Іспит
з Математичної Статистики

Захаров Дмитро

10 грудня, 2024

Варіант 5

Зміст

1 Питання 1. Багатовимірний нормальний розподіл

Умова. Багатовимірний нормальний закон розподілу: теорема про щільність розподілу. Приклад.

Відповідь. В попередніх курсах ми вивчали одновимірний нормальний розподіл $\mathcal{N}(\mu, \sigma)$, що задано згідно густині $f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$. Проте, природа багатьох задач багатовимірна, тому природньо розширити цей розподіл на багатовимірний випадок. Як і з одновимірним випадком, спочатку розглянемо *стандартний* багатовимірний нормальний розподіл.

Definition 1.1. Випадковий вектор $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ має *стандартний багатовимірний нормальний розподіл*, якщо випадкові величини $\xi_1, \ldots, \xi_n$ є незалежними в сукупності випадковими величинами, що мають стандартний нормальний розподіл: $\xi_i \sim \mathcal{N}(0, 1)$.

Для такого випадку, за допомогою критерія незалежності неперервних випадкових величин, щільність розподілу $\boldsymbol{\xi}$ можна легко подати у вигляді:

$$f_\xi(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\mathbf{x}^\top \mathbf{x}}.$$

В таку випадку, ми формально записуємо $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{E}_{n\times n})$, щоб підкреслити, що вектор $\boldsymbol{\xi}$ має нульовий вектор середніх та одиничну коваріаційну матрицю — тобто має стандартний багатовимірний нормальний розподіл. Введемо поняття загального багатовимірного нормального розподілу.

Definition 1.2. Вважаємо, що n-вимірний випадковий вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ має нормальний розподіл, якщо існують вектор $\boldsymbol{\beta} \in \mathbb{R}^n$ та невироджена квадратна матриця $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ така, що $\boldsymbol{\xi} = \boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\eta}$, де $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{E}_{n\times n})$.

Проте, яка густина буде у такого розподілу? Для цього наводимо **теорему про щільність багатовимірного нормального розподілу**.

Theorem 1.3. Випадковий вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ має нормальний розподіл тоді і тільки тоді, коли його щільність має вигляд:

$$f_\xi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \cdot (\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\}.$$

Тут, $\Sigma \in \mathbb{R}^{n\times n}$ — симетрична додатно визначена матриця. Причому, якщо $\boldsymbol{\xi} = \boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\eta}$, то $\boldsymbol{\mu} = \boldsymbol{\beta}$ та $\Sigma = \boldsymbol{P}\boldsymbol{P}^\top$. Більш того, $\mathrm{Cov}[\boldsymbol{\xi}] = \Sigma$ та $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\xi}]$.

Доведення. Доведемо **необхідність**. Отже нехай $\boldsymbol{\xi} = \boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\eta}$, де $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{E}_{n\times n})$. Для обчислення щільності $\boldsymbol{\xi}$, розглянемо множину $\Pi(\mathbf{x}) = (-\infty, x_1) \times \cdots \times (-\infty, x_n)$ і множину $\Pi(\mathbf{x}) - \boldsymbol{\beta} = \{\mathbf{z} - \boldsymbol{\beta} : \mathbf{z} \in \Pi(\mathbf{x})\}$. Побудуємо функцію розподілу $\boldsymbol{\xi}$:

$$F_\xi(\mathbf{x}) = \Pr[\boldsymbol{\xi} \in \Pi(\mathbf{x})] = \Pr[\boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\eta} \in \Pi(\mathbf{x})] = \Pr[\boldsymbol{\eta} \in \boldsymbol{P}^{-1}(\Pi(\mathbf{x}) - \boldsymbol{\beta})]$$

Далі скористаємось означенням функції розподілу:

$$F_\xi(\mathbf{x}) = \int_{\boldsymbol{P}^{-1}(\Pi(x)-\boldsymbol{\beta})} f_\eta(\mathbf{y})d\mathbf{y} = \int_{\Pi(x)} f_\eta\left(\boldsymbol{P}^{-1}(\mathbf{u}-\boldsymbol{\beta})\right)|\det \boldsymbol{P}^{-1}|d\mathbf{u}$$

Тут ми використали заміну змінних $\mathbf{u} = \boldsymbol{\beta} + \boldsymbol{P}\mathbf{x}$ з Якобіаном $\det \boldsymbol{P}^{-1}$. Далі залишається лише продиференціювати вираз:

$$\begin{aligned}
f_\xi(\mathbf{x}) &= \frac{\partial F_\xi}{\partial \mathbf{x}} = f_\eta(\boldsymbol{P}^{-1}(\mathbf{x}-\boldsymbol{\beta}))|\det \boldsymbol{P}^{-1}| \\
&= \frac{1}{|\det \boldsymbol{P}|(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{P}^{-1}(\mathbf{x}-\boldsymbol{\beta}))^\top \boldsymbol{P}^{-1}(\mathbf{x}-\boldsymbol{\beta})\right\} \\
&= \frac{1}{|\det \boldsymbol{P}|(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\beta})^\top (\boldsymbol{P}^{-1})^\top \boldsymbol{P}^{-1}(\mathbf{x}-\boldsymbol{\beta})\right\} \\
&= \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\beta})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\beta})\right\},
\end{aligned}$$

де $\Sigma = \boldsymbol{P}\boldsymbol{P}^\top$. Симетричність матриці Σ очевидна з її представлення, а додатна визначеність з того, що $\mathbf{x}^\top\Sigma\mathbf{x} = \mathbf{x}^\top\boldsymbol{P}\boldsymbol{P}^\top\mathbf{x} = (\boldsymbol{P}^\top\mathbf{x})^\top\boldsymbol{P}^\top\mathbf{x} = \|\boldsymbol{P}^\top\mathbf{x}\|_2^2 > 0$.

Доведемо **достатність**. Нехай $\boldsymbol{\xi}$ має щільність $f_\xi(\mathbf{x})$ з вигляду, як в теоремі. Діагоналізуємо матрицю коваріації: $\boldsymbol{U}^\top\Sigma\boldsymbol{U} = \Lambda$, де $\Lambda = \operatorname{diag}\{\lambda_1, \dots, \lambda_n\}$. Нехай $\Lambda^{1/2} := \operatorname{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\}$. В такому разі, $\Sigma = \boldsymbol{U}\Lambda^{1/2}\Lambda^{1/2}\boldsymbol{U}^\top = \boldsymbol{P}\boldsymbol{P}^\top$, де $\boldsymbol{P} = \boldsymbol{U}\Lambda^{1/2}$.

Розглянемо випадковий вектор $\boldsymbol{\eta} = -\boldsymbol{P}^{-1}\boldsymbol{\mu} + \boldsymbol{P}^{-1}\boldsymbol{\xi}$. Щільність $\boldsymbol{\xi}$ ми знаємо, потрібно знайти щільність $\boldsymbol{\eta}$. Для цього, достатньо провести ті самі викладки, що і у випадку необхідності, себто:

$$F_\eta(\mathbf{x}) = \Pr[\boldsymbol{\xi} \in \boldsymbol{P}(\Pi(\mathbf{x}) + \boldsymbol{P}^{-1}\boldsymbol{\mu})] = \int_{\Pi(x)} f_\xi(\boldsymbol{P}(\mathbf{u} + \boldsymbol{P}^{-1}\boldsymbol{\mu}))|\det \boldsymbol{P}|d\mathbf{u}$$

В такому разі густина:

$$f_\eta(\mathbf{x}) = \frac{\partial F_\eta}{\partial \mathbf{x}} = f_\xi(\boldsymbol{P}\mathbf{x} + \boldsymbol{\mu})|\det \boldsymbol{P}|$$

Після підстановки відомої щільності, отримаємо:

$$f_\eta(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top\Lambda^{1/2}\Lambda^{-1}\Lambda^{1/2}\mathbf{x}\right\} = \frac{1}{(2\pi)^{n/2}}e^{-x^\top x/2},$$

звідки $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{E}_{n\times n})$. Нарешті, залишилось знайти математичне сподівання і коваріаційну матрицю:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\xi}] &= \boldsymbol{\mu} + \mathbb{E}[\boldsymbol{P}\boldsymbol{\eta}] = \boldsymbol{\mu} + \boldsymbol{P}\mathbb{E}[\boldsymbol{\eta}] = \boldsymbol{\mu}, \\
\operatorname{Cov}[\boldsymbol{\xi}] &= \operatorname{Cov}[\boldsymbol{\mu} + \boldsymbol{P}\boldsymbol{\eta}] = \operatorname{Cov}[\boldsymbol{P}\boldsymbol{\eta}] = \boldsymbol{P}\operatorname{Cov}[\boldsymbol{\eta}]\boldsymbol{P}^\top = \boldsymbol{P}\boldsymbol{P}^\top = \Sigma.
\end{aligned}$$

Example. Для двовимірного випадкового вектора $\boldsymbol{\xi} = (\xi_1, \xi_2)$ з математичними сподіваннями μ_1, μ_2 та дисперсіями σ_1^2, σ_2^2 і коефіцієнтом кореляції ρ, маємо матрицю коваріації $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ та вектор $\boldsymbol{\mu} = (\mu_1, \mu_2)$. Густина:

$$f_\xi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1-\rho^2)}\right\}, \quad z_i = \frac{x_i - \mu_i}{\sigma_i}, \quad i = 1, 2.$$

2 Питання 2. Довірчий інтервал

Умова. При $N = 1000$ випробуваннях Бернуллі подія E відбулась $m = 40$ разів. Для довірчої ймовірності $\alpha = 0.9$ побудувати довірчий інтервал $\mathcal{I}_\alpha$ для ймовірності p події.

Відповідь. Як було доведено на лекціях, довічний інтервал для ймовірності p настання події E в схемі Бернуллі з N випробуваннями має вигляд:

$$\mathcal{I}_\alpha = \left(\frac{N\hat{p} + \frac{1}{2}z_{\alpha/2}^2 - z_{\alpha/2}\sqrt{N\hat{p}(1-\hat{p}) + \frac{1}{4}z_{\alpha/2}^2}}{N + z_{\alpha/2}^2}, \frac{N\hat{p} + \frac{1}{2}z_{\alpha/2}^2 + z_{\alpha/2}\sqrt{N\hat{p}(1-\hat{p}) + \frac{1}{4}z_{\alpha/2}^2}}{N + z_{\alpha/2}^2} \right),$$

де $z_{\alpha/2} = \Phi_0^{-1}\left(\frac{\alpha}{2}\right)$, $\hat{p} = \frac{m}{N}$. Залишається підставити значення. Маємо $\hat{p} = \frac{40}{1000} = 0.04$. Значення $\alpha/2 = 0.45$, а отже за таблицею $z_{\alpha/2} = \Phi_0^{-1}(0.45) \approx 1.65$. Всі числа є, підставляємо. Якщо $\mathcal{I}_\alpha = (\ell_\alpha, u_\alpha)$, то:

$$\ell_\alpha = \frac{1000 \cdot 0.04 + \frac{1}{2} \cdot 1.65^2 - 1.65 \cdot \sqrt{1000 \cdot 0.04(1 - 0.04) + \frac{1}{4} \cdot 1.65^2}}{1000 + 1.65^2} \approx 0.031,$$
$$u_\alpha = \frac{1000 \cdot 0.04 + \frac{1}{2} \cdot 1.65^2 + 1.65 \cdot \sqrt{1000 \cdot 0.04(1 - 0.04) + \frac{1}{4} \cdot 1.65^2}}{1000 + 1.65^2} \approx 0.052.$$

Таким чином, довірчий інтервал має вигляд $\Pr[0.031 < p < 0.052] = 0.9$.
Відповідь. $\mathcal{I}_{0.9} = (0.031, 0.052)$.

Коментар. Відмітимо, що $N \gg z_{\alpha/2}$, тому формулу інтервалу можна спростити до:

$$\mathcal{I}_\alpha \approx \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right)$$

Ідея виведення формули. На лекції доведено, що $\lim_{n\to\infty} \Pr\left[\frac{\hat{p}-p}{\sqrt{p(1-p)/\sqrt{n}}} < x\right] = \Phi(x)$, себто випадкова величина $\xi = \sqrt{n}(\hat{p}-p)/\sqrt{p(1-p)}$ має асимптотичний стандартний нормальний розподіл. Задамо довірчу ймовірність α; з рівності $\Pr[|\xi| < z_{\alpha/2}] = \alpha$ для $z_{\alpha/2} = \Phi_0^{-1}(\alpha/2)$ по суті все зводиться до перевірки того, що $|\xi| < z_{\alpha/2}$. Або, якщо підставити ξ, то маємо $|\sqrt{n}(\hat{p} - p)/\sqrt{p(1-p)}| < z_{\alpha/2}$. Це аналогічно перевірці $n(\hat{p} - p)^2 \leq p(1-p)z_{\alpha/2}^2$, звідки і отримуємо формулу з розв'язання.

3 Питання 3. Статистична значущість кореляції

Умова. З двовимірної генеральної сукупності $(X, Y)^\top$ вилучено вибірку об'єму $n = 20$ і одержано вибіркову оцінку коефіцієнту кореляції: $\overline{r} = -0.2$. Перевірити гіпотезу про статистичну значущість коефіцієнту кореляції між випадковими величинами X, Y.

Відповідь. З'ясуємо, чи значуще коефіцієнт кореляції відрізняється від 0. Вводимо у розглядання дві гіпотези:

1. $\mathcal{H}_0$: $r = 0$ (основна гіпотеза).
2. $\mathcal{H}_1$: $r \neq 0$ (альтернативна гіпотеза).

Вводимо статистику:

$$t = \frac{\overline{r}\sqrt{n-2}}{\sqrt{1-\overline{r}^2}}.$$

Якщо справедлива гіпотеза $\mathcal{H}_0$, то статистика t розподілена за законом розподілу Стьюдента з $n-2$ ступенями свободи. Введемо рівень значущості q. Правило перевірки наступне.

Правило перевірки. Якщо $|t| < t_{n-2,q}$, то з довірчою ймовірністю $\alpha = 1 - q$ приймаємо гіпотезу $\mathcal{H}_0$ про те, що коефіцієнт кореляції $r = 0$. Якщо ж $|t| \geq t_{n-2,q}$, то на рівні значущості q відхиляємо цю гіпотезу та приймаємо альтернативну гіпотезу $\mathcal{H}_1$ $(r \neq 0)$.

Обрахуємо значення статистики t:

$$t = \frac{-0.2\sqrt{20-2}}{\sqrt{1-0.2^2}} \approx -0.866, \quad t_{18,0.05} = 2.1.$$

Отже, $|t| = 0.866 < 2.1$. Отже, з довірчою ймовірністю 0.95 ми приймаємо гіпотезу $\mathcal{H}_0$ про те, що коефіцієнт кореляції $r = 0$.

4 Питання 4. Рівність середніх

Умова. З генеральної сукупності X, яка має нормальний закон розподілу з дисперсією $\sigma_X^2 = 4$, вилучено вибірку об'єму $n_X = 20$ і підраховано вибіркове середнє $\overline{\mu}_X = -2$. З генеральної сукупності Y, яка має нормальний закон розподілу зі середньоквадратичним відхиленням $\sigma_Y = 1$, вилучено вибірку об'єму $n_Y = 10$ і підраховано вибіркове середнє $\overline{\mu}_Y = -2.5$. Перевірити гіпотезу про рівність середніх сукупностей X, Y.

Відповідь. Нехай μ_X, μ_Y — середні сукупностей X, Y. Висуваємо дві гіпотези:

1. $\mathcal{H}_0$: $\mu_X = \mu_Y$ (основна гіпотеза).
2. $\mathcal{H}_1$: $\mu_X \neq \mu_Y$ (альтернативна гіпотеза).

Маємо випадок відомих дисперсій σ_X^2, σ_Y^2. Введемо випадкову величину:

$$\xi = \overline{\mu}_X - \overline{\mu}_Y$$

На лекції було доведено, що $\xi \sim \mathcal{N}\left(\mu_X - \mu_Y, \frac{1}{n_X}\sigma_X^2 + \frac{1}{n_Y}\sigma_Y^2\right)$. Таким чином, розглядаємо величину

$$z = \frac{\overline{\mu}_X - \overline{\mu}_Y - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n_X}\sigma_X^2 + \frac{1}{n_Y}\sigma_Y^2}} \sim \mathcal{N}(0, 1).$$

Отже, якщо гіпотеза $\mathcal{H}_0$ правильна, то $z = \frac{\overline{\mu}_X - \overline{\mu}_Y}{\sqrt{\frac{1}{n_X}\sigma_X^2 + \frac{1}{n_Y}\sigma_Y^2}} \sim \mathcal{N}(0, 1)$. Отже правило гіпотези наступне.

Правило гіпотези. Якщо $|z| < \Phi_0^{-1}(\alpha/2)$, то з довірчою ймовірністю α приймаємо гіпотезу $\mathcal{H}_0$ про те, що $\mu_X = \mu_Y$. Якщо ж $|z| \geq \Phi_0^{-1}(\alpha/2)$, то на рівні значущості $q = 1 - \alpha$ відхиляємо цю гіпотезу та приймаємо альтернативну гіпотезу $\mathcal{H}_1$ про те, що $\mu_X \neq \mu_Y$.

Оберемо довірчу ймовірність $\alpha = 0.95$. Для нашого конкретного випадку, маємо:

$$z = \frac{-2 + 2.5}{\sqrt{\frac{4}{20} + \frac{1}{10}}} \approx 0.913, \quad \Phi_0^{-1}\left(\frac{\alpha}{2}\right) = \Phi_0^{-1}(0.475) = 1.96.$$

Отже, $|z| = 0.913 < \Phi_0^{-1}(0.475) = 1.96$. Отже, з довірчою ймовірністю 0.95 ми приймаємо гіпотезу $\mathcal{H}_0$ про рівність середніх сукупностей X, Y.