

Перевірка статистичних гіпотез. Гіпотези на параметри нормального розподілу.

1 Теоретичні відомості

Що таке статистична гіпотеза? Види гіпотез.

Нехай $X = (X_1, \dots, X_n)$ – кратна вибірка з розподілом спостережень

$$F(t) = \mathbf{P}_\theta (X_1 < t), \theta \in \Theta,$$

де $X_j \in \mathbb{R}^d$. Статистичною гіпотезою називають припущення про розподіл спостережень F (припущення про можливі значення невідомого параметра θ).

Якщо статистична гіпотеза однозначно визначає розподіл, то таку гіпотезу називають *простою*. В іншому разі, статистична гіпотеза називається *складною*.

Приклад 1. Нехай в моделі вище $d = 1$, $X_j \sim N(\mu, \sigma^2)$, де $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ вважається невідомим. Розглянемо таку статистичну гіпотезу:

$$H_1 : \mu = 1, \sigma^2 = 1.$$

Гіпотеза H_1 однозначно визначає розподіл вибірки X (значення параметра задано повністю). Тепер розглянемо іншу статистичну гіпотезу:

$$H_2 : \mu = 1.$$

Гіпотеза H_2 неоднозначно визначає розподіл вибірки X : конкретних значень для $\sigma^2 > 0$ не вказано, тому для кожного значення σ^2 маємо інший розподіл даних (маємо іншу модель).

Приклад 2. Нехай в моделі вище $d = 1$, X_j є абсолютно неперервними в.в., $\theta = F(\cdot) \in \Theta$, де Θ є простором усіх неперервних функцій розподілу. Розглянемо такі гіпотези:

$$H_1 : F(t) = \mathbf{P}(\xi < t), \xi \sim U[0, 5],$$

$$H_2 : F(t) = \mathbf{P}(\xi < t), \xi \sim U[a, b], -\infty < a < b < \infty,$$

$$H_3 : F(t) \neq \mathbf{P}(\xi < t), \xi \sim U[a, b], -\infty < a < b < \infty.$$

Гіпотеза H_1 однозначно визначає розподіл вибірки X (вигляд функції розподілу $F(t)$ задано повністю відомою функцією $\mathbf{P}(\xi < t)$).

Гіпотеза H_2 неоднозначно визначає розподіл вибірки X (вважається, що $F(t)$ належить до сім'ї функцій розподілів, що відповідають рівномірно розподіленій на $[a, b]$ випадковій величині ξ , $-\infty < a < b < \infty$). При різних a, b маємо різні імовірнісні розподіли.

Гіпотеза H_3 неоднозначно визначає розподіл вибірки X ($F(t)$ не належить до вищеописаної сім'ї імовірнісних розподілів з H_2). Звісно, що крім рівномірного розподілу існують інші неперервні розподіли, наприклад $N(\mu, \sigma^2)$, для якого теж можна перебирати значення μ, σ^2 .

Задача перевірки двох гіпотез. Статистичний тест

Продовжуємо розглядати кратну вибірку $X = (X_1, \dots, X_n)$ з розподілом спостережень

$$F(t) = \mathbf{P}_\theta(X_1 < t), \quad \theta \in \Theta,$$

де $X_j \in \mathbb{R}^d$ та θ вважається невідомим, S є вибіркоvim простором (множина, в якій X приймає значення незалежно від θ).

Для θ потрібно перевірити наступні статистичні гіпотези:

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1,$$

де $\Theta_j \subset \Theta$, $j = 0, 1$ та $\Theta_0 \cap \Theta_1 = \emptyset$.

Задачею перевірки статистичних гіпотез є побудова такого *правила*, яке на основі вхідних даних $X = x$ приймає одну з висунутих гіпотез, та його безпосереднє використання на досліджуваних даних (прийняття висновку щодо розподілу даних).

Гіпотезу H_0 називають основною гіпотезою, а H_1 – альтернативою.

Статистичним правилом будемо називати функцію $\pi(x) : S \rightarrow \{0, 1\}$. Якщо $\pi(x) = 0$, тоді тест приймає основну гіпотезу. Інакше, якщо $\pi(x) = 1$, то тест приймає альтернативу.

Для побудови статистичного тесту π часто розглядається підхід, який базується на основі деякої допоміжної статистики $S(x)$, яка за виконання основної гіпотези H_0 імовірно прийматиме значення в одній множині, а за виконання альтернативи H_1 – в іншій множині.

Тобто ідея полягає в тому, що тест буде дивитися на інформацію відносно $S(x)$: чи буде достатньо підстав для прийняття основної гіпотези або її відхилення.

Введемо множину D – така множина значень $S(x)$, коли треба приймати альтернативу. Цю множину називають критичною областю тесту. Тоді статистичний тест $\pi(x)$ на основі $(S(x), D)$ можна побудувати так:

$$\pi(x) = \begin{cases} 0, & S(x) \notin D, \\ 1, & S(x) \in D. \end{cases} \quad (1)$$

Приклад (статистичного тесту). У рамках вищеописаної моделі, $d = 1$ та $X_j \sim N(\mu, \sigma_0^2)$, де μ вважається невідомим, а $\sigma_0^2 > 0$ відоме. На параметр μ висунуто такі дві гіпотези:

$$H_0 : \mu < 0, \quad H_1 : \mu \geq 0.$$

Можемо побудувати статистичний тест з наступних міркувань. Оцінити μ ми можемо використавши вибіркове середнє $S(X) = \frac{1}{n} \sum_{j=1}^n X_j$. Оцінка $S(X)$ для досить великого обсягу n зосереджується навколо μ , відносно якого висунуто припущення. Вважаємо, що коли $S(X) < 0$, то варто приймати основну гіпотезу, інакше – альтернативу. Отже,

$$\pi(x) = \begin{cases} 0, & S(x) < 0, \\ 1, & S(x) \geq 0, \end{cases} = \begin{cases} 0, & S(x) \notin D, \\ 1, & S(x) \in D, \end{cases} \quad x \in \mathbb{R}^n,$$

де $D = [0, \infty)$ – критична область тесту. Подобається вам такий тест?

Характеризація якості статистичних тестів

Тут мова піде про те, наскільки хорошим є статистичний тест $\pi(x)$ для перевірки двох гіпотез.

Помилки першого та другого роду. Імовірності помилок

Ніхто не виключає випадок, коли тест $\pi(x)$ може помилково відпрацювати, тобто прийняти хибну гіпотезу.

Помилкою першого роду тесту $\pi(x)$ називають прийняття тестом альтернативи H_1 , коли вірна основна гіпотеза H_0 . Імовірністю помилки першого роду відповідно є

$$\alpha_{\pi}(\theta) = \mathbf{P}_{\theta}(\pi(X) = 1), \theta \in \Theta_0.$$

Помилкою другого роду тесту $\pi(x)$ називають прийняття тестом основної гіпотези H_0 , коли вірна альтернатива H_1 . Імовірністю помилки другого роду відповідно є

$$\beta_{\pi}(\theta) = \mathbf{P}_{\theta}(\pi(X) = 0), \theta \in \Theta_1.$$

Приклад (помилки тесту). Нехай лікар має деякий тест $\pi(x)$ для перевірки двох гіпотез:

H_0 : Вагітна людина, H_1 : Не вагітна людина.

Тобто, якщо $\pi(x) = 0$, то тест вважає що людина не вагітна, інакше вважає вагітною. Тоді

- Приклад помилки першого роду: чоловік проходить тест на вагітність та в результаті є підстави вважати, що він вагітний.
- Приклад помилки другого роду: вагітна жінка проходить тест на вагітність та в результаті недостатньо підстав для того, щоб вважати що вона вагітна.

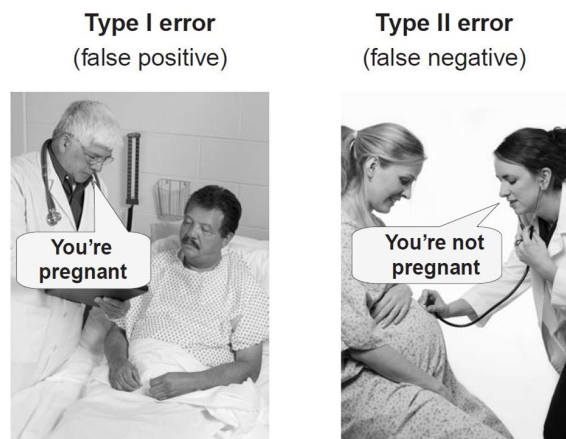


Рис. 1: Власне, приклад.

Рівень значущості тесту. Потужність тесту

Рівнем значущості тесту $\pi(x)$ називається найбільш можлива імовірність помилки першого роду:

$$\alpha_{\pi}^* = \sup_{\theta \in \Theta_0} \alpha_{\pi}(\theta).$$

Потужністю тесту $\pi(x)$ називають імовірність правильно прийняти альтернативу:

$$\varphi_{\pi}(\theta) = 1 - \beta_{\pi}(\theta) = \mathbf{P}_{\theta}(\pi(X) = 1), \theta \in \Theta_1.$$

Для вибору кращого статистичного тесту можна іти з міркувань збільшення його потужності, фіксуючи при цьому рівень значущості $\alpha_{\pi}^* := \alpha^*$ тесту.

Про досягнутий рівень значущості

Розглянемо статистичний тест $\pi(x)$ з (1). Нехай за вибіркою $X = x$ маємо деяке значення статистики тесту $S(x) = \tilde{s}$.

Є така штука, як досягнутий рівень значущості статистичного тесту (або ж p-value). Це імовірність того, що статистика тесту $S(X)$ буде 'перевищувати' обчислене значення \tilde{s} , коли основна гіпотеза вірна. Приклади, коли $S(x) \in \mathbb{R}$:

1. Якщо $D = (c, \infty)$, тоді p-value = $\mathbf{P}_{\theta}(S(X) > \tilde{s})$, $\theta \in \Theta_0$,
2. Якщо $D = (-\infty, c)$, тоді p-value = $\mathbf{P}_{\theta}(S(X) < \tilde{s})$, $\theta \in \Theta_0$,
3. Якщо $D = (-\infty, c) \cup (c, +\infty)$, тоді p-value = $\mathbf{P}_{\theta}(|S(X)| > |\tilde{s}|)$, $\theta \in \Theta_0$.

Тобто наскільки імовірно отримати значення статистики 'більше' за обчислене, коли основна гіпотеза вірна. Тоді, в термінах досягнутого рівня значущості p-value,

$$\pi = \begin{cases} 0, & \text{p-value} > \alpha^*, \\ 1, & \text{p-value} \leq \alpha^*. \end{cases}$$

Перейдемо до задач на перевірку гіпотез про параметри нормального розподілу.

2 Задачі

2.1 Задача 1

Номинальний опір резисторів 2000 Ом. Для контролю добрана партія кількістю 12 резисторів. Після вимірювання опору кожного зразка з середньоквадратичним відхиленням 5 Ом отримано такі значення:

2130, 2090, 2030, 2080, 1920, 2020, 2015, 2000, 2045, 1940, 1980, 1970.

Чи можна відхилення від номіналу (2000 Ом) розглядати як випадкові (допустимі), чи, навпаки, результати свідчать про те, що опір резисторів істотно відрізняється від номіналу? Використати нормальну модель.

Розв'язання.

За умовою задачі, маємо кратну вибірку з $n = 12$ спостережень, де опір резисторів має нормальний розподіл $X_j \sim N(\mu, \sigma^2)$, де μ – невідома та $\sigma^2 = 5^2 = 25$.

В задачі питають, чи істотно відрізняється від номіналу опір резисторів. В термінах моделі, якщо різниці від номінального значення немає, то $\mu = 2000$ Ом.

Тоді відповісти на питання задачі – це перевірити такі дві статистичні гіпотези:

$$H_0 : \mu = 2000, H_1 : \mu \neq 2000.$$

Побудуємо статистичний тест використовуючи статистику від нормальних спостережень. Відомо, що вибіркове середнє $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ при належному нормуванні має розподіл

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0, 1).$$

В якості статистики тесту розглянемо

$$S(X) = \sqrt{n}(\bar{X}_n - 2000)/\sigma.$$

Виберемо критичну область статистики $S(X)$. Якщо основна гіпотеза H_0 вірна, тоді можна очікувати, що значення $S(X)$ будуть несуттєво відрізнятися від нуля. Критичними будуть ті значення статистики $S(X)$, які будуть 'досить' великими.

Легко бачити, що за виконання основної гіпотези $S(X) \sim N(0, 1)$. Також відмітимо, що за виконання основної гіпотези, $S(X)$ потрапляє в інтервал $(-z_\alpha, z_\alpha)$ з імовірністю $1 - \alpha$, де $z_\alpha = Q^{N(0,1)}(1 - \alpha/2)$. Вважатимемо підставою приймати альтернативу буде ситуація, коли $S(X)$ буде лежати за межами $(-z_\alpha, z_\alpha)$.

Отже, статистичний тест для перевірки висунутих гіпотез має вигляд:

$$\pi(x) = \begin{cases} 0, & S(x) \in (-z_\alpha, z_\alpha), \\ 1, & S(x) \in (-\infty, -z_\alpha] \cup [z_\alpha, +\infty). \end{cases}$$

Рівнем значущості в цій задачі є $\mathbf{P}(|N(0, 1)| > z_\alpha) = \alpha$ (побачити чому). Будемо вважати, що $\alpha = 0.05$. Обчислимо за наявними даними $X = x$ статистику та поріг тесту:

$$\begin{aligned} z_\alpha &\approx 1.959964, \\ \bar{X}_n &\approx 2018.333, \\ \sqrt{n} &\approx 3.464102, \\ \Rightarrow S(X) &\approx 12.70171. \end{aligned}$$

Легко бачити, що $S(X) \approx 12.70171 > 1.959964 = z_\alpha$, тобто значення $S(X)$ потрапляє в критичну область тесту. Отже, з рівнем значущості $\alpha = 0.05$ ми не можемо відхилити альтернативу H_1 (маємо підстави на її користь).

2.2 Задача 2

В результаті вимірювання 10 зразків твердості сплаву отримано значення в умовних одиницях:

$$12.1, 13.7, 11.0, 11.6, 11.9, 13.9, 11.5, 12.9, 13.0, 10.5.$$

Припустимо, що твердість сплаву розподілена нормально. Чи можна вважати, що дисперсія розподілу твердості становить 2.25?

Розв'язання.

В задачі сказано, що маємо кратну вибірку з $n = 10$ спостережень, де розподіл твердості сплаву є гауссовим $X_j \sim N(\mu, \sigma^2)$, де μ та σ^2 невідомі. Відповідти на питання задачі – це перевірити гіпотези на параметр σ^2 :

$$H_0 : \sigma^2 = 2.25, H_1 : \sigma^2 \neq 2.25.$$

Побудуємо статистичний тест використовуючи статистики від нормальних спостережень. Відомо, що вибіркова дисперсія $\hat{s}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \frac{1}{n} \sum_{i=1}^n X_i)^2$ при належному нормуванні має розподіл

$$\frac{n-1}{\sigma^2} \hat{s}_n^2 \sim \chi_{n-1}^2.$$

В якості статистику тесту розглянемо

$$S(X) = \frac{n-1}{2.25} \hat{s}_n^2.$$

Очевидно, що за виконання основної гіпотези H_0 , $S(X) \sim \chi_{n-1}^2$. Підберемо такий інтервал (h_-, h_+) , в який $S(X)$ за виконання H_0 потрапить з імовірністю $1 - \alpha$:

$$\mathbf{P}_\theta(h_- < S(X) < h_+) = 1 - \alpha.$$

Вважаємо, що імовірності вийти за межі інтервалу у лівий та правий бік однакові. Тоді

$$h_- = Q^{\chi_{n-1}^2}(\alpha/2), h_+ = Q^{\chi_{n-1}^2}(1 - \alpha/2).$$

Побудуємо на основі цього статистичний тест. Підставою прийняти альтернативу буде ситуація, коли значення $S(X)$ лежатиме за межами інтервалу (h_-, h_+) .

Тобто

$$\pi(X) = \begin{cases} 0, & S(x) \in (h_-, h_+), \\ 1, & S(x) \in (0, h_-] \cup [h_+, +\infty). \end{cases}$$

Рівень значущості такого тесту є $\mathbf{P}(\chi_{n-1}^2 \leq h_-) + \mathbf{P}(\chi_{n-1}^2 \geq h_+) = \alpha/2 + \alpha/2 = \alpha$. Покладемо рівень значущості $\alpha := 0.05$. Обчислимо за наявними даними $X = x$ статистику та поріг тесту:

$$\begin{aligned} h_- &\approx 2.700389, \quad h_+ \approx 19.02277, \\ n - 1 &= 9, \\ \hat{s}_n^2 &\approx 1.283222, \\ S(X) &\approx 5.132889. \end{aligned}$$

Значення статистики тесту лежить за межами критичної області. Отже, маємо підстави для прийняття основної гіпотези H_0 з рівнем значущості $\alpha = 0.05$.