\theta_x-\theta_y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Тогда

$$\mathrm{P}\left((\overline{X}-\overline{Y}) - t_{1-\frac{\alpha}{2}}S\sqrt{\frac{1}{n}+\frac{1}{m}} < (\theta_x-\theta_y) < (\overline{X}-\overline{Y}) - t_{\frac{\alpha}{2}}S\sqrt{\frac{1}{n}+\frac{1}{m}}\right) = 1-\alpha$$

и доверительный интервал есть

$$\left((\overline{X}-\overline{Y}) - t_{1-\frac{\alpha}{2}}S\sqrt{\frac{1}{n}+\frac{1}{m}}, (\overline{X}-\overline{Y}) - t_{\frac{\alpha}{2}}S\sqrt{\frac{1}{n}+\frac{1}{m}}\right).$$

Задача 5а ($\mathcal{N}(0,\theta)$)

Пусть $X_1,\ldots,X_n \sim \mathcal{N}(0,\theta)$. Построим точный доверительный интервал на основе оценки максимального правдоподобия $T = \overline{X^2}$ (см. домашнюю работу 3). Поскольку случайные величины $\frac{X_i}{\sqrt{\theta}}$ имеют стандартное нормальное распределение, то случайная величина $\frac{nT}{\theta}$ имеет χ^2_n-распределение. Обозначим через $c_{\alpha,n}$ — квантиль уровня α для функции распределения случайной величины с распределением χ^2_n. Тогда

$$\mathrm{P}\left(c_{1-\frac{\alpha}{2},n} < \frac{nT}{\theta} < c_{\frac{\alpha}{2},n}\right) = 1-\alpha.$$

Отсюда

$$\mathrm{P}\left(\frac{nT}{c_{\frac{\alpha}{2},n}} < \theta < \frac{nT}{c_{1-\frac{\alpha}{2},n}}\right) = 1-\alpha$$

и точный доверительный интервал есть

$$\left(\frac{nT}{c_{\frac{\alpha}{2},n}}, \frac{nT}{c_{1-\frac{\alpha}{2},n}}\right).$$

Задача 5б ($\mathcal{N}(\theta_1,\theta)$)

Пусть $X_1,\ldots,X_n \sim \mathcal{N}(\theta_1,\theta)$. Построим точный доверительный интервал на основе оценки максимального правдоподобия (для θ): $T = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})$ (см. вычисление после решений задач, где $a=\theta_1$ и $\theta=\sigma^2$). Поскольку случайные величины $Y_i = \frac{X_i-\theta_1}{\sqrt{\theta}}$ имеют стандартное нормальное распределение, то случайная величина

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2 = \sum_{i=1}^{n}\left(\frac{X_i-\theta_i}{\sqrt{\theta}} - \frac{\overline{X}-\theta_i}{\sqrt{\theta}}\right)^2 = \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\theta} = \frac{nT}{\theta}$$

имеет χ^2_{n-1}-распределение. Обозначим через $c_{\alpha,n}$ — квантиль уровня α для функции распределения случайной величины с распределением χ^2_n. Тогда

$$\mathrm{P}\left(c_{1-\frac{\alpha}{2},n-1} < \frac{nT}{\theta} < c_{\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

Отсюда

$$\mathrm{P}\left(\frac{nT}{c_{\frac{\alpha}{2},n-1}} < \theta < \frac{nT}{c_{\frac{1-\alpha}{2},n-1}}\right) = 1-\alpha$$

и точный доверительный интервал есть

$$\left(\frac{nT}{c_{\frac{\alpha}{2},n-1}}, \frac{nT}{c_{1-\frac{\alpha}{2},n-1}}\right).$$

Задача 5в ($\mathcal{N}(\theta_1, \theta)$)

Пусть $X_1, \dots, X_n, Y \sim \mathcal{N}(\theta_1, \theta)$. Построим аналог точного доверительного интервала для Y. Случайная величина $\overline{X} - Y$ имеет распределение $\mathcal{N}(0, \frac{n+1}{n}\theta)$, а случайные величины $X_i - \overline{X} = (X_i - Y) - (\overline{X} - Y)$ — распределение $\mathcal{N}(0, \frac{n-1}{n}\theta)$. Тогда следующая случайная величина имеет распределение Стьюдента t_{n-1}:

$$\frac{\sqrt{\frac{n}{(n+1)\theta}}(\overline{X}-Y)}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}\frac{n}{(n-1)\theta}(X_i-\overline{X})^2}} = \frac{\sqrt{n-1}}{\sqrt{n+1}}T,$$

где $T = \frac{\overline{X}-Y}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}}$. Теперь

$$\mathrm{P}\left(c_{1-\frac{\alpha}{2},n-1} < \frac{\sqrt{n-1}}{\sqrt{n+1}} \cdot \frac{\overline{X}-Y}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}} < c_{\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

Отсюда

$$\mathrm{P}\left(\overline{X} - c_{\frac{\alpha}{2},n-1}\frac{\sqrt{n+1}}{\sqrt{n-1}} \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} < Y < \overline{X} - c_{1-\frac{\alpha}{2},n-1}\frac{\sqrt{n+1}}{\sqrt{n-1}} \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}\right) = 1-\alpha$$

и точный доверительный интервал есть

$$\left(\overline{X} - c_{\frac{\alpha}{2},n-1}\frac{\sqrt{n+1}}{\sqrt{n-1}} \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}, \overline{X} - c_{1-\frac{\alpha}{2},n-1}\frac{\sqrt{n+1}}{\sqrt{n-1}} \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}\right).$$

Вычисление ОМП для $\mathcal{N}(a, \sigma^2)$

Найдём оценку максимального правдоподобия для случайных величин с распределением $\mathcal{N}(a, \sigma^2)$. Имеем формулу для плотности $f_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$. Тогда

$$f_\theta(x_1, \dots, x_n) = \prod_{j=1}^{n} f_\theta(x_j) = (2\pi)^{\frac{-n}{2}} \cdot \frac{1}{\sigma^n} \cdot e^{-\frac{1}{2\sigma^2}\sum\limits_{j=1}^{n}(x_j-a)^2}.$$

Далее будем исследовать точки максимума логарифма

$$M(a, \sigma) = \ln\left(\frac{1}{\sigma^n} \cdot e^{-\frac{1}{2\sigma^2}\sum\limits_{j=1}^{n}(x_j-a)^2}\right) = -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-a)^2.$$

Вычислим частные производные по a и σ:

$$\frac{\partial M}{\partial a}\Big|_a = -\frac{1}{\sigma^2}\left(na - \sum_{j=1}^{n} x_j\right),$$

$$\frac{\partial M}{\partial \sigma}\Big|_\sigma = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(x_j-a)^2.$$

Тогда приравнивая их к 0, получаем $a = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ и $\hat{a} = \overline{X}$, а также $\sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - a)^2$, откуда $\hat{\sigma} = \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}$.