Решения задач по математической статистике

КОНСТАНТИН ЗЮБИН

Задача 1а

Пусть $X_1, \dots, X_n \sim \mathcal{N}(\theta, 1)$ и $\theta \sim \mathcal{N}(\mu, 1)$. Вычислим апостериорную плотность θ. Имеем

$$\pi_{\theta|X_1,\dots,X_n}(t, x_1, \dots, x_n) = \frac{f_{\theta,X_1,\dots,X_n}(t,x_1,\dots,x_n)}{f_{X_1,\dots,X_n}(x_1,\dots,x_n)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2}}\prod\limits_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i-t)^2}{2}}}{f_{X_1,\dots,X_n}(x_1,\dots,x_n)} =$$

$$= C(x_1, \dots, x_n, \mu)e^{-\frac{t^2-2t\mu}{2}-\frac{\sum\limits_{i=1}^{n}(x_i-t)^2}{2}} = C(x_1, \dots, x_n, \mu)e^{-\frac{1}{2}\left((n+1)t^2-2t\sum\limits_{i=1}^{n}x_i-2t\mu+\sum\limits_{i=1}^{n}x_i^2\right)} =$$

$$= \tilde{C}(x_1, \dots, x_n, \mu)e^{-\frac{n+1}{2}\left(t-\frac{1}{n+1}\sum\limits_{i=1}^{n}x_i-\frac{\mu}{n+1}\right)^2}.$$

Тогда $\pi_{\theta|X_1,\dots,X_n}$ — плотность случайной величины $\hat{\theta}$ с распределением

$$\mathcal{N}\left(\frac{n\overline{X}+\mu}{n+1}, \frac{1}{n+1}\right).$$

Задача 1б

Вычислим байесовскую оценка для абсолютного риска. Имеем

$$\frac{1}{2} = F_{\hat{\theta}}(q) = \Phi_{\frac{n\overline{X}+\mu}{n+1},\frac{1}{n+1}}(q).$$

Отсюда $q - \frac{n\overline{X}+\mu}{n+1} = 0$ и $q = \frac{n\overline{X}+\mu}{n+1}$.

Вычислим байесовскую оценку для среднеквадратичного риска:

$$S = \mathrm{E}\,\hat{\theta} = \frac{n\overline{X}+\mu}{n+1}.$$

Вычислим байесовский доверительный интервал. В условиях задачи имеем

$$q_\alpha = \frac{z_\alpha}{\sqrt{n+1}} + \frac{n\overline{X}+\mu}{n+1}.$$

Тогда доверительный интервал $(q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}})$ равен

$$\left(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n+1}} + \frac{n\overline{X}+\mu}{n+1}, \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n+1}} + \frac{n\overline{X}+\mu}{n+1}\right).$$

Задача 2аб

Пусть $X_1, \dots, X_n \sim \mathcal{N}(0, \frac{1}{\theta})$ и $\theta \sim \Gamma(2, \beta)$. Вычислим апостериорную плотность θ. Имеем

$$\pi_{\theta|X_1,\dots,X_n}(t, x_1, \dots, x_n) = \frac{f_{\theta,X_1,\dots,X_n}(t,x_1,\dots,x_n)}{f_{X_1,\dots,X_n}(x_1,\dots,x_n)} = \frac{\frac{t}{\beta^2}e^{-\frac{t}{\beta}}I(t>0)\prod\limits_{i=1}^{n}\frac{\sqrt{t}}{\sqrt{2\pi}}e^{-\frac{x_i^2 t}{2}}}{f_{X_1,\dots,X_n}(x_1,\dots,x_n)} =$$

$$= C(x_1, \dots, x_n, \beta)t^{1+\frac{n}{2}}e^{-\frac{t}{2}(\frac{2}{\beta}+\sum\limits_{i=1}^{n}x_i^2)}.$$

Тогда $\pi_{\theta|X_1,\dots,X_n}$ — плотность случайной величины $\hat{\theta}$ с распределением

$$\Gamma\left(2 + \frac{n}{2}, \left(\frac{1}{\beta} + \frac{1}{2}S_n^2\right)^{-1}\right),$$

где $S_n^2 = \sum\limits_{i=1}^{n} X_i^2$.

Вычислим байесовскую оценку для среднеквадратичного риска:

$$S = \mathrm{E}\,\hat{\theta} = \frac{2+\frac{n}{2}}{\frac{1}{\beta}+\frac{1}{2}S_n^2}.$$

Задача 3аб

Пусть $X_1, \ldots, X_n \sim \mathrm{R}[0,\theta]$ и $\theta \sim \mathrm{B}(\alpha, 1)$. Вычислим апостериорную плотность θ. Имеем

$$\pi_{\theta|X_1,\ldots,X_n}(t, x_1, \ldots, x_n) = \frac{f_{\theta,X_1,\ldots,X_n}(t,x_1,\ldots,x_n)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)} = \frac{\frac{1}{B(\alpha,1)}t^{\alpha-1}I(0<t<1)\prod\limits_{i=1}^{n}\frac{1}{t}I(0<x_i<t)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)} =$$

$$= C(x_1, \ldots, x_n, \alpha)t^{\alpha-n-1}I(\max_{i=1\ldots n} x_i < t < 1).$$

Тогда $\pi_{\theta|X_1,\ldots,X_n} = C(X_1, \ldots, X_n, \alpha)t^{\alpha-n-1}I(X_{(n)} < t < 1)$ — плотность случайной величины $\hat{\theta}$. Для $C(X_1, \ldots, X_n)$ имеем выражение:

$$C(X_1, \ldots, X_n) = \int\limits_{-\infty}^{+\infty} t^{\alpha-n-1}I(X_{(n)} < t < 1)dt = \int\limits_{X_{(n)}}^{1} t^{\alpha-n-1}dt = t^{\alpha-n}\Big|_{X_{(n)}}^{1} = 1 - X_{(n)}^{\alpha-n}.$$

Таким образом,

$$\pi_{\theta|X_1,\ldots,X_n} = \frac{t^{\alpha-n-1}I(X_{(n)}<t<1)}{1-X_{(n)}^{\alpha-n}}.$$

Вычислим квантили этого распределения. Имеем

$$\beta = \int\limits_{-\infty}^{q_\beta} \frac{t^{\alpha-n-1}I(X_{(n)}<t<1)}{1-X_{(n)}^{\alpha-n}}dt = \int\limits_{X_{(n)}}^{q_\beta} \frac{t^{\alpha-n-1}}{1-X_{(n)}^{\alpha-n}}dt = \frac{q_\beta^{\alpha-n}-X_{(n)}^{\alpha-n}}{1-X_{(n)}^{\alpha-n}}.$$

Отсюда

$$q_\beta = \left(\beta(1 - X_{(n)}^{\alpha-n}) + X_{(n)}^{\alpha-n}\right)^{\frac{1}{\alpha-n}} = \left(\beta + (1-\beta)X_{(n)}^{\alpha-n}\right)^{\frac{1}{\alpha-n}}.$$

Тогда байесовский доверительный интервал для $1-\beta$ есть

$$\left(\left(\frac{\beta}{2} + (1-\frac{\beta}{2})X_{(n)}^{\alpha-n}\right)^{\frac{1}{\alpha-n}}, \left(1 - \frac{\beta}{2} + \frac{\beta}{2}X_{(n)}^{\alpha-n}\right)^{\frac{1}{\alpha-n}}\right).$$

Байесовская оценка для абсолютного риска равна

$$q = \left(\frac{1}{2} + \frac{1}{2}X_{(n)}^{\alpha-n}\right)^{\frac{1}{\alpha-n}}.$$

Задача 4аб

Пусть T — достаточная статистика. По критерию факторизации имеется разложение

$$L(t, x_1, \ldots, x_n) = g(t, T(x_1, \ldots, x_n))h(x_1, \ldots, x_n)$$

для некоторых g и h. Тогда апостериорная плотность равна

$$\pi_{\theta|X_1,\ldots,X_n}(t, x_1, \ldots, x_n) = \frac{f_{\theta,X_1,\ldots,X_n}(t,x_1,\ldots,x_n)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)} = \frac{g(t,T)h(x_1,\ldots,x_n)}{\int\limits_{\Theta} g(t,T)h(x_1,\ldots,x_n)dt} = \frac{h(x_1,\ldots,x_n)}{\int\limits_{\Theta} g(t,T)h(x_1,\ldots,x_n)dt}\cdot g(t,T) = \frac{g(t,T)}{\int\limits_{\Theta} g(t,T)dt}.$$

Таким образом, параметры апостериорного распределения выражаются через достаточную статистику T.

В задаче 1 в качестве такой статистики выступает $\overline{X}$. В задаче 2 в качестве такой статистики выступает $S_n^2 = \sum X_i^2$. В задаче 3 в качестве такой статистики выступает $X_{(n)}$. Все эти оценки возникают как оценки максимального правдоподобия или пропорциональные им (задача 2) и были посчитаны в 3-й домашней работе (такие оценки являются достаточными).