Міністерство освіти і науки України
Прикарпатський національний університет
імені Василя Стефаника

Р.В. Шевчук

МАТЕМАТИЧНА СТАТИСТИКА

Конспект лекцій
для студентів напрямів підготовки
«статистика», «математика» та
«прикладна математика»
вищих навчальних закладів

Івано-Франківськ
2015

Шевчук Р.В. Математична статистика. Конспект лекцій. – Івано-Франківськ: ???, 2015. – 80 с.

У вигляді конспекту лекцій викладено основи математичної статистики. Матеріал поділено на чотири розділи: «Точкові оцінки», «Інтервальні оцінки», «Перевірка статистичних гіпотез», «Нормальна лінійна регресія». У перших двох розділах відображено загальні поняття і методи теорії оцінювання параметрів розподілів. Третій розділ знайомить читача з теорією перевірки статистичних гіпотез, а четвертий розділ — з елементарними поняттями регресійного аналізу.

Конспект лекцій призначений для студентів напрямів підготовки «статистика», «математика» та «прикладна математика» вищих навчальних закладів.

Розділ 1

Точкові оцінки

1.1 Статистики та оцінки

Задача математичної статистики — отримати конкретні висновки з есперементальних даних. Вихідним матеріалом для статистичного дослідження реального явища служить набір результатів спостережень над ним або ж результатів спеціально поставлених експериментів. У випадку, коли результати спотережень є числовими, їх можна розглядати як значення випадкової величини. Ця випадкова величина може бути як дискретною, так і неперервною.

Нехай в стохастичному експерименті спостерігається деяка випадкова величина. Повторимо цей експеримент в незмінних умовах n разів. З організованим таким чином дослідженням пов'язаний n-вимірний випадковий вектор

$$\mathbf{X} = (X_1, X_2, \ldots, X_n),$$

де випадкові величини $X_i,\ i = \overline{1,n}$, незалежні в сукупності і однаково розподілені.

Означення. Випадковий вектор $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ зі значеннями в просторі $\mathbb{R}^n$, утворений послідовністю незалежних, однаково розподілених випадкових величин X_1, $X_2, \ldots, X_n$, кожна з яких має розподіл G, називають вибіркою обсягом n із розподілу G.

Простір $\mathbb{R}^n$, у якому вибірка набуває значень, будемо називати вибірковим простором.

Далі матимемо справу з вибірками, розподіли (функції розподілу) яких залежать від невідомого параметра θ, який може набувати довільних значень з деякої множини можливих значень $\Theta \subset \mathbb{R}$.

Задача оцінювання параметрів розподілів. Нехай $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ — реалізація вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ з розподілу $F(\cdot;\theta)$. Значення параметра θ в розподілі $F(\cdot;\theta)$ невідоме, і його необхідно оцінити (наближено визначити) за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$. У цьому й полягає задача оцінювання параметрів розподілів.

Сформульовану задачу можна розглядати як задачу відшукування такої функції $h : \mathbb{R}^n \to \Theta$, що $h(\mathbf{x}) \approx \theta$, де $\mathbf{x}$ — реалізація вибірки $\mathbf{X}$. Значення $h(\mathbf{x})$ ми й будемо використовувати як θ. Оскільки для кожної реалізації $\mathbf{x}$ значення $h(\mathbf{x})$, яке використовується як θ, буде своє, оцінку параметра θ можна вважати випадковою величиною.

Означення. Борелєву функцію $h(\cdot)$, задану на вибірковому просторі $\mathbb{R}^n$, зі значеннями в Θ — множині можливих значень параметра θ — будемо називати статистикою, а $\hat{\theta} = h(\mathbf{X}) = h(X_1, X_2, \ldots, X_n)$ — борелєву функцію від вибірки зі значеннями в Θ — оцінкою (точковою оцінкою) параметра θ.

1.2 Властивості оцінок

Основне питання задачі оцінювання параметрів розподілів — наскільки великою є похибка при використанні оцінки $\hat{\theta}$ як значення θ. Кількісно міру похибки при заміні θ на $\hat{\theta}$ (міру розсіювання $\hat{\theta}$ відносно θ) будемо описувати величиною

$$\delta^2 = \mathrm{M}|\hat{\theta} - \theta|^2,$$

яку називають середньоквадратичною похибкою оцінки $\hat{\theta}$.

Серед усіх оцінок з однією і тією ж дисперсією $\mathrm{D}\hat{\theta}$ найменшу середньоквадратичну похибку мають оцінки, для яких $\mathrm{M}\hat{\theta} = \theta$. Останнє випливає з рівностей

$$\delta^2 = \mathrm{M}|\hat{\theta} - \theta|^2 = \mathrm{M}(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) = \mathrm{M}\hat{\theta}^2 - 2\theta\mathrm{M}\hat{\theta} + \theta^2 =$$
$$= \mathrm{M}\hat{\theta}^2 - (\mathrm{M}\hat{\theta})^2 + (\mathrm{M}\hat{\theta})^2 - 2\theta\mathrm{M}\hat{\theta} + \theta^2 = \mathrm{D}\hat{\theta} + (\mathrm{M}\hat{\theta} - \theta)^2.$$

Відзначимо також, що якщо $\mathrm{M}\hat{\theta} = \theta$, то середньоквадратична похибка δ^2 перетворюється в дисперсію оцінки $\hat{\theta}$.

Означення. Оцінка $\hat{\theta}$ параметра θ називається незміщеною, якщо $\mathrm{M}\hat{\theta} = \theta$.

Наочно незміщеність оцінки $\hat{\theta}$ параметра θ означає, що при багаторазовому використанні оцінки $\hat{\theta}$ як значення θ, тобто, при багаторазовій заміні θ на $\hat{\theta}$, середнє значення похибки $\hat{\theta} - \theta$ дорівнює нулеві. Незміщеність вказує на відсутність систематичної похибки.

Для оцінювання параметра θ можна запропонувати багато оцінок $\hat{\theta}$ з $\mathrm{M}\hat{\theta} = \theta$. Із сукупності таких оцінок природно вибрати ті, що мають найменшу дисперсію.

Означення. Якщо в деякому класі незміщених оцінок параметра θ, які мають скінченні дисперсії, існує така оцінка $\hat{\theta}$, що $\mathrm{D}\hat{\theta} \leq \mathrm{D}\tilde{\theta}$ для всіх оцінок $\tilde{\theta}$ із цього класу, то кажуть, що оцінка $\hat{\theta}$ є ефективною в даному класі оцінок.

Іншими словами, дисперсія ефективної оцінки параметра в деякому класі незміщених оцінок є мінімальною серед дисперсій всіх оцінок з цього класу.

Ефективну оцінку в класі всіх незміщених оцінок будемо називати просто ефективною оцінкою (без слів «в класі незміщених оцінок»). У літературі з математичної статистики замість терміну «ефективна оцінка» можна також зустріти такі назви, як «незміщена оцінка з мінімальною дисперсією» або «оптимальна оцінка».

Теорема (про єдиність ефективної оцінки). *Якщо* $\hat{\theta}$ *і* $\tilde{\theta}$ *— ефективні оцінки параметра* θ*, то з ймовірністю 1 вони співпадають:* $\mathrm{P}\{\hat{\theta} = \tilde{\theta}\} = 1$.

◀ За умовою теореми $\mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta}$. Нехай $\theta^* = \frac{1}{2}(\hat{\theta} + \tilde{\theta})$. Тоді

$$\mathrm{D}\theta^* = \frac{1}{4}(\mathrm{D}\hat{\theta} + \mathrm{D}\tilde{\theta}) + \frac{1}{2}\mathrm{cov}(\hat{\theta}, \tilde{\theta}) = \frac{1}{2}(\mathrm{D}\hat{\theta} + \mathrm{cov}(\hat{\theta}, \tilde{\theta})).$$

Оскільки

$$|\mathrm{cov}(\hat{\theta}, \tilde{\theta})| \le \sqrt{\mathrm{D}\hat{\theta}\mathrm{D}\tilde{\theta}} = \mathrm{D}\hat{\theta},$$

виконується нерівність

$$\mathrm{D}\theta^* = \frac{1}{2}|\mathrm{D}\hat{\theta} + \mathrm{cov}(\hat{\theta}, \tilde{\theta})| \le \frac{1}{2}(\mathrm{D}\hat{\theta} + \mathrm{D}\hat{\theta}) \le \mathrm{D}\hat{\theta}.$$

З цієї нерівності і того, що $\hat{\theta}$ — ефективна оцінка, випливає, що $\mathrm{D}\theta^* = \mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta}$. Тоді

$$\mathrm{cov}(\hat{\theta}, \tilde{\theta}) = \mathrm{D}\hat{\theta} = \mathrm{D}\tilde{\theta} = \sqrt{\mathrm{D}\hat{\theta}\mathrm{D}\tilde{\theta}}.$$

Остання рівність дозволяє нам стверджувати, що випадкові величини $\hat{\theta}$ і $\tilde{\theta}$ з ймовірністю 1 лінійно залежні, тобто

$\mathrm{P}\{\hat{\theta} = k\tilde{\theta} + b\} = 1$ для деяких сталих k і b. Враховуючи те, що

$$\mathrm{D}\hat{\theta} = \mathrm{cov}(k\tilde{\theta} + b, \tilde{\theta}) = k\mathrm{D}\tilde{\theta} = k\mathrm{D}\hat{\theta},$$

отримуємо $k = 1$. З умови незміщеності оцінок випливає, що $b = 0$:

$$\mathrm{M}\tilde{\theta} = \mathrm{M}\hat{\theta} = \mathrm{M}(\tilde{\theta} + b) = \mathrm{M}\tilde{\theta} + b.$$

Таким чином, $\mathrm{P}\{\hat{\theta} = \tilde{\theta}\} = 1$. ▶

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі щільністю $f(x, \theta)$, $x \in \mathbb{R}$, $\theta \in \Theta$. Позначимо через

$$f_{\mathbf{X}}(\mathbf{x}; \theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; \theta)$$

щільність розподілу випадкового вектора $\mathbf{X}$. Далі, у цьому параграфі, при розгляді функції $f_{\mathbf{X}}(\mathbf{x}, \theta)$, будемо застосовувати диференціювання за параметром під знаком інтеграла, залежного від параметру. Тому будемо вважати, що Θ — інтервал на прямій $\mathbb{R}$, а функція $f_{\mathbf{X}}(\mathbf{x}, \theta)$, $\mathbf{x} \in \mathbb{R}^n$, $\theta \in \Theta$, задовольняє умови регулярності, які забезпечують законність вказаних операцій.

Означення. Функцію

$$I(\theta) = \mathrm{M}\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{X}; \theta)}{\partial \theta}\right)^2$$

(якщо вона визначена) називають кількістю інформації за Фішером.

Теорема (нерівність Крамера-Рао). Якщо $\hat{\theta}$ — незміщена оцінка параметра θ і виконуються умови регулярності, то має місце нерівність Крамера-Рао

$$\mathrm{D}\hat{\theta} \geq \frac{1}{I(\theta)}, \tag{1.1}$$

де $I(\theta)$ — кількість інформації за Фішером.

◀ Нехай $f(x;\theta) > 0$ при $x \in A \subset \mathbb{R}$ і $f(x;\theta) = 0$ при $x \notin A$. Тоді щільність

$$f_{\mathbf{X}}(\mathbf{x},\theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

розподілу вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ відмінна від нуля на множині $B = A \times A \times \ldots \times A \subset \mathbb{R}^n$. Оскільки

$$\int\limits_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 1,$$

то

$$\frac{\partial}{\partial\theta} \int\limits_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \frac{\partial}{\partial\theta} \int\limits_{B} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} \frac{\partial f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} d\mathbf{x} =$$

$$= \int\limits_{B} \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 0. \tag{1.2}$$

З умови незміщеності оцінки $\hat{\theta} = h(\mathbf{X}) = h(X_1,\ X_2, \ldots, X_n)$ випливає, що

$$\mathrm{M}\hat{\theta} = \int\limits_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \int\limits_{B} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \theta.$$

Тоді

$$\frac{\partial}{\partial\theta} \int\limits_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = \frac{\partial}{\partial\theta} \int\limits_{B} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} =$$

$$= \int\limits_{B} h(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(x,\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta) d\mathbf{x} = 1. \tag{1.3}$$

Віднявши від (1.3) рівність (1.2), помножену на параметр θ, отримуємо

$$\int\limits_B (h(\mathbf{x})-\theta)\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} f_{\mathbf{X}}(\mathbf{x},\theta)d\mathbf{x} = 1.$$

Скориставшись нерівністю Коші-Буняковського, встановлюємо, що

$$1 \le \int\limits_B (h(\mathbf{x})-\theta)^2 f_{\mathbf{X}}(\mathbf{x},\theta)d\mathbf{x}\times$$
$$\times\int\limits_B \left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta}\right)^2 f_{\mathbf{X}}(\mathbf{x},\theta)d\mathbf{x} = \mathrm{D}\hat{\theta} I(\theta). \tag{1.4}$$

Це доводить твердження теореми. ▶

За умов регулярності нерівність (1.1) визначає нижню межу дисперсій всіх незміщених оцінок параметра θ. Величину

$$e(\theta) = \frac{1}{I(\theta)\mathrm{D}\hat{\theta}}$$

називають показником ефективності в розумінні Крамера-Рао. Із (1.1) випливає, що $0 < e(\theta) \le 1$ для будь-якої незміщеної оцінки параметра θ.

Означення. Незміщену оцінку $\hat{\theta}$ параметра θ називають ефективною в розумінні Крамера-Рао, якщо показник ефективності $e(\theta) = 1$.

Зауваження. Рівність

$$\mathrm{D}\hat{\theta} = \frac{1}{I(\theta)}$$

має місце тоді й тільки тоді, коли

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x},\theta)}{\partial\theta} = c(\theta)(h(\mathbf{x})-\theta) \tag{1.5}$$

при всіх $x \in \mathbb{R}^n$. Це випливає з необхідних і достатніх умов перетворення в рівність нерівності Коші-Буняковського. Останнє співвідношення, при виконанні умов регулярності, є критерієм ефективності незміщеної оцінки.

Зауваження. З теореми про єдиність ефективної оцінки в випливає, що ефективна оцінка в розумінні Крамера-Рао є ефективною.

Зауваження. Нерівність Крамера-Рао має місце і в дискретному випадку. Якщо розподіл випадкового вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ дискретний, то $f_{\mathbf{X}}(\mathbf{x}, \theta)$ — це ймовірність події $\{\mathbf{X} = \mathbf{x}\}$. При цьому $f_{\mathbf{X}}(\mathbf{x}, \theta) > 0$ лише для скінченної або зліченної множини точок $\mathbf{x} \in \mathbb{R}^n$. Доведення нерівності (1.1) у цьому випадку є повторенням доведення останньої теореми з очевидними змінами.

Часто можна розглядати не одну оцінку $\hat{\theta} = h(\mathbf{X}) = h(X_1, X_2, \ldots, X_n)$, побудовану за вибіркою $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, а послідовність оцінок $\hat{\theta}_n = h_n(X_1,\ X_2, \ldots,\ X_n)$, $n = 1, 2, \ldots$. У цій ситуації природно говорити про асимптотичну поведінку послідовності оцінок $\hat{\theta}_n$.

Означення. Послідовність оцінок $\hat{\theta}_n,\ n = 1, 2, \ldots$, параметра θ називається консистентною (слушною, змістовною), якщо вона збігається за ймовірністю до θ, тобто для кожного $\varepsilon > 0$ $\mathrm{P}\{|\hat{\theta}_n - \theta| > \varepsilon\} \to 0$ при $n \to \infty$.

Означення. Послідовність оцінок $\hat{\theta}_n,\ n = 1, 2, \ldots$, параметра θ називається асимптотично незміщеною, якщо $\mathrm{M}\hat{\theta}_n \to \theta$ при $n \to \infty$.

Означення. Послідовність незміщених оцінок $\hat{\theta}_n$, $n = 1, 2, \ldots$, параметра θ називається асимптотично ефективною, якщо $e(\theta) \to 1$ при $n \to \infty$.

Зауваження. Надалі поняття «оцінка» будемо використовувати також у широкому розумінні як послідовність оцінок.

Приклад (оцінка математичного сподівання нормального розподілу). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу з невідомим математичним сподіванням a і відомою дисперсією σ^2. Перевірити незміщеність, консистентність та ефективність оцінки

$$\hat{a} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

параметра a.

За умовою $\mathrm{M}X_i = a$, $i = \overline{1,n}$. Тоді, враховуючи лінійність математичного сподівання, отримуємо

$$\mathrm{M}\bar{X} = \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}X_i =$$
$$= \frac{1}{n}\sum_{i=1}^{n} a = \frac{1}{n} na = a, \tag{1.6}$$

а це й означає, що $\bar{X}$ є неміщеною оцінкою параметра a.

Оскільки $X_1, X_2, \ldots, X_n$ — незалежні та однаково розподілені випадкові величини зі скінченною дисперсією σ^2, то згідно з законом великих чисел у формі Чебишева для будь-якого $\varepsilon > 0$

$$\mathrm{P}\{|\bar{X} - a| > \varepsilon\} \to 0, \quad n \to \infty. \tag{1.7}$$

З останнього співвідношення випливає консистентність оцінки $\bar{X}$.

Покажемо тепер, що $\bar{X}$ є ефективною оцінкою параметра a. Для цього достатньо перевірити виконання рівності (1.5). Щільність розподілу випадкового вектора $\mathbf{X} = (X_1,$

$X_2, \ldots, X_n)$ має вигляд

$$f_{\mathbf{X}}(\mathbf{x}; a) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; a) =$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\} =$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right\}.$$

Тоді

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}; a)}{\partial a} = \frac{\partial}{\partial a}\left(-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right) =$$
$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial a}(x_i - a)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) =$$
$$= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - na\right) = \frac{n}{\sigma^2}(\bar{x} - a).$$

Звідси випливає, що $\hat{a} = \bar{X}$ — ефективна оцінка параметра a.

Приклад (оцінка дисперсії нормального розподілу при відомому математичному сподіванні). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу $N(a, \sigma^2)$, і нехай математичне сподівання a — відоме, а дисперсія σ^2 — невідома. Перевірити незміщеність, консистентність та ефективність оцінки

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2.$$

параметра σ^2.

Незміщеність оцінки $\hat{\sigma^2}$ випливає з наступних співвідношень:

$$\mathrm{M}\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}(X_i - a)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}(X_i - \mathrm{M}X_i)^2 =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2.$$

Покажемо, що $\hat{\sigma^2}$ є консистентною оцінкою. Згідно з другою нерівністю Чебишева

$$\mathrm{P}\left\{|\hat{\sigma^2} - \sigma^2| > \varepsilon\right\} \leq \frac{\mathrm{D}\hat{\sigma^2}}{\varepsilon^2}.$$

Знайдемо дисперсію оцінки $\hat{\sigma^2}$:

$$\mathrm{D}\hat{\sigma^2} = \mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2\right) =$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}(X_i - a)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}(X_i - \mathrm{M}X_i)^2 =$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\left(\mathrm{M}(X_i - \mathrm{M}X_i)^4 - \left(\mathrm{M}(X_i - \mathrm{M}X_i)^2\right)^2\right) =$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}(3\sigma^4 - \sigma^4) = \frac{2\sigma^4}{n}.$$

Тоді

$$\mathrm{P}\left\{|\hat{\sigma^2} - \sigma^2| > \varepsilon\right\} \leq \frac{2\sigma^4}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Це доводить консистентність оцінки $\hat{\sigma^2}$.

Доведемо ефективність оцінки $\hat{\sigma^2}$. Для цього позначимо $\theta = \sigma^2$ і перевіримо виконання рівності (1.5) для щільності $f_{\mathbf{X}}(\mathbf{x}; \theta)$, $\mathbf{x} \in \mathbb{R}^n$, випадкового вектора $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Враховуючи результат попереднього прикладу, маємо

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left(-\frac{n}{2} \ln(2\pi\theta) - \frac{1}{2\theta} \sum_{i=1}^{n} (x_i - a)^2 \right) =$$
$$= -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - a)^2 = \frac{n}{2\theta^2} \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - a)^2 - \theta \right).$$

З останньої рівності випливає, що $\hat{\sigma^2}$ — ефективна оцінка параметра σ^2.

Зауваження. Наведені у цьому параграфі властивості оцінок можна узагальнити на випадок, коли θ — векторний параметр, тобто $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

Вправа. Довести, що коли дисперсія незміщеної оцінки прямує до нуля при $n \to \infty$, то ця оцінка є конзистентною.

1.3 Емпірична функція розподілу

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу F. Припустимо, що функція розподілу $F(x) = \mathrm{P}\{X_i < x\}$, $x \in \mathbb{R}$, $i = \overline{1, n}$, невідома і її необхідно оцінити.

Означення. Емпіричною функцією розподілу називається функція $\hat{F}_n$ визначена на $\mathbb{R}$ рівністю

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i < x\}},$$

де $\mathbb{I}_{\{X_i < x\}}$ — індикатор події $\{X_i < x\}$.

Зауваження. При кожному фіксованому $x \in \mathbb{R}$ емпірична функція розподілу $\hat{F}_n(x)$ як функція випадкового вектора $(X_1, X_2, \ldots, X_n)$ є випадковою величиною.

У процесі побудови графіка емпіричної функції розподілу (а точніше її реалізації) ми стикаємося з задачею впорядкування точок її стрибків — вибіркових значень.

Нехай $(x_1, x_2, \ldots, x_n)$ — реалізація вибірки $(X_1, X_2, \ldots, X_n)$. Розташуємо числа $x_1, x_2, \ldots, x_n$ у порядку зростання:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}, \tag{1.8}$$

де $x_{(1)}$ — найменше, $x_{(n)}$ — найбільше серед значень x_1, $x_2, \ldots, x_n$.

Означення. Послідовність чисел $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, які задовольняють умову (1.8), називають варіаційним рядом послідовності $x_1, x_2, \ldots, x_n$ або просто варіаційним рядом. Число $x_{(i)}$, $i = \overline{1,n}$, називають i-м членом варіаційного ряду.

Позначимо через $X_{(i)}$, $i = \overline{1,n}$, випадкову величину, яка при кожній реалізації вибірки $(X_1, X_2, \ldots, X_n)$ набуває значення, що дорівнює i-му члену варіаційного ряду.

Означення. Послідовність випадкових величин $X_{(1)}$, $X_{(2)}, \ldots, X_{(n)}$ називають варіаційним рядом послідовності $X_1, X_2, \ldots, X_n$. При цьому $X_{(i)}$, $i = \overline{1,n}$, називають i-м членом варіаційного ряду вибірки.

Нехай серед членів варіаційного ряду $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ є r різних чисел, які також впорядковано за зростанням. Позначимо їх $z_{(1)}, z_{(2)}, \ldots, z_{(r)}$. Припустимо, що кожне з них повторюється відповідно $n_1, n_2, \ldots, n_r$ разів, при цьому очевидно, що $\sum\limits_{i=1}^{r} n_i = n$. Тоді реалізацію емпіричної функції розподілу $\hat{F}_n(x)$ можна подати у вигляді:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i:\ z_{(i)}<x} n_i = \begin{cases} 0, & x \le z_{(1)}; \\ \frac{n_1}{n}, & z_{(1)} < x \le z_{(2)}; \\ \frac{n_1+n_2}{n}, & z_{(2)} < x \le z_{(3)}; \\ \dots \\ 1, & x > z_{(r)}. \end{cases} \qquad (1.9)$$

Із вищенаведеного випливає, що при фіксованих значеннях вибірки, $\hat{F}_n(x)$ як функція аргументу x є дискретною функцією розподілу. Розподіл $\left(z_{(i)}, \frac{n_i}{n}\right)$, $i = \overline{1, r}$, що їй відповідає, називається емпіричним.

Теорема (про властивості емпіричної функції розподілу). Для кожного $x \in \mathbb{R}$ значення емпіричної функції розподілу $\hat{F}_n(x)$ має такі властивості:

1) випадкова величина $n\hat{F}_n(x)$ має біноміальний розподіл з парараметрами n, $F(x)$:
$$\mathrm{P}\left\{\hat{F}_n(x) = \frac{k}{n}\right\} = C_n^k F^k(x)(1 - F(x))^{n-k}, \quad k = \overline{0, n};$$

2) $\hat{F}_n(x)$ є незміщеною оцінкою значення функції розподілу $F(x)$:
$$\mathrm{M}\hat{F}_n(x) = F(x);$$

3) $\hat{F}_n(x)$ є консистентною оцінкою значення функції розподілу $F(x)$: для кожного $\varepsilon > 0$
$$\mathrm{P}(|\hat{F}_n(x) - F(x)| > \varepsilon) \to 0, \quad n \to \infty.$$

◀ Зафіксуємо довільне $x \in \mathbb{R}$. Розглянемо послідовність із n випробувань, в яких i-м успіхом називається подія $\{X_i < x\}$, $i = \overline{1,n}$. Оскільки $X_1, X_2, \ldots, X_n$ незалежні та однаково розподілені випадкові величини, то дана послідовність є схемою випробувань Бернуллі, причому ймовірність успіху не залежить від номера випробування і дорівнює $p = \mathrm{P}\{X_i < x\} = F(x)$. Випадкова величина $n\hat{F}_n(x) = \sum\limits_{i=1}^{n} \mathbb{I}_{\{X_i<x\}}$ — це кількість успіхів в n випробуваннях Бернуллі. Отже,

$$\mathrm{P}\left\{n\hat{F}_n(x) = k\right\} = C_n^k p^k (1-p)^{n-k} = \\ = C_n^k F^k(x)(1-F(x))^{n-k}, \quad k = \overline{0,n},$$

що і доводить 1).

Властивість 2) виводиться з лінійності математичного сподівання та з формули для математичного сподівання індикаторної величини:

$$\mathrm{M}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}(\mathbb{I}_{\{X_i<x\}}) = \frac{1}{n}\sum_{i=1}^{n} p = p = F(x).$$

Доведемо властивість 3). За означенням величина $\hat{F}_n(x)$ є відносною частотою успіхів в n випробуваннях за схемою Бернуллі. Тоді згідно з теоремою Бернуллі закону великих чисел $\hat{F}_n(x)$ збігається за ймовірністю при $n \to \infty$ до ймовірності успіху в одному випробуванні $p = F(x)$. Це означає, що $\hat{F}_n(x)$ є консистентною оцінкою для $F(x)$. ►

1.4 Вибіркові моменти

Нехай X — випадкова величина з функцією розподілу $F(x)$, $x \in \mathbb{R}$. Нагадаємо, що знаючи F, можна записати математичне сподівання функції $g(X)$ у вигляді

$$\mathrm{M}g(X) = \int\limits_{-\infty}^{\infty} g(x)F(dx),$$

де останній інтеграл є інтегралом Стільтьєса. У випадку, коли $g(X) = X^k$, або $g(X) = (X - \mathrm{M}X)^k$, $k \geq 0$, отримуємо відповідно початкові моменти m_k і центральні моменти m_k^0 порядку k випадкової величини X:

$$m_k = \mathrm{M}X^k = \int\limits_{-\infty}^{\infty} x^k F(dx),$$

$$m_k^0 = \mathrm{M}(X - \mathrm{M}X)^k = \int\limits_{-\infty}^{\infty} (x - \mathrm{M}X)^k F(dx).$$

Зокрема, при $g(X) = X$ і $g(X) = (X - \mathrm{M}X)^2$ отримуємо формули відповідно для математичного сподівання і дисперсії випадкової величини X.

В математичній статистиці всі ці числові характеристики, а також функцію розподілу випадкової величини X називають теоретичними.

Відзначимо, що між теоретичними центральними і початковими моментами існує простий зв'язок. Справді,

$$m_k^0 = \mathrm{M}(X - \mathrm{M}X)^k = \mathrm{M}\sum_{i=0}^{k} C_k^i X^i(-\mathrm{M}X)^{k-i} =$$

$$= \sum_{i=0}^{k} C_k^i \mathrm{M}X^i(-\mathrm{M}X)^{k-i} = \sum_{i=0}^{k}(-1)^{k-i} C_k^i m_i (m_1)^{k-i} =$$

$$= \sum_{i=2}^{k}(-1)^{k-i} C_k^i m_i (m_1)^{k-i} + (-1)^{k-1}(k-1)(m_1)^k. \quad (1.10)$$

Випишемо цей звязок між моментами для перших чотирьох значень k:

$$m_0^0 = 1, \quad m_1^0 = 0, \quad m_2^0 = m_2 - (m_1)^2,$$
$$m_3^0 = m_3 - 3m_2m_1 + 2(m_1)^2,$$
$$m_4^0 = m_4 - 4m_3m_1 + 6m_2(m_1)^2 - 3(m_1)^4.$$

Нехай $\mathbf{X} = (X_1, X_2 \ldots, X_n)$ — вибірка, $\hat{F}_n(x),\ x \in \mathbb{R}$, — емпірична функція розподілу, побудована за реалізацією $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ вибірки $\mathbf{X}$. Позначимо через $\widetilde{X}$ випадкову величину з функцією розподілу $\hat{F}_n$. За означенням початкові і центральні моменти k-го порядку випадкової величини $\widetilde{X}$ відповідно дорівнюють

$$\hat{m}_k(\mathbf{x}) = \mathrm{M}\widetilde{X}^k = \frac{1}{n}\sum_{i=1}^{r} z_{(i)}^k n_i = \frac{1}{n}\sum_{i=1}^{n} x_i^k,$$
$$\hat{m}_k^0(\mathbf{x}) = \mathrm{M}(\widetilde{X} - \mathrm{M}\widetilde{X})^k =$$
$$= \frac{1}{n}\sum_{i=1}^{r}(z_{(i)} - \mathrm{M}\widetilde{X})^k n_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k,$$

де $\bar{x} = \hat{m}_1(\mathbf{x}),\ z_{(i)},\ i = \overline{1,r}$ — можливі значення випадкової величини $\widetilde{X}$, а $\frac{n_i}{n}$ — ймовірності цих значень (див. (1.9)).

Означення. Випадкова величина

$$\hat{m}_k = \hat{m}_k(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

називається вибірковим початковим моментом порядку k. Зокрема, вибірковий початковий момент першого порядку $\bar{X} = \hat{m}_1(\mathbf{X})$ називається вибірковим середнім.

Означення. Випадкова величина

$$\hat{m}_k^0 = \hat{m}_k^0(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

називається вибірковим центральним моментом порядку k. Зокрема, вибірковий центральний момент другого порядку $\hat{\sigma^2} = \hat{m}_2^0(\mathbf{X})$ називається вибірковою дисперсією.

Зауваження. При кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ моменти $\hat{m}_k$, $\hat{m}_k^0$ є теоретичними моментами порядку k випадкової величини $\widetilde{X}$, і тому вони задовольняють всі властивості теоретичних моментів.

З урахуванням останнього зауваження, наприклад, можна стверджувати, що для вибіркових початкових і центральних моментів виконуються співвідношення (1.10). Зокрема, має місце тотожність

$$\hat{\sigma^2} = \hat{m}_2 - (\hat{m}_1)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

Також відзначимо, що центральні моменти є інваріантними відносно зсувів, тобто для довільної сталої c

$$\begin{aligned} \hat{m}_k^0 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k = \\ &= \frac{1}{n}\sum_{i=1}^{n}\left((X_i - c) - \frac{1}{n}\sum_{i=1}^{n}(X_i - c)\right)^k. \end{aligned} \tag{1.11}$$

Теорема (про властивості вибіркового середнього). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі скінченною дисперсією, яку позначимо через σ^2. Тоді оцінка $\bar{X}$ математичного сподівання $a = \mathrm{M}X_1$ є незміщеною, консистентною та ефективною в класі всіх лінійних оцінок, тобто оцінок вигляду

$$\widetilde{\theta} = \sum_{i=1}^{n} \alpha_i X_i,$$

де $\alpha_1, \alpha_2, \ldots, \alpha_n$ — довільні сталі такі, що $\sum\limits_{i=1}^{n} \alpha_i = 1$.

◀ Незміщеність і консистентність оцінки $\bar{X}$ випливає з рівності (1.6) і співвідношення (1.7) відповідно.

Доведемо, що вибіркове середнє $\bar{X}$ є ефективною оцінкою параметра a в класі всіх лінійних оцінок. Для цього достатньо показати, що

$$\mathrm{D}\widetilde{\theta} = \mathrm{D}\sum_{i=1}^{n} \alpha_i X_i = \sum_{i=1}^{n} \alpha_i^2 \mathrm{D}X_i = \sigma^2 \sum_{i=1}^{n} \alpha_i^2,$$

де $\sigma^2 \equiv \mathrm{D}X_1$, набуває свого мінімального значення при $\alpha_i = \frac{1}{n}$, тобто коли $\widetilde{\theta} = \bar{X}$.

Для знаходження умовного мінімуму функції

$$g(\alpha_1, \alpha_2, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i^2$$

при умові

$$\sum_{i=1}^{n} \alpha_i = 1$$

складемо функцію Лагранжа

$$L(\alpha_1, \alpha_2, \ldots, \alpha_n; \lambda) = \sum_{i=1}^{n} \alpha_i^2 + \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right),$$

де λ — множник Лагранжа. Запишемо необхідні умови існування умовного екстремуму:

$$\begin{cases} \frac{\partial L}{\partial \alpha_i} = 2\alpha_i + \lambda = 0, \quad i = \overline{1, n}, \\ \frac{\partial L}{\partial \lambda} = \sum\limits_{i=1}^{n} \alpha_i - 1 = 0. \end{cases}$$

Розв'язавши цю систему, знаходимо $\lambda = -\frac{2}{n}$ і $\alpha_i = \frac{1}{n}$, $i = \overline{1,n}$, і переконуємося в тому, що при цих значеннях аргументів функція $g(\alpha_1, \alpha_2, \ldots, \alpha_n)$ має умовний мінімум. ▶

Теорема (про властивості вибіркової дисперсії). Якщо $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу зі скінченною дисперсією σ^2, то вибіркова дисперсія $\hat{\sigma^2}$ — зміщена консистентна оцінка σ^2.

◀ Враховуючи те, що оцінка $\hat{\sigma^2}$ інваріантна відносно зсувів (див. формулу (1.11)), зокрема, не змінюється при відніманні від X_i їх спільного середнього $a = \mathrm{M}X_i$, можна вважати, що $a = \mathrm{M}X_i = 0$. Тоді

$$\mathrm{M}\hat{\sigma^2} = \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}X_i^2 - \mathrm{M}\bar{X}^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathrm{D}X_i - \mathrm{D}\bar{X} = \frac{1}{n}n\sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2,$$

тобто $\hat{\sigma^2}$ — зміщена оцінка для дисперсії.

Покажемо, що $\hat{\sigma^2}$ є консистентною оцінкою. Доведення проведемо для випадку, коли існують моменти даного розподілу до четвертого порядку включно. Згідно з другою нерівністю Чебишева

$$\mathrm{P}\left\{\left|\hat{\sigma^2} - \frac{n-1}{n}\sigma^2\right| > \varepsilon\right\} \leq \frac{\mathrm{D}\hat{\sigma^2}}{\varepsilon^2}. \qquad (1.12)$$

Для знаходження дисперсії оцінки $\hat{\sigma^2}$ скористаємося формулою

$$\mathrm{D}\hat{\sigma^2} = \mathrm{M}(\hat{\sigma^2})^2 - (\mathrm{M}\hat{\sigma^2})^2.$$

Оскільки

$$(\hat{\sigma^2})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right)^2 =$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2\right)^2 - 2\bar{X}^2\frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}^4,$$

то виконується рівність

$$\mathrm{M}(\hat{\sigma^2})^2 = \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^2\right)^2 -$$

$$- \frac{2}{n}\mathrm{M}\left(\bar{X}^2\sum_{i=1}^{n} X_i^2\right) + \mathrm{M}\bar{X}^4. \tag{1.13}$$

Для першого доданка у формулі (1.13) маємо

$$\mathrm{M}\left(\sum_{i=1}^{n} X_i^2\right)^2 = \mathrm{M}\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\right) =$$

$$= \sum_{i=1}^{n} \mathrm{M}X_i^4 + \sum_{i\neq j} \mathrm{M}X_i^2 \mathrm{M}X_j^2 =$$

$$= \sum_{i=1}^{n} m_4^0 + \sum_{i\neq j} \sigma^2\sigma^2 = nm_4^0 + n(n-1)\sigma^4. \tag{1.14}$$

Спростимо другий доданок у формулі (1.13):

$$\mathrm{M}\left(\bar{X}^2\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n} X_j\right)^2\sum_{i=1}^{n} X_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n} X_j^2 + \sum_{j\neq k} X_j X_k\right)\sum_{i=1}^{n} X_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\sum_{j=1}^{n} X_j^2 \sum_{i=1}^{n} X_i^2\right) + \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^2 \sum_{j\neq k} X_j X_k\right).$$

Враховуючи те, що

$$\mathrm{M}\sum_{i:\ i\neq k} X_i X_k = \sum_{i:\ i\neq k} \mathrm{M}X_i \mathrm{M}X_k = 0,$$

отримуємо

$$\sum_{i=1}^{n} X_i^2 \sum_{j\neq k} X_j X_k =$$
$$= \sum_{i\neq j, i\neq k, j\neq k} \mathrm{M}(X_i^2 X_j X_k) + 2\sum_{i\neq j} X_i^3 X_j = 0.$$

Звідси

$$\mathrm{M}\left(\bar{X}^2 \sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\sum_{j=1}^{n} X_j^2 \sum_{i=1}^{n} X_i^2\right) =$$
$$= \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\right) =$$
$$= \frac{1}{n^2}\left(nm_4^0 + n(n-1)\sigma^4\right). \tag{1.15}$$

Аналогічно можна показати, що

$$\mathrm{M}\bar{X}^4 = \frac{1}{n^4}\left(\sum_{i=1}^{n} X_i\right)^4 = \frac{m_4^0 + 3(n-1)\sigma^4}{n^3}. \tag{1.16}$$

Підставивши (1.14), (1.15), (1.16) у (1.13), отримуємо

$$\mathrm{M}(\hat{\sigma^2})^2 = \frac{nm_4^0 + n(n-1)\sigma^4}{n^2} -$$
$$- \frac{2(nm_4^0 + n(n-1)\sigma^4)}{n^3} + \frac{m_4^0 - 3\sigma^4}{n^3} =$$

$$= \sigma^4 + \frac{m_4^0 - 3\sigma^4}{n} - \frac{2m_4^0 - 5\sigma^4}{n^2} + \frac{m_4 + 3(n-1)\sigma^4}{n^3}.$$

Оскільки $\mathrm{M}\hat{\sigma^2} = \sigma^2 - \frac{\sigma^2}{n}$, остаточно виводимо

$$\mathrm{D}\hat{\sigma^2} = \frac{m_4^0 - \sigma^4}{n} - \frac{2(m_4^0 - 2\sigma^4)}{n^2} + \frac{m_4^0 - 3\sigma^4}{n^3}.$$

Підставивши вираз для $\mathrm{D}\hat{\sigma^2}$ у другу нерівність Чебишева (1.12) і перейшовши до границі при $n \to \infty$, встановлюємо, що вибіркова дисперсія $\hat{\sigma^2}$ є консистентною оцінкою для σ^2. ▶

Зауваження. З теореми випливає, що оцінка

$$S^2 = \frac{n}{n-1}\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

є незміщеною і консистентною оцінкою дисперсії σ^2. Її називають виправленою вибірковою дисперсією.

Справді,

$$\mathrm{M}S^2 = \frac{n}{n-1}\mathrm{M}\hat{\sigma^2} = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2,$$
$$\mathrm{D}S^2 = \frac{n^2}{(n-1)^2}\mathrm{D}\hat{\sigma^2} \to 0, \quad n \to \infty,$$

а це й означає незміщеність і консистентність S^2.

Зауваження. Можна довести, що вибіркові початкові і центральні моменти є консистентними оцінками відповідних теоретичних моментів, якщо тільки вони існують. Однак ці оцінки, крім $\bar{X}$, є зміщеними.

1.5 Метод моментів

Метод моментів був запропонований англійським статистиком К. Пірсоном і є одним із перших загальних методів оцінювання невідомих параметрів розподілів. Цей метод полягає у наступному.

Нехай $\mathbf{X} = (X_1, X_2, \dots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$. Розподіл $F(\cdot, \theta)$ залежить від невідомого векторного параметра $\theta = (\theta_1, \theta_2, \dots, \theta_s)$, який набуває значень із множини $\Theta \subset \mathbb{R}^s$. Необхідно знайти оцінку параметра θ за реалізацією $\mathbf{x} = (x_1, x_2, \dots, x_n)$ вибірки $\mathbf{X}$.

Будемо вважати, що для даного розподілу існують перші s моментів, які, очевидно, є функціями векторного аргументу θ:

$$m_k = m_k(\theta) = \int\limits_{-\infty}^{\infty} x^k F(dx; \theta), \quad k = \overline{1, s}.$$

Розглянемо вибіркові моменти

$$\hat{m}_k = \hat{m}_k(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

Нагадаємо, що вибіркові моменти $\hat{m}_k$ є консистентними оцінками відповідних теоретичних моментів m_k. Це означає, що при великому обсязі вибірки n значення оцінок $\hat{m}_k = \hat{m}_k(\mathbf{x})$, де $\mathbf{x}$ — реалізація вибірки $\mathbf{X}$, є близькими до відповідних моментів m_k.

Згідно з методом моментів в якості точкової оцінки $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n)$ векторного параметра θ беруть оцінку, значення якої для довільної реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ отримують як розв'язок системи рівнянь

$$\hat{m}_k = m_k(\theta), \quad k = \overline{1, s}.$$

Можна показати, що за умови неперервної залежності розв'язку цієї системи від $\hat{m}_k,\ k = \overline{1, s}$, оцінка, отримана методом моментів, є конзистентною і має асимптотично нормальний розподіл, тобто її розподіл при $n \to \infty$ прямує до нормального.

Приклад. Знайти оцінки параметрів a та σ^2 нормального розподілу

$$f(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}.$$

Для нормального розподілу з параметрами a та σ^2 теоретичні моменти $m_1(a, \sigma^2)$ та $m_2(a, \sigma^2)$ відповідно дорівнюють

$$m_1(a, \sigma^2) = a, \quad m_2(a, \sigma^2) = m_2^0(a, \sigma^2) + (m_1(a, \sigma^2))^2 = \sigma^2 + a^2.$$

Нехай $\mathbf{x} = (x_1, x_2, \ldots x_n)$ — реалізація вибірки $\mathbf{X}$ із розподілу $N(a, \sigma^2)$. Тоді

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2.$$

Прирівнюючи значення вибіркових моментів $\hat{m}_1$ та $\hat{m}_2$ до відповідних теоретичних моментів $m_1(a, \sigma^2)$ та $m_2(a, \sigma^2)$, отримуємо систему рівнянь

$$\begin{cases} \frac{1}{n} \sum\limits_{i=1}^{n} x_i = a, \\ \frac{1}{n} \sum\limits_{i=1}^{n} x_i^2 = \sigma^2 + a^2, \end{cases}$$

звідки знаходимо значення оцінок

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2.$$

Отже, оцінками методу моментів для математичного a і дисперсії σ^2 нормального розподілу є відповідно вибіркове середнє

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

і вибіркова дисперсія

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

1.6 Метод максимальної правдоподібності

Одним із найбільш універсальних методів оцінювання параметрів розподілів є метод максимальної правдоподібності, запропонований англійським статистиком Р. Фішером.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$, де $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$. Параметр θ — невідомий, і його необхідно оцінити за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$. Як і раніше, через $f_{\mathbf{X}}(\mathbf{x}; \theta) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n; \theta)$ будемо позначати щільність розподілу випадкового вектора $\mathbf{X}$ у випадку, коли він абсолютно неперервний, і ймовірність події $\{\mathbf{X} = \mathbf{x}\}$, якщо вектор $\mathbf{X}$ — дискретний.

Означення. Функцією правдоподібності вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ будемо називати випадкову функцію $L(\theta)$ аргументу $\theta \in \Theta \subset \mathbb{R}^s$, що визначається рівністю

$$L(\theta) = f_{\mathbf{X}}(\mathbf{X}; \theta).$$

Метод максимальної правдоподібносі полягає в тому, що за оцінку параметра θ вибирається оцінка $\hat{\theta} = (\hat{\theta}_1,$

$\hat{\theta}_2, \ldots, \hat{\theta}_n)$, значення якої при кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ максимізує функцію правдоподібності $L(\theta)$:

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{x}, \theta).$$

Точкова оцінка, знайдена у такий спосіб, називається оцінкою максимальної правдоподібності.

Відзначимо, що функції $L(\theta)$ і $\ln L(\theta)$ досягають найбільшого значення в одній і тій самій точці. А знаходити точку, в якій функція $\ln L(\theta)$ досягає найбільшого значення, часто зручніше.

Логарифм від функції правдоподібності називають логарифмічною функцією правдоподібності.

Якщо функція $L(\theta)$ диференційовна при будь-якій реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ і максимум $L(\theta)$ досягається у внутрішній точці множини Θ, то значення оцінки максимальної правдоподібності задовольняє рівняння (необхідні умови екстремуму)

$$\frac{\partial \ln L(\theta)}{\partial \theta_k} = 0, \quad k = \overline{1, s}. \tag{1.17}$$

Рівняння (1.17) називають рівняннями правдоподібності.

Можна показати, що за умов регулярності функції $f_{\mathbf{X}}(\mathbf{x}; \theta)$ оцінка максимальної правдоподібності $\hat{\theta}_n$ є консистентною, асимптотично ефективною і має асимптотично нормальний розподіл.

Приклад. Знайти оцінку максимальної правдоподібності ймовірності p успіху у схемі n випробувань Бернуллі.

Розглянемо вибірку $\mathbf{X} = (X_1, X_2, \ldots X_n)$ із дискретного розподілу $(0; 1-p)$, $(1; p)$ (X_i — індикатор успіху в i-му випробуванні Бернуллі). Для будь-якої реалізації $\mathbf{x}$ вибірки

X функція правдоподібності $L(p)$ має вигляд:

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

Знайшовши

$$\ln L(p) = \sum_{i=1}^{n} \left(x_i \ln p + (1-x_i)\ln(1-p)\right),$$

отримуємо рівняння правдоподібності

$$\frac{\partial \ln L(p)}{\partial p} = \sum_{i=1}^{n} \left(\frac{x_i}{p} - \frac{1-x_i}{1-p}\right) = 0.$$

Розв'язком цього рівняння є

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{k}{n},$$

де k — кількість успіхів в схемі Бернуллі. Нескладно переконатися в тому, що $\hat{p}$ є точкою максимуму функції $L(p)$. Таким чином, оцінкою максимальної правдоподібності ймовірності p є відносна частота успіхів в n випробуваннях.

Розділ 2

Інтервальні оцінки

2.1 Поняття про інтервальне оцінювання

Для оцінювання невідомих параметрів розподілів поряд із розглянутими вище точковими оцінками використовуються також інтервальні оцінки. На відміну від точкової оцінки, інтервальна оцінка дозволяє отримати ймовірнісну характеристику точності оцінювання невідомого параметра.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot, \theta)$. Значення параметра $\theta \in \mathbb{R}$ в розподілі $F(\cdot, \theta)$ невідоме і його необхідно оцінити.

Означення. Інтервальною оцінкою (надійним інтервалом) параметра θ з рівнем надійності $1-\alpha$ називається випадковий інтервал $(\hat{\theta}_1, \hat{\theta}_2)$, де $\hat{\theta}_1$, $\hat{\theta}_2$ є функціями вибірки $\mathbf{X}$, який з ймовірністю $1-\alpha$ містить значення оцінюваного параметра θ, тобто

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha. \qquad (2.1)$$

Ймовірнісною характеристикою точності інтервального оцінювання параметра θ є випадкова величина

$$l = \hat{\theta}_2 - \hat{\theta}_1,$$

яка для будь-якої реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ є довжиною інтервалу $(\hat{\theta}_1, \hat{\theta}_2)$.

В деяких ситуаціях (наприклад, при розгляді дискретних випадкових величин) замість рівності (2.1) вдається лише отримати нерівність

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} \geq 1 - \alpha,$$

тобто побудувати інтервальну оцінку параметра θ з рівнем надійності не меншим, ніж $1 - \alpha$. Інколи виникає потреба оцінити параметр θ тільки зліва або тільки справа. При цьому, якщо

$$\mathrm{P}\{\hat{\theta}_1 < \theta\} = 1 - \alpha,$$

то інтервал $(\hat{\theta}_1, \infty)$ називають правостороннім надійним інтервалом, а у випадку, коли

$$P\{\theta < \hat{\theta}_2\} = 1 - \alpha,$$

інтервал $(-\infty, \theta_2)$ називають лівостороннім надійним інтервалом з рівнем надійності $1 - \alpha$.

Зауваження. Поняття інтервальної оцінки тривіально поширюються на випадок векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

2.2 Загальний метод опорних величин

Загальним методом побудови інтервальних оцінок для θ є метод опорних величин. Він полягає у знаходженні опорної випадкової величини вигляду $h(\mathbf{X};\theta) = h(X_1, X_2, \ldots, X_n; \theta)$, що має такі властивості:

1) вона є функцією вибірки $\mathbf{X}$ та невідомого параметра θ,

2) має повністю відомий розподіл,

3) монотонно залежить від θ.

Виходячи з відомого розподілу цієї випадкової величини, знаходять такі значення h_1, h_2, що

$$\mathrm{P}\{h_1 < h(\mathbf{X};\theta) < h_2\} = 1 - \alpha, \tag{2.2}$$

при всіх можливих значеннях θ. Далі, з урахуванням властивості 3), розв'язують відносно θ нерівність у фігурних дужках виразу (2.2) і, таким чином, знаходять випадкові величини $\hat{\theta}_1, \hat{\theta}_2$ такі, що

$$\mathrm{P}\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \mathrm{P}\{h_1 < h(\mathbf{X};\theta) < h_2\} = 1 - \alpha.$$

Отже, $(\hat{\theta}_1, \hat{\theta}_2)$ — шуканий надійний інтервал для θ з рівнем надійності $1 - \alpha$.

Якщо розподіл опорної величини лише асимптотично наближається до відомого, то, розглядаючи вибірку великого обсягу n, відповідні асимптотичні надійні інтервали будують, виходячи з наближених рівностей

$$\mathrm{P}\{h_1 < h_n(\mathbf{X};\theta) < h_2\} \approx 1 - \alpha.$$

2.3 Деякі спеціальні розподіли

Нижче будуть визначені деякі спеціальні розподіли, які мають важливе значення в математичній статистиці і використовуватимуться нами у параграфі 2.4 цього розділу для побудови інтервальних оцінок параметрів нормального розподілу, а також у наступних розділах даної праці.

Означення. Випадкова величина χ_n^2 має χ^2-розподіл або розподіл Пірсона з n ступенями свободи, якщо її можна подати у вигляді

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2,$$

де $X_1, X_2, \ldots, X_n$ — незалежні в сукупності стандартні нормальні величини.

Щільність розподілу випадкової величини χ_n^2 задається формулою:

$$f_{\chi_n^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

де $\Gamma(p) = \int\limits_0^\infty t^{p-1}e^{-t}dt,\ p > 0,$ — гамма-функція Ейлера.

Знайдемо математичне сподівання та дисперсію:

$$\mathrm{M}\chi_n^2 = \mathrm{M}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathrm{M}X_i^2 = n,$$

$$\mathrm{D}\chi_n^2 = \mathrm{D}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathrm{D}X_i^2 = \sum_{i=1}^{n} \left(\mathrm{M}X_i^4 - (\mathrm{M}X_i^2)^2\right) = 2n.$$

Означення. Випадкова величина τ_n має t-розподіл або розподіл Стьюдента з n ступенями свободи, якщо її можна подати у вигляді

$$\tau_n = \frac{X}{\sqrt{\chi_n^2/n}},$$

де X — стандартна нормальна величина, а χ^2_n — незалежна від неї величина з χ^2-розподілом та n ступенями свободи.

Щільність розподілу випадкової величини τ_n має вигляд:

$$f_{\tau_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

Можна показати, що

$$\mathrm{M}\tau_n = 0, \quad n \geq 2,$$

$$\mathrm{D}\tau_n = \frac{n}{n-2}, \quad n > 2.$$

Означення. Випадкова величина $\phi_{n,m}$ має F-розподіл або розподіл Фішера з n, m ступенями свободи, якщо її можна подати у вигляді

$$\phi_{n,m} = \frac{\chi^2_n / n}{\chi^2_m / m},$$

де χ^2_n, χ^2_m — незалежні величини з χ^2-розподілом та n, m ступенями свободи відповідно.

Для розподілу Фішера

$$f_{\phi_{n,m}}(x) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

$$\mathrm{M}\phi_{n,m} = \frac{m}{m-2}, \quad m > 2,$$

$$\mathrm{D}\phi_{n,m} = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, \quad m > 4.$$

Теорема (про вибіркові моменти нормальної вибірки). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу $N(a, \sigma^2)$. Тоді вибіркове середнє $\bar{X}$ та виправлена вибіркова дисперсія S^2 — незалежні. До того ж, величина $\bar{X}$ має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$, а величина $(n-1)\frac{S^2}{\sigma^2}$ — χ^2-розподіл з $n-1$ ступенями свободи.

Перш ніж доводити цю теорему, нагадаємо деякі поняття з лінійної алгебри.

Матриця A розміру $n \times n$ називається ортогональною, якщо вона невироджена та $A^T = A^{-1}$, де A^T — транспонована матриця. Це рівносильно тому, що $A^T A = AA^T = E$, а $|\det A| = 1$ з огляду на рівність

$$(\det A)^2 = \det A \det A = \det A^T \det A = \det(A^T A) = 1.$$

Якщо $\|\mathbf{x}\| = \left(\sum\limits_{i=1}^n x_i^2\right)^{\frac{1}{2}}$ — евклідова норма в $\mathbb{R}^n$, то $\|A\mathbf{x}\| = \|\mathbf{x}\|$. Справді,

$$\|A\mathbf{x}\| = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|.$$

Нехай тепер $\mathbf{X} = (X_1, X_2, \cdot, X_n)$ — вибірка з розподілу $N(0, \sigma^2)$. Тоді

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right\},$$

а для випадкового вектора $Y = g(\mathbf{X}) = A\mathbf{X} + \mathbf{c}$, використавши відому формулу

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \cdot J_{g^{-1}}(y)$$

(J_h — якобіан відображення h, тобто визначник матриці $\left(\frac{\partial h_i}{\partial y_j}\right)_{i,j=1}^{n}$), отримуємо

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(A^{-1}(\mathbf{y}-\mathbf{c})) = f_{\mathbf{X}}(A^{T}(\mathbf{y}-\mathbf{c})) = \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|A^T(\mathbf{y}-\mathbf{c})\|^2\right\} = \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{c}\|^2\right\} = \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i-c_i)^2}{2\sigma^2}\right\}. \qquad (2.3)
\end{aligned}
$$

Таким чином, компоненти $Y_1, Y_2, \ldots, Y_n$ випадкового вектора $\mathbf{Y}$ незалежні і мають розподіл $N(c_i, \sigma^2)$. Більше того, якщо $\mathbf{Z} = \mathbf{X} + \mathbf{d}$, а $\mathbf{Y} = g(\mathbf{Z}) = A(\mathbf{X}+\mathbf{d}) + \mathbf{c} = A\mathbf{X} + (A\mathbf{d} + \mathbf{c})$, то $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{y} - (A\mathbf{d}+\mathbf{c}))$. Отже, ми довели наступне твердження.

Теорема (про ортогональне перетворення нормального вектора) Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — випадковий вектор, компоненти якого незалежні та нормально розподілені з однаковою дисперсією σ^2, а $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = A\mathbf{X} + \mathbf{c}$ з ортогональною матрицею $A = (a_{ij})$. Тоді $Y_1, Y_2, \ldots, Y_n$ незалежні та мають нормальний розподіл з математичним сподіванням $\mathrm{M}Y_i = \sum\limits_{j=1}^{n} a_{ij}\mathrm{M}X_j + c_i$ та дисперсією σ^2.

Доведемо тепер теорему про вибіркові моменти нормальної вибірки.

◀ Побудуємо яку-небудь ортогональну матрицю A з першим рядком

$$
\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)
$$

і покладемо $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = A\mathbf{X}$. Тоді за теоремою про ортогональне перетворення нормального вектора випадкові величини $Y_1, Y_2, \ldots, Y_n$ незалежні і нормально розподілені з математичним сподіванням $\mathrm{M}Y_i = a\sum\limits_{j=1}^{n} a_{ij}$ та дисперсією σ^2. Оскільки матриця A ортогональна, при $i > 1$ маємо

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{1j}a_{ij} = 0,$$

а тому

$$\mathrm{M}Y_1 = a\sqrt{n}, \quad \mathrm{M}Y_2 = \mathrm{M}Y_3 = \ldots = \mathrm{M}Y_n = 0.$$

Розглянемо випадкову величину $\frac{1}{\sigma^2}\sum\limits_{i=2}^{n} Y_i^2 = \sum\limits_{i=2}^{n}\left(\frac{Y_i}{\sigma}\right)^2$. Ця величина має χ^2-розподіл з $n-1$ ступенями свободи, як сума квадратів $n-1$ стандартних нормальних величин. З рівності

$$\bar{X} = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)(X_1, X_2, \ldots X_n)^T = \frac{Y_1}{\sqrt{n}}.$$

випливає, що випадкові величини $\bar{X}$ та $\frac{1}{\sigma^2}\sum\limits_{i=2}^{n} Y_i^2$ незалежні, причому вибіркове середнє $\bar{X}$ має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$. Для того, щоб завершити доведення теореми, достатньо показати, що $(n-1)S^2 = \sum\limits_{i=2}^{n} Y_i^2$, а це неважко перевірити з огляду на ортогональність матриці A:

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \|X\|^2 - Y_1^2 =$$

$$= \|AX\|^2 - Y_1^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2. \quad \blacktriangleright$$

Теорема (про статистику Стьюдента від нормальної вибірки) Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу $N(a, \sigma^2)$, а S^2 — виправлена вибіркова дисперсія. Тоді випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S}$$

має розподіл Стьюдента з $n-1$ ступенями свободи.

◀ Випадкову величину τ_{n-1} запишемо у вигляді

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S} = \frac{\frac{\bar{X}-a}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\frac{\bar{X}-a}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}.$$

З теореми про вибіркові моменти нормальної вибірки випливає, що випадкові величини $\frac{\bar{X}-a}{\sigma/\sqrt{n}}$ та $\frac{(n-1)S^2}{\sigma^2}$ — незалежні, до того ж, вони мають стандартний нормальний розподіл та χ^2-розподіл з $n-1$ ступенями свободи відповідно. Це й означає, що випадкова величина τ_{n-1} має розподіл Стьюдента з $n-1$ ступенями свободи. ▶

2.4 Інтервальні оцінки параметрів нормального розподілу

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з нормального розподілу $N(a, \sigma^2)$. Розглянемо різні варіанти побудови інтервальних оцінок для математичного сподівання a і дисперсії σ^2.

Надійний інтервал для математичного сподівання при відомій дисперсії. Точковою оцінкою математичного сподівання є вибіркове середнє $\bar{X}$. За теоремою про вибіркові моменти нормальної вибірки ця випадкова величина має розподіл $N\left(a, \frac{\sigma^2}{n}\right)$. Тому випадкова величина

$$Z = h(\mathbf{X}; a) = \frac{\bar{X} - a}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\bar{X} - a}{\sigma}$$

має стандартний нормальний розподіл $N(0, 1)$. Це означає, що

$$\mathrm{P}\left\{-t < \sqrt{n}\frac{\bar{X} - a}{\sigma} < t\right\} = F_0(t) - F_0(-t) = 2F_0(t) - 1,$$

де $F_0(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} \exp\{-\frac{z^2}{2}\}dz$ — функція розподілу $N(0, 1)$. Інакше кажучи,

$$\mathrm{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}t < a < \bar{X} + \frac{\sigma}{\sqrt{n}}t\right\} = 2F_0(t) - 1.$$

Зафіксуємо рівень надійності $1 - \alpha$ і знайдемо таке t, щоб виконувалася рівність $2F_0(t) - 1 = 1 - \alpha$. Отримуємо $t = u_{1-\frac{\alpha}{2}}$ — квантиль рівня $1 - \frac{\alpha}{2}$ стандартного нормального розподілу. Тоді

$$\mathrm{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} < a < \bar{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right\} = 1 - \alpha.$$

Отже, надійний інтервал для математичного сподівання нормального розподілу при відомій дисперсії σ^2 має вигляд:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}},\ \bar{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right).$$

Надійний інтервал для математичного сподівання при невідомій дисперсії. Поряд з точковою оцінкою $\bar{X}$ математичного сподівання a розглянемо оцінку S^2 для дисперсії σ^2. За теоремою про статистику Стьюдента від нормальної вибірки випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a}{S}$$

має розподіл Стьюдента з $n-1$ ступенями свободи. Позначимо через $F_{\tau_{n-1}}(t)$ функцію розподілу величини τ_{n-1}. Тоді

$$\begin{aligned}
&\mathrm{P}\left\{-t < \sqrt{n}\frac{\bar{X} - a}{S} < t\right\} = F_{\tau_{n-1}}(t) - F_{\tau_{n-1}}(-t) = \\
&= 2F_{\tau_{n-1}}(t) - 1; \\
&\mathrm{P}\left\{\bar{X} - \frac{S}{\sqrt{n}}t < a < \bar{X} + \frac{S}{\sqrt{n}}t\right\} = 2F_{\tau_{n-1}}(t) - 1.
\end{aligned}$$

Розв'язавши рівняння $2F_{\tau_{n-1}}(t) - 1 = 1 - \alpha$, де $1-\alpha$ — рівень надійності, отримуємо $t = t_{1-\frac{\alpha}{2}}(n-1)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n-1$ ступенями свободи. Тоді

$$\mathrm{P}\left\{\bar{X} - \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1) < a < \bar{X} + \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\right\} = 1-\alpha.$$

Таким чином, надійний інтервал для математичного сподівання нормального розподілу при невідомій дисперсії має вигляд:

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

Надійний інтервал для дисперсії при відомому математичному сподіванні. Найкращою точковою оцінкою дисперсії σ^2 при відомому математичному сподіванні a є $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$ (див. приклади параграфу 1.2). Розглянемо випадкову величину

$$n\frac{\hat{\sigma^2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - a}{\sigma}\right)^2.$$

Ця випадкова величина є сумою квадратів n стандартних нормальних величин і тому має χ^2-розподіл з n ступенями свободи. Тоді при заданому рівні надійності $1 - \alpha$ маємо

$$1 - \alpha = \mathrm{P}\left\{\sigma_1^2 < \sigma^2 < \sigma_2^2\right\} = \mathrm{P}\left\{n\frac{\hat{\sigma^2}}{\sigma_2^2} < n\frac{\hat{\sigma^2}}{\sigma^2} < n\frac{\hat{\sigma^2}}{\sigma_1^2}\right\} =$$
$$= F_{\chi_n^2}\left(n\frac{\hat{\sigma^2}}{\sigma_1^2}\right) - F_{\chi_n^2}\left(n\frac{\hat{\sigma^2}}{\sigma_2^2}\right), \tag{2.4}$$

де $F_{\chi_n^2}(t)$ — функція χ^2-розподілу з n ступенями свободи. Виберемо σ_1^2 і σ_2^2 так, щоб

$$F_{\chi_n^2}\left(n\frac{\hat{\sigma^2}}{\sigma_1^2}\right) = 1 - \frac{\alpha}{2}, \quad F_{\chi_n^2}\left(n\frac{\hat{\sigma^2}}{\sigma_2^2}\right) = \frac{\alpha}{2}.$$

Отримаємо

$$\sigma_1^2 = \frac{n\hat{\sigma^2}}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \quad \sigma_2^2 = \frac{n\hat{\sigma^2}}{\chi^2_{\frac{\alpha}{2}}(n)},$$

де $\chi^2_{1-\frac{\alpha}{2}}(n)$ та $\chi^2_{\frac{\alpha}{2}}(n)$ — квантилі χ^2 розподілу з n ступенями свободи рівня $1 - \frac{\alpha}{2}$ та $\frac{\alpha}{2}$ відповідно.

Отже, надійний інтервал для дисперсії нормального розподілу при відомому математичному сподіванні a має вигляд:

$$\left(\frac{n\hat{\sigma^2}}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \ \frac{n\hat{\sigma^2}}{\chi^2_{\frac{\alpha}{2}}(n)} \right).$$

Надійний інтервал для дисперсії при невідомому математичному сподіванні. Точковою оцінкою дисперсії σ^2 при невідомому математичному сподіванні є виправлена вибіркова дисперсія S^2. За теоремою про вибіркові моменти нормальної вибірки випадкова величина

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

має χ^2-розподіл з $n-1$ ступенями свободи. Тоді

$$\begin{aligned} 1-\alpha &= \mathrm{P}\left\{\sigma_1^2 < \sigma^2 < \sigma_2^2\right\} = \\ &= \mathrm{P}\left\{(n-1)\frac{S^2}{\sigma_2^2} < (n-1)\frac{S^2}{\sigma^2} < (n-1)\frac{S^2}{\sigma_1^2}\right\} = \\ &= F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_1^2}\right) - F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_2^2}\right). \end{aligned}$$

Межі надійного інтервалу σ_1^2 і σ_2^2 виберемо так, щоб

$$F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_1^2}\right) = 1 - \frac{\alpha}{2}, \quad F_{\chi^2_{n-1}}\left((n-1)\frac{S^2}{\sigma_2^2}\right) = \frac{\alpha}{2}.$$

Отримаємо

$$\sigma_1^2 = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \quad \sigma_2^2 = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}.$$

Таким чином, надійний інтервал для дисперсії нормального розподілу при невідомому математичному сподіванні має вигляд:

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \; \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right).$$

Вправа. Побудувати лівосторонні та правосторонні надійні інтервали для параметрів нормального розподілу.

Розділ 3

Перевірка статистичних гіпотез

3.1 Задачі перевірки статистичних гіпотез

В цьому розділі розглядається ще один клас задач математичної статистики, пов'язаних з перевіркою статистичних гіпотез.

У теорії перевірки статистичних гіпотез прийнято наступні означення та домовленості. Статистичною гіпотезою називається довільне припущення про розподіл випадкової величини, яке перевіряється за результатами спостережень цієї величини. Гіпотези щодо розподілів, які однозначно їх визначають, називаються простими, у протилежному разі — складними. Гіпотеза, яка підлягає перевірці, називається основною чи нульовою і позначається H_0. Гіпотеза, що суперечить H_0, називається альтернативною чи конкуруючою гіпотезою. Вона позначається H_1. Правило, за яким приймається рішення на користь однієї із вказаних гіпотез, називається статистичним критерієм.

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із розподілу G, стосовно якого висувається гіпотеза H_0 при альтернативній гіпотезі H_1. Перевірка гіпотези H_0 зводиться до наступного. За реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ підраховується значення деякої борелєвої функції $h(\mathbf{x})$, заданої на вибірковому просторі $\mathbb{R}^n$. Ця функція називається статистикою критерію. Вона вибирається так, щоб за припущення H_0 розподіл величини $h(\mathbf{X})$ був відомим. Нехай D — множина значень статистики h, а $D_k \subset D$ — така множина, що, у випадку істинності H_0, ймовірність потрапляння значення випадкової величини $h(\mathbf{X})$ в D_k дорівнює наперед заданому числу α, яке називається рівнем значущості критерію, тобто

$$\mathrm{P}_{H_0}\{h(\mathbf{X}) \in D_k\} = \alpha. \tag{3.1}$$

При цьому рівень значущості α вибирається достатньо малим, щоб за припущення H_0 подію, яка полягає в тому, що $h(\mathbf{X}) \in D_k$ можна було б вважати такою, що в одноразовому експерименті практично не відбувається. Якщо підраховане за реалізацією вибірки значення $h(\mathbf{x})$ все ж потрапило в область D_k, то це можна пояснити тим, що нульова гіпотеза хибна і тому її треба відхилити. У цьому випадку приймається альтернативна гіпотеза H_1. У випадку, коли значення $h(\mathbf{x})$ не потрапило в D_k, гіпотеза H_0 не відхиляється.

Множина D_k, на якій відхиляється нульова гіпотеза, називається критичною областю статистики критерію. У загальному випадку з критичною областю D_k статистики критерію можна пов'язати критичну область вибірки

$$W_k = \{\mathbf{x} \in \mathbb{R}^n : \ h(\mathbf{x}) \in D_k\}.$$

При потраплянні реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ у критичну область W_k нульова гіпотеза відхиляється, у протилежному випадку — не відхиляється.

Зауваження. У зв'язку зі статистичним характером перевірки гіпотез зауважимо таке. Перевірка гіпотези ні в якому разі не є доведенням того, що вона істинна чи хибна. Невдача у відхиленні H_0 означає лише те, що немає достатньо вагомих підстав для відхилення H_0.

При використанні будь-якого статистичного критерію можливі помилки. Розрізняють помилки двох видів.

Означення. Помилка, яка полягає в тому, що гіпотеза H_0 відхиляєтья, коли вона справджується, називається помилкою першого роду.

Означення. Помилка, яка полягає в тому, що гіпотеза H_0 не відхиляєтья, коли правильною є альтернативна гіпотеза H_1, називається помилкою другого роду.

Ймовірність помилки першого роду співпадає з рівнем значущості критерію α і тому може бути зроблена як завгодно малою. Ймовірність помилки другого роду дорівнює

$$\beta = \mathrm{P}_{H_1}\{h(\mathbf{X}) \in D \setminus D_k)\} = \mathrm{P}_{H_1}\{\mathbf{X} \in \mathbb{R}^n \setminus W_k\}.$$

Означення. Потужністю критерію називається ймовірність відсутності помилки другого роду, тобто ймовірність відхилення хибної гіпотези H_0:

$$1 - \beta = \mathrm{P}_{H_1}\{h(\mathbf{X}) \in D_k)\} = \mathrm{P}_{H_1}\{\mathbf{X} \in W_k\}.$$

Відзначимо, що і рівень значущості, і потужність критерію одночасно монотонно (у напрямку зростання) залежать від критичної області W_k вибірки $\mathbf{X}$. Зокрема, при $W_k = \varnothing$ ймовірність помилки першого роду нульова, другого роду дорівнює одиниці, а потужність — нульова. При

$W_k = \mathbb{R}^n$ має місце протилежне: похибка першого роду і потужність одночасно дорівнюють одиниці. Тому задачу відшукування оптимального критерію можна звести до задачі умовної оптимізації, яка полягає у знаходженні найбільшої потужності при обмеженні на ймовірність помилки першого роду.

Зауваження. Кожен критерій перевірки статистичної гіпотези можна задавати парою $(h(\mathbf{X}), D_k)$, що утворена статистикою критерію h та її критичною областю D_k (або ж критичною областю W_k вибірки $\mathbf{X}$).

3.2 Найбільш потужні критерії

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із невідомого розподілу F, стосовно якого висувається гіпотеза H_0 при альтернативній гіпотезі H_1.

Означення Статистичний критерій із критичною областю W_k^* вибірки є найбільш потужним критерієм рівня α, якщо його рівень значущості дорівнює α, причому довільний критерій із критичною областю W_k і рівнем значущості α має не більшу потужність, ніж W_k^*:

$$\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} \le \mathrm{P}_{H_1}\{\mathbf{X} \in W_k^*\}.$$

Задача відшукування найбільш потужного критерію не завжди має розв'язок. Однак у випадку простих основної та альтернативної гіпотез найбільш потужний критерій існує. Припустимо, що основна гіпотеза та альтернативна гіпотеза є простими:

$$H_0 : F = G_0, \quad H_1 : F = G_1,$$

де обидва розподіли G_0 та G_1 повністю визначені і абсолютно неперервні. Нехай $G_{\mathbf{X},i}(B) = \mathrm{P}_{H_i}\{\mathbf{X} \in B\}$ (B —

будь-яка борелєва підмножина $\mathbb{R}^n$) — розподіл випадкового вектора $\mathbf{X}$, що відповідає розподілу $G_i,\ i = 0, 1$. Тоді існує щільність $l_{01} \geq 0$ така, що

$$G_{\mathbf{X},1}(B) = \int\limits_B l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}). \tag{3.2}$$

До того ж, випадкову величину $l_{01}(\mathbf{X})$ можна подати у вигляді відношення правдоподібностей

$$l_{01}(\mathbf{X}) = \frac{L_1(\mathbf{X})}{L_0(\mathbf{X})},$$

де $L_i(\mathbf{X})$ — функція правдоподібності вибірки $\mathbf{X}$, яка відповідає розподілу $G_i,\ i = 0, 1$.

Теорема (критерій Неймана-Пірсона). Нехай при заданому рівні значущості $\alpha \in (0, 1)$ існує стала l_α така, що

$$\mathrm{P}_{H_0}\{l_{01}(\mathbf{X}) \geq l_\alpha\} = \alpha.$$

Тоді критерій відношення правдоподібностей $(l_{01}(\mathbf{X}), [l_\alpha, \infty))$ із критичною областю вибірки

$$W_k^* = \{\mathbf{x} \in \mathbb{R}^n :\ l_{01}(\mathbf{x}) \geq l_\alpha\}$$

є найбільш потужним критерієм рівня α.

◀ Розглянемо довільний критерій із критичною областю W_k і рівнем значущості α:

$$\mathrm{P}_{H_0}\{\mathbf{X} \in W_k\} = G_{\mathbf{X},0}(W_k) = \alpha.$$

Обчислимо його потужність з урахуванням формули (3.2):

$$\begin{aligned} &\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} = G_{\mathbf{X},1}(W_k) = \\ &= G_{\mathbf{X},1}(W_k \setminus W_k^*) + G_{\mathbf{X},1}(W_k \cap W_k^*) = \\ &= G_{\mathbf{X},1}(W_k \setminus W_k^*) + G_{\mathbf{X},1}(W_k^*) - G_{\mathbf{X},1}(W_k^* \setminus W_k) = \end{aligned}$$

$$= \int\limits_{W_k \setminus W_k^*} l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}) + G_{\mathbf{X},1}(W_k^*) - \int\limits_{W_k^* \setminus W_k} l_{01}(\mathbf{x}) G_{\mathbf{X},0}(d\mathbf{x}).$$

За означенням критичної області W_k^* при всіх $x \in W_k^*$ виконується нерівність $l_{01}(x) \geq l_\alpha$. Це означає, що якщо $x \notin W_k^*$, то $l_{01}(x) < l_\alpha$. Тоді

$$\begin{aligned} &\mathrm{P}_{H_1}\{\mathbf{X} \in W_k\} \leq l_\alpha G_{\mathbf{X},0}(W_k \setminus W_k^*) + \\ &+ G_{\mathbf{X},1}(W_k^*) - l_\alpha G_{\mathbf{X},0}(W_k^* \setminus W_k) = \\ &= l_\alpha G_{\mathbf{X},0}(W_k) + G_{\mathbf{X},1}(W_k^*) - l_\alpha G_{\mathbf{X},0}(W_k^*) = \\ &= \alpha l_\alpha + G_{\mathbf{X},1}(W_k^*) - \alpha l_\alpha = G_{\mathbf{X},1}(W_k^*) = \mathrm{P}_{H_1}\{\mathbf{X} \in W_k^*\}. \end{aligned}$$

Теорему доведено. ►

Приклад (відшукування найбільш потужного критерію). Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із нормального розподілу з невідомим математичним сподівання a і відомою дисперсією σ^2. Побудувати найбільш потужний критерій для випадку двох простих основної і альтернативної гіпотез

$$H_0 : a = a_0, \quad H_1 : a = a_1,$$

де a_0 і a_1 деякі задані значення.

Для визначеності будемо вважати, що $a_0 < a_1$.

Відношення правдоподібностей має вигляд

$$l_{01}(\mathbf{X}) = \frac{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X_i - a_1)^2}{2\sigma^2}\right\}}{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X_i - a_0)^2}{2\sigma^2}\right\}} =$$

$$= \frac{\exp\left\{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (X_i - a_1)^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (X_i - a_0)^2\right\}} =$$

$$
\begin{aligned}
&= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((X_i - a_1)^2 - (X_i - a_0)^2\right)\right\} = \\
&= \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(a_1 - a_0)(2X_i - (a_1 + a_0))\right\} = \\
&= \exp\left\{\frac{n}{\sigma^2}(a_1 - a_0)\bar{X} - \frac{n}{2\sigma^2}(a_1^2 - a_0^2)\right\},
\end{aligned}
$$

де $\bar{X}$ — вибіркове середнє. Нерівність

$$
l_{01}(\mathbf{X}) = \exp\left\{\frac{n}{\sigma^2}(a_1 - a_0)\bar{X} - \frac{n}{2\sigma^2}(a_1^2 - a_0^2)\right\} \geq l_\alpha
$$

еквівалентна нерівності

$$
\sqrt{n}\frac{\bar{X} - a_0}{\sigma} \geq \frac{\sigma}{\sqrt{n}(a_1 - a_0)}\ln l_\alpha + \frac{\sqrt{n}}{2\sigma}(a_1 - a_0). \tag{3.3}
$$

З теореми про вибіркові моменти нормальної вибірки випливає, що випадкова величина

$$
Z = \sqrt{n}\frac{\bar{X} - a_0}{\sigma}
$$

у лівій частині нерівності (3.3), за припущення H_0, має стандартний нормальний розподіл. Праву частину нерівності позначимо через C. Розв'язавши рівняння $\mathrm{P}_{H_0}\{Z \geq C\} = \alpha$, отримуємо $C = u_{1-\alpha}$ — квантиль рівня $1 - \alpha$ розподілу $N(0, 1)$. Тоді стала l_α, для якої

$$
\mathrm{P}_{H_0}\{l_{01} \geq l_\alpha\} = \alpha,
$$

знаходиться з умови

$$
\frac{\sigma}{\sqrt{n}(a_1 - a_0)}\ln l_\alpha + \frac{\sqrt{n}}{2\sigma}(a_1 - a_0) = u_{1-\alpha}
$$

і вона дорівнює

$$l_\alpha = \exp\left\{\sqrt{n}\frac{a_1 - a_0}{\sigma}\left(u_{1-\alpha} - \frac{\sqrt{n}}{2\sigma}(a_1 - a_0)\right)\right\}$$

Таким чином, ми знайшли сталу l_α для якої

$$\mathrm{P}_{H_0}\{l_{0,1}(\mathbf{X}) \geq l_\alpha\} = \mathrm{P}_{H_0}\{Z \geq u_{1-\alpha}\} = \alpha$$

і, отже, критерій $(Z,\ [u_{1-\alpha}, +\infty))$ із статистикою Z та критичною областю $D_k = [u_{1-\alpha}, +\infty)$ є найбільш потужним критерієм рівня α для перевірки гіпотези H_0 при альтернативній гіпотезі H_1.

3.3 Перевірка гіпотез про значення параметрів і надійні інтервали

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка з розподілу $F(\cdot;\theta)$, який залежить від невідомого параметра θ. Розглянемо гіпотезу $H_0 : \theta = \theta_0$, де θ_0 — гіпотетичне значення параметра θ. Альтернативною будемо вважати гіпотезу $H_1 : \theta \neq \theta_0$. Необхідно за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ зробити висновок про гіпотезу H_0, а саме: відхиляти її чи ні.

Перевірку гіпотези H_0 можна проводити, базуючись на інтервальних оцінках невідомого параметра θ.

Між знаходженням інтервальних оцінок параметрів і перевіркою гіпотез про їх можливі значення існує тісний взаємозв'язок. Це, по суті, два різні способи формулювання однієї й тієї ж задачі. Надійний інтервал для невідомого параметра можна розглядати як множину прийнятних значень для його гіпотетичного значення, а доповнення цього інтервалу — як відповідну критичну область для нульової гіпотези. Тобто, для перевірки гіпотези про значення невідомого параметра достатньо побудувати надійний інтервал для параметра і перевірити, чи потрапляє в нього гіпотетичне значення параметра.

Проілюструємо сказане, розглянувши надійний інтервал $(\hat{\theta}_1, \hat{\theta}_2)$ для параметра θ з рівнем надійності $1 - \alpha$. Будемо вважати, що цей інтервал побудовано загальним методом опорних величин. Тоді для будь-якої реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ виконується співвідношення

$$\hat{\theta}_1 < \theta < \hat{\theta}_2 \Longleftrightarrow h_1 < h(\mathbf{x}; \theta) < h_2,$$

де $h(\mathbf{x}; \theta)$ — значення деякої опорної величини $h(\mathbf{X}; \theta)$, а h_1, h_2 — значення, знайдені з умови

$$\mathrm{P}\{h_1 < h(\mathbf{X}; \theta) < h_2\} = 1 - \alpha.$$

Тоді $h(\mathbf{X}; \theta_0)$ та $D_k = D \setminus (h_1, h_2)$ є відповідно статистикою критерію та критичною областю для перевірки гіпотези $H_0 : \theta = \theta_0$ при заданому рівні значущості α і при альтернативній гіпотезі $H_1 : \ \theta \neq \theta_0$. Це означає, що у випадку, коли $h(\mathbf{x}; \theta_0) \in D_k$ (або $\theta_0 \notin (\hat{\theta}_1, \hat{\theta}_2)$), гіпотеза H_0 відхиляється і приймається альтернативна гіпотеза H_1.

На завершення, для гіпотези H_0 розглянемо ще два варіанти альтернативних гіпотез: $H_1' : \theta < \theta_0$, $H_1'' : \theta > \theta_0$. Критерії з альтернативними гіпотезами H_1' та H_1'' називаються односторонніми — лівостороннім та правостороннім відповідно. Якщо альтернативною є гіпотеза H_1', то основна гіпотеза H_0 відхиляється у випадку, коли гіпотетичне значення θ_0 невідомого параметра θ не потрапляє у лівосторонній надійний інтервал $(-\infty, \hat{\theta}_2)$ для θ з рівнем надійності $1 - \alpha$; якщо альтернативною є гіпотеза H_1'', то

H_0 відхиляється у випадку, коли θ_0 не потрапляє у відповідний правосторонній надійний інтервал $(\hat{\theta}_1, \infty)$.

Зауваження. Наведені у цьому параграфі міркування поширюються на випадок векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

3.4 Перевірка гіпотез про значення параметрів нормального розподілу

Нехай $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ — реалізація вибірки $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ із розподілу $N(a, \sigma^2)$. Побудуємо критерії для перевірки різних варіантів гіпотез про значення математичного сподівання a і дисперсії σ^2.

Перевірка гіпотези про значення математичного сподівання при відомій дисперсії. Перевіримо гіпотезу $H_0 : a = a_0$ при альтернативній гіпотезі $H_1 : a \neq a_0$. Точковою оцінкою невідомого математичного сподівання a є вибіркове середнє $\bar{X}$. З теореми про вибіркові моменти нормальної вибірки випливає, що величина

$$Z = \sqrt{n}\frac{\bar{X} - a_0}{\sigma},$$

за припущення H_0, має стандартний нормальний розподіл $N(0, 1)$. Це означає, що

$$\mathrm{P}_{H_0}\left\{|Z| \geq u_{1-\frac{\alpha}{2}}\right\} = \alpha,$$

де $u_{1-\frac{\alpha}{2}}$ — квантиль рівня $1 - \frac{\alpha}{2}$ стандартного нормального розподілу.

Тоді критерій для перевірки гіпотези H_0 при альтернативній гіпотезі H_1 полягає в тому, що гіпотеза H_0 відхиляється, якщо значення $|Z|$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $D_k = [u_{1-\frac{\alpha}{2}}, +\infty)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки при відхиленні H_0 дорівнює α.

Якщо альтернативною є гіпотеза $H_1' : a < a_0$ чи $H_1'' : a > a_0$, то критична область статистики Z має вигляд $D_k' = (-\infty, u_\alpha]$ чи $D_k'' = [u_{1-\alpha}, +\infty)$ відповідно.

Перевірка гіпотези про значення математичного сподівання при невідомій дисперсії. Перевіримо тепер сформульовану вище гіпотезу H_0 при альтернативній гіпотезі H_1 у випадку, коли дисперсія σ^2 невідома. Поряд з точковою оцінкою $\bar{X}$ математичного сподівання a розглянемо оцінку S^2 для дисперсії σ^2. Згідно з теоремою про статистику Стьюдента від нормальної вибірки, випадкова величина

$$\tau_{n-1} = \sqrt{n}\frac{\bar{X} - a_0}{S},$$

за гіпотези H_0, має розподіл Стьюдента з $n-1$ ступенями свободи. Тому

$$\mathrm{P}_{H_0}\left\{|\tau_{n-1}| \geq t_{1-\frac{\alpha}{2}}(n-1)\right\} = \alpha, \tag{3.4}$$

де $t_{1-\frac{\alpha}{2}}(n-1)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n-1$ ступенями свободи.

Таким чином, критерій з рівнем значущості α для перевірки гіпотези H_0 при альтернативній гіпотезі H_1 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення $|\tau_{n-1}|$ потрапляє у критичну область $D_k = [t_{1-\frac{\alpha}{2}}(n-1), +\infty)$, і не відхиляється у протилежному разі.

У випадку, коли альтернативною є одна з гіпотез $H_1' : a < a_0$, $H_1'' : a > a_0$, критична область статистики Z є відповідно однією з областей: $D'_k = (-\infty, t_\alpha(n-1)]$, $D''_k = [t_{1-\alpha}(n-1), +\infty)$.

Перевірка гіпотези про значення дисперсії при відомому математичному сподіванні. Перевіримо гіпотезу $H_0 : \sigma^2 = \sigma_0^2$ при альтернативній гіпотезі $H_1 : \sigma^2 \neq \sigma_0^2$. Нехай математичне сподівання a — відоме. Найкращою точковою оцінкою невідомої дисперсії σ^2 при відомому математичному сподіванні a є $\hat{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^2$. За гіпотези H_0 випадкова величина

$$\chi_n^2 = n\frac{\hat{\sigma^2}}{\sigma_0^2}$$

має χ^2-розподіл з n ступенями свободи, як сума квадратів n стандартних нормальних величин. Тому

$$\mathrm{P}_{H_0}\left\{\chi^2_{\frac{\alpha}{2}}(n) < \chi_n^2 < \chi^2_{1-\frac{\alpha}{2}}(n)\right\} = 1 - \alpha,$$

або

$$\mathrm{P}_{H_0}\{\chi_n^2 \in D_k\} = \alpha, \quad D_k = [0, \chi^2_{\frac{\alpha}{2}}(n)] \cup [\chi^2_{1-\frac{\alpha}{2}}(n), +\infty),$$

де $\chi^2_\gamma(n)$ — квантиль рівня γ розподілу χ^2 з n ступенями свободи.

Отже, критерій з рівнем значущості α для перевірки перевірки гіпотези H_0 при альтернативній гіпотезі H_1 полягає в тому, що гіпотеза H_0 відхиляється, якщо значення статистики χ_n^2 потрапляє у критичну область D_k, і не відхиляється у протилежному разі.

Якщо альтернативною є гіпотеза $H'_1 : \sigma^2 < \sigma_0^2$ або гіпотеза $H''_1 : \sigma^2 > \sigma_0^2$, то відповідні критичні області статистики χ_n^2 виглядають так: $D'_k = [0, \chi^2_\alpha(n)]$, $D''_k = [\chi^2_{1-\alpha}(n), +\infty)$.

Перевірка гіпотези про значення дисперсії при невідомому математичному сподіванні. Перевіримо тепер гіпотезу H_0 при альтернативній гіпотезі H_1 у випадку, коли a — невідоме. Точковою оцінкою дисперсії σ^2 при невідомому математичному сподіванні є виправлена вибіркова дисперсія S^2. Якщо справджується гіпотеза H_0, то за теоремою про вибіркові моменти нормальної вибірки випадкова величина

$$\chi^2_{n-1} = (n-1)\frac{S^2}{\sigma_0^2}$$

має χ^2-розподіл з $n-1$ ступенями свободи. Тоді

$$\begin{aligned}
&\mathrm{P}_{H_0}\left\{\chi^2_{n-1} \in D_k\right\} = \alpha, \\
&D_k = [0, \chi^2_{\frac{\alpha}{2}}(n-1)] \cup [\chi^2_{1-\frac{\alpha}{2}}(n-1), +\infty).
\end{aligned}$$

Таким чином, правило перевірки гіпотези H_0 при альтернативній гіпотезі H_1 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення статистики χ^2_n потрапляє у критичну область D_k, і не відхиляється у протилежному разі. При цьому, ймовірність помилки першого роду дорівнює α.

При альтернативній гіпотезі $H_1' : \sigma^2 < \sigma_0^2$ чи $H_1'' : \sigma^2 > \sigma_0^2$ критична область статистики критерію є однією з областей $D_k' = [0, \chi^2_\alpha(n-1)]$, $D_k'' = [\chi^2_{1-\alpha}(n-1), +\infty)$ відповідно.

3.5 Критерій Стьюдента

Нехай $X = (X_1, X_2, \dots, X_n)$ та $Y = (Y_1, Y_2, \dots, Y_m)$ — незалежні вибірки з розподілів $N(a_1, \sigma^2)$ та $N(a_2, \sigma^2)$ відповідно. Математичні сподівання a_1, a_2 — невідомі, дисперсія σ^2 — однакова для обидвох розподілів і також невідома. Щодо параметрів a_1 та a_2 висувається гіпотеза $H_0 : a_1 - a_2 = a_0$, де a_0 — деяка відома стала. Зокрема, при $a_0 = 0$, гіпотеза H_0 полягає в тому, що математичні сподівання a_1, a_2 рівні між собою, або, що те саме, $\mathbf{X}$ та $\mathbf{Y}$ є вибірками з одного й того ж нормального розподілу. Альтернативною будемо вважати гіпотезу $H_1 : a_1 - a_2 \neq a_0$. Необхідно за реалізаціями $\mathbf{x}$ та $\mathbf{y}$ цих вибірок зробити висновок: відхиляти чи не відхиляти нульову гіпотезу.

Побудуємо критерій Стьюдента для перевірки гіпотези H_0. Нехай $\bar{X}, \bar{Y}$ — відповідні вибіркові середні для $\mathbf{X}$ та $\mathbf{Y}$, а $S_1^2,\ S_2^2$ — виправлені вибіркові дисперсії. Тоді з незалежності вибірок і теореми про вибіркові моменти нормальної вибірки випливає, що величини $\bar{X}, \bar{Y}, S_1^2, S_2^2$ незалежні. До того ж, величини $\bar{X}, \bar{Y}$ мають нормальний розподіл $N\left(a_1, \frac{\sigma^2}{n}\right)$, $N\left(a_2, \frac{\sigma^2}{m}\right)$ відповідно, а величини $(n-1)\frac{S_1^2}{\sigma^2}$, $(m-1)\frac{S_2^2}{\sigma^2}$ — розподіли χ^2 з $n-1$ та $m-1$ ступенями свободи відповідно. Тому випадкова величина

$$\sqrt{\frac{nm}{n+m}}\frac{\bar{X}-\bar{Y}-a_0}{\sigma}$$

за припущення H_0 має стандартний нормальний розподіл і не залежить від суми

$$(n-1)\frac{S_1^2}{\sigma^2} + (m-1)\frac{S_2^2}{\sigma^2},$$

яка за означенням має χ^2-розподіл з $(n-1)+(m-1) = n+m-2$ ступенями свободи.

За означенням розподілу Стьюдента величина, що утворена відношенням

$$\tau_{n+m-2} = \sqrt{\frac{nm}{n+m}}\frac{\bar{X}-\bar{Y}-a_0}{\left(\sqrt{(n-1)S_1^2 + (m-1)S_2^2)/(n+m-2}\right)},$$

у випадку істинності гіпотези H_0, має розподіл Стьюдента з $n+m-2$ ступенями свободи і тому

$$\mathrm{P}_{H_0}\left\{|\tau_{n+m-2}| \geq t_{1-\frac{\alpha}{2}}(n+m-2)\right\} = \alpha,$$

де $t_{1-\frac{\alpha}{2}}(n+m-2)$ — квантиль рівня $1-\frac{\alpha}{2}$ розподілу Стьюдента з $n+m-2$ ступенями свободи.

Отже, критерій Стьюдента переверки гіпотези H_0 можна сформулювати так: гіпотеза H_0 відхиляється, якщо значення $|\tau_{n+m-2}|$, підраховане за реалізаціями $\mathbf{x}, \mathbf{y}$ вибірок $\mathbf{X}, \mathbf{Y}$ відповідно, потрапляє у критичну область $D_k = \left[t_{1-\frac{\alpha}{2}}(n+m-2), +\infty\right)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки при відхиленні H_0 (помилки першого роду) дорівнює α.

Вправа. Побудувати критерій для перевірки гіпотези про рівність математичних сподівань двох незалежних нормальних вибірок з відомими дисперсіями, що є різними.

3.6 Критерій Фішера

Нехай $X = (X_1, X_2, \ldots, X_n)$ та $Y = (Y_1, Y_2, \ldots, Y_m)$ — незалежні вибірки з розподілів $N(a_1, \sigma_1^2)$ та $N(a_2, \sigma_2^2)$ відповідно. Параметри $a_1, a_2, \sigma_1^2, \sigma_2^2$ — невідомі. Щодо дисперсій σ_1^2 та σ_2^2 висувається гіпотеза $H_0 : \sigma_1^2 = \sigma_2^2$, тобто гіпотеза, яка полягає в тому, що $\mathbf{X}$ та $\mathbf{Y}$ є вибірками з нормального розподілу з однією й тією самою дисперсією. Альтернативною будемо вважати гіпотезу $H_1 : \sigma_1^2 \neq \sigma_2^2$. Необхідно за реалізаціями $\mathbf{x}$ та $\mathbf{y}$ цих вибірок зробити висновок: відхиляти чи не відхиляти нульову гіпотезу.

Побудуємо критерій Фішера для перевірки гіпотези H_0. Нехай S_1^2 та S_2^2 — виправлені вибіркові дисперсії для вибірок $\mathbf{X}$ та $\mathbf{Y}$ відповідно. Тоді з незалежності вибірок випливає, що величини S_1^2, S_2^2 незалежні. За теоремою про вибіркові моменти нормальної вибірки величини $(n-1)\frac{S_1^2}{\sigma_1^2}$ та $(m-1)\frac{S_2^2}{\sigma_2^2}$ мають розподіли χ^2 з $n-1$ та $m-1$ ступенями свободи відповідно і тому, у випадку істинності H_0, відношення

$$\phi_{n-1,m-1} = \frac{S_1^2}{S_2^2}$$

має розподіл Фішера з $n-1,\ m-1$ ступенями свободи. Це означає, що

$$\mathrm{P}_{H_0}\left\{f_{\frac{\alpha}{2}}(n-1,m-1) < \phi_{n-1,m-1} < f_{1-\frac{\alpha}{2}}(n-1,m-1)\right\} = \\ = 1-\alpha,$$

або

$$\mathrm{P}_{H_0}\left\{\phi_{n-1,m-1} \in D_k\right\} = \alpha, \\ D_k = [0, f_{\frac{\alpha}{2}}(n-1,m-1)] \cup [f_{1-\frac{\alpha}{2}}(n-1,m-1), +\infty),$$

де $f_\gamma(n-1,m-1)$ — квантиль рівня γ розподілу Фішера з $n-1,\ m-1$ ступенями свободи.

Отже, критерій Фішера з рівнем значущості α полягає в тому, що якщо значення $\phi_{n-1,m-1}$, підраховане за реалізаціями $\mathbf{x}, \mathbf{y}$ вибірок $\mathbf{X}, \mathbf{Y}$ відповідно, потрапляє у критичну область D_k, то гіпотеза H_0 відхиляється і не відхиляється у протилежному разі.

Вправа. Сформулювати критерій Фішера для випадку, коли a_1, a_2 — відомі.

3.7 Критерій χ^2

Нехай $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ — вибірка із невідомого розподілу F, стосовно якого висувається гіпотеза $H_0 : F =$

G, де G належить заданому класу розподілів (зокрема, G може бути повністю визначеним розподілом). Інакше кажучи, гіпотезу H_0 можна сформулювати так: $\mathbf{X}$ є вибіркою з розподілу G. Необхідно за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ зробити висновок про гіпотезу H_0, а саме: відхиляти її чи ні.

Критерій χ^2 (гіпотетичний розподіл визначений). Ідея побудови критерію для перевірки гіпотези H_0 грунтується на тому, що емпіричний розподіл $\hat{F}_n(x)$, побудований за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$ з F, мало відрізняється від розподілу F. Справді, за теоремою про властивості емпіричної функції розподілу, для кожного $x \in \mathbb{R}$ значення емпіричної функції розподілу $\hat{F}_n(x)$ є незміщеною і консистентною оцінкою $F(x)$. Тому, якщо ввести відхилення $D(\hat{F}_n, G)$ емпіричного розподілу від гіпотетичного, причому так, щоб воно набувало малих значень, коли гіпотеза H_0 справджується, і великих, коли вона не справджується, то гіпотезу H_0 природно відхиляти або не відхиляти залежно від того, якого значення набуло відхилення $D(\hat{F}_n, G)$ для конкретної реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ — великого чи малого.

Відхилення $D(\hat{F}_n, G)$ можна будувати різними способами. Критерій χ^2 побудований на підставі відхилення $\hat{F}_n$ від G, запропонованого К. Пірсоном. Це відхилення будується так: область значень гіпотетичого розподілу G розбивається на скінченну кількість неперетинних множин $\Delta_i,\ i = \overline{1,k}$, і як відхилення розглядається

$$D(\hat{F}_n, G) = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{\nu_i}{n} - p_i \right) = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i},$$

де p_i — ймовірність того, що випадкова величина X_j потрапить до множини Δ_i обчислена за гіпотетичним розподілом G, тобто $p_i = \mathrm{P}\{X_j \in \Delta_i\} = G(\Delta_i)$; $\nu_i = \nu_i(\mathbf{X}) = \sum\limits_{j=1}^{n} \mathbb{I}_{\{X_j \in \Delta_i\}}$ — випадкова величина, яка при кожній реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ дорівнює кількості значень серед x_1, $x_2, \ldots, x_n$, що потрапили до Δ_i.

Розглянемо випадкову величину $\hat{p}_i = \frac{\nu_i}{n} = \frac{1}{n} \sum\limits_{j=1}^{n} I_{\{X_j \in \Delta_i\}}$, $i = \overline{1,k}$. При фіксованих значеннях вибірки $\hat{p}_i$ — це ймовірність того, що випадкова величина X_j потрапить до множини Δ_i, обчислена за емпіричним розподілом $\hat{F}_n$. Цілком аналогічно доведенню теореми про властивості емпіричної функції розподілу встановлюється, що випадкова величина $\hat{p}_i$ є незміщеною та консистентною оцінкою ймовірності $\mathrm{P}\{X_j \in \Delta_i\} = F(\Delta_i)$. Це означає, що у випадку істинності H_0 ця оцінка є незміщеною та консистентною для p_i, і тому відхилення $D(\hat{F}_n, G)$ у цьому випадку мінімальне (порівняно з відхиленнями $D(\hat{F}_n, G^*)$, коли розподіли G^* відмінні від F).

Теорема Пірсона. *Якщо справджується гіпотеза H_0, то для всіх $x > 0$*

$$\lim_{n\to\infty} \mathrm{P}\{D(\hat{F}_n, G) < x\} = F_{\chi^2_{k-1}}(x),$$

де $F_{\chi^2_{k-1}}(x)$ — функція χ^2-розподілу з $k-1$ ступенями свободи.

◀ Доведення проведемо для випадку $k = 2$. У цьому випадку $\nu_2 = n - \nu_1$, $p_2 = 1 - p_1$. Випадкову величину $D(\hat{F}_n, G)$ подамо у вигляді

$$D(\hat{F}_n, G) = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \frac{(\nu_1 - np_1)^2}{np_1} +$$

$$+ \frac{(n - \nu_1 - n(1-p_1))^2}{n(1-p_1)} = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(-\nu_1 + np_1)^2}{n(1-p_1)} =$$

$$= \frac{(\nu_1 - np_1)^2}{n}\left(\frac{1}{p_1} + \frac{1}{1-p_1}\right) =$$
$$= \frac{(\nu_1 - np_1)^2}{np_1(1-p_1)} = \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2.$$

Оскільки величина ν_1 є сумою n незалежних однаково розподілених випадкових величин: $\nu_1 = \sum\limits_{j=1}^{n} I_{\{X_j \in \Delta_1\}}$ і її моменти

$$\mathrm{M}\nu_1 = \sum_{j=1}^{n} \mathrm{M}I_{\{X_j \in \Delta_1\}} = np_1,$$
$$\mathrm{D}\nu_1 = \sum_{j=1}^{n} \mathrm{D}I_{\{X_j \in \Delta_1\}} = np_1(1-p_1),$$

за центральною граничною теоремою для всіх $x \in \mathbb{R}$

$$\mathrm{P}\left\{\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} < x\right\} \longrightarrow F_0(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}dt,$$

при $n \to \infty$. Останнє співвідношення означає, що граничний розподіл випадкової величини $Z_n = \frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}$ співпадає зі стандартним нормальним розподілом. Тоді граничний розподіл випадкової величини Z_n^2 співпадає із χ^2 розподілом з одним ступенем свободи. Отже, для всіх $x > 0$

$$\mathrm{P}\{D(\hat{F}_n, G) < x\} = \mathrm{P}\{Z_n^2 < x\} \longrightarrow F_{\chi^2_{k-1}}(x),$$

при $n \to \infty$, а це й треба було довести. ▶

З теореми Пірсона випливає, що при великих обсягах вибірки

$$\mathrm{P}\{D(\hat{F}_n, G) \geq \chi^2_{1-\alpha}(k-1) | H_0\} \approx \alpha,$$

де α — заданий рівень значущості, $\chi^2_{1-\alpha}(k-1)$ — квантиль рівня $1-\alpha$ розподілу χ^2 з $k-1$ ступенями свободи. Отже критерій χ^2 з рівнем значущості α полягає у тому, що гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $[\chi^2_{1-\alpha}(k-1), +\infty)$, і не відхиляється у протилежному разі.

Зауваження. Критерієм χ^2 можна користуватися у випадку, коли $np_i \geq 10$, $i = \overline{1,k}$. Якщо для деяких i ця умова не виконується, то відповідні множини Δ_i потрібно об'єднати з сусідніми.

Критерій χ^2 (гіпотетичний розподіл невизначений). Нехай розподіл G визначений з точністю до невідомого векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$, який набуває значень із множини $\Theta \subset \mathbb{R}^s$. У цьому випадку гіпотеза $H_0 : F(\cdot) = G(\cdot;\theta)$ означає, що $\mathbf{X}$ є вибіркою із розподілу, що належить до класу розподілів $G(\cdot;\theta)$, $\theta \in \Theta$.

Оскільки параметр θ невідомий, замість нього будемо використовувати значення його оцінки $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, і отже, як гіпотетичний розглядати розподіл $G(\cdot;\hat{\theta})$. Р. Фішер встановив, що коли гіпотеза H_0 справджується і оцінка $\hat{\theta}$ є такою, що її значення при кожній реалізації x вибірки X мінімізує відхилення

$$D(\hat{F}_n, G) = \sum_{i=1}^{k} \frac{(\nu_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})},$$

між $\hat{F}_n$ і G, то розподіл $D(\hat{F}_n, G)$ при $n \to \infty$ прямує до χ^2-розподілу з $(k-1-s)$ ступенями свободи, де $s = \dim(\theta)$ — розмірність векторного параметра θ.

Отже, у випадку, коли шукану оцінку $\hat{\theta}$ знайдено, критерієм χ^2 з рівнем значущості α можна користуватися у такому формулюванні: гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$, підраховане за реалізацією $\mathbf{x}$ вибірки $\mathbf{X}$, потрапляє у критичну область $[\chi^2_{1-\alpha}(k-1-s), +\infty)$, і не відхиляється у протилежному разі.

Зауваження. Оцінку $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$, що мінімізує відхилення $D(\hat{F}_n, G)$ можна отримати як оцінку максимальної правдоподібності для θ, побудовану за частотами $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$. Функція правдоподібності має вигляд:

$$L(\theta) = \prod_{i=1}^{k} (p_i(\theta))^{\nu_i},$$

де

$$\sum_{i=1}^{k} p_i(\theta) = 1, \quad \sum_{i=1}^{k} \nu_i(\theta) = n.$$

Критерій χ^2 як критерій незалежності. Нехай X та Y — дві дискретні випадкові величини, які можуть набувати значення $a_1, a_2, \ldots a_s$ та $b_1, b_2, \ldots b_r$ відповідно. За результатами n спостережень $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ випадкового вектора (X, Y) потрібно перевірити гіпотезу H_0 : випадкові величини X та Y — незалежні. Позначимо

$$p_{ij} = \mathrm{P}\{X = a_i, Y = b_j\}, \quad i = \overline{1, s},\ j = \overline{1, r},$$

$$p_{i\bullet} = \mathrm{P}\{X = a_i\} = \sum_{j=1}^{r} p_{ij}, \quad p_{\bullet j} = \mathrm{P}\{Y = b_j\} = \sum_{i=1}^{s} p_{ij}.$$

Тоді гіпотезу H_0 можна сформулювати так:

$$H_0 : p_{ij} = p_{i\bullet} p_{\bullet j}, \quad i = \overline{1, s},\ j = \overline{1, r},$$

або так: вибірка[1]

$$(\mathbf{X},\mathbf{Y}) = ((X_1,Y_1),(X_2,Y_2),\ldots(X_n,Y_n))$$

є вибіркою із розподілу $G:\ \ ((a_i,b_j);p_{i\bullet}p_{\bullet j}),\ i=\overline{1,s},\ j=\overline{1,r}$, визначеного з точністю до $(s+r-2)$-вимірного векторного параметра

$$\theta = (p_{1\bullet},p_{2\bullet},\ldots p_{s-1\bullet},p_{\bullet 1},p_{\bullet 2},\ldots,p_{\bullet r-1})$$

(значення $p_{s\bullet}$ та $p_{\bullet r}$ знаходяться з рівностей $p_{s\bullet} = 1-\sum\limits_{i=1}^{s-1} p_{i\bullet}$, $p_{\bullet r} = 1-\sum\limits_{j=1}^{r-1} p_{\bullet j}$).

Для перевірки цієї гіпотези скористаємося критерієм χ^2. Область значень гіпотетичного розподілу розіб'ємо на множини $\Delta_{ij},\ i=\overline{1,s},\ j=\overline{1,r}$, кожна з яких складається лише з однієї точки (a_i,b_j). Нехай ν_{ij} — кількість значень серед (x_1,y_1), $(x_2,y_2),\ldots(x_n,y_n)$, що потрапили до Δ_{ij}. За частотами (ν_{ij}) знайдемо оцінку максимальної правдоподібності невідомого векторного параметра θ. У випадку істинності H_0, функція правдоподібності $L(\theta)$ має вигляд:

$$L(\theta) = \prod_{i=1}^{s}\prod_{j=1}^{r} p_{ij}^{\nu_{ij}} = \prod_{i=1}^{s} p_{i\bullet}^{\nu_{i\bullet}} \prod_{j=1}^{r} p_{\bullet j}^{\nu_{\bullet j}},$$

де

$$\nu_{i\bullet} = \sum_{j=1}^{r}\nu_{ij}, \quad \nu_{\bullet j} = \sum_{i=1}^{s}\nu_{ij}.$$

[1] Під багатовимірною вибіркою розуміємо впорядкований набір незалежних і однаково розподілених випадкових векторів $X_1,X_2,\ldots,X_n$. Усі вищенаведені результати поширюються на випадок багатовимірної вибірки $\mathbf{X}$.

Тоді

$$\ln L(\theta) = \sum_{i=1}^{s} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r} \nu_{\bullet j} \ln p_{\bullet j} =$$

$$= \sum_{i=1}^{s-1} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r-1} \nu_{\bullet j} \ln p_{\bullet j} + \nu_{s\bullet} \ln p_{s\bullet} + \nu_{\bullet r} \ln p_{\bullet r} =$$

$$= \sum_{i=1}^{s-1} \nu_{i\bullet} \ln p_{i\bullet} + \sum_{j=1}^{r-1} \nu_{\bullet j} \ln p_{\bullet j} +$$

$$+ \left(n - \sum_{i=1}^{s-1} \nu_{i\bullet} \right) \ln \left(1 - \sum_{i=1}^{s-1} p_{i\bullet} \right) +$$

$$+ \left(n - \sum_{j=1}^{r-1} \nu_{\bullet j} \right) \ln \left(1 - \sum_{j=1}^{r-1} p_{\bullet j} \right).$$

Продиференціювавши останню рівність за $p_{1\bullet}, p_{2\bullet}, \ldots, p_{s-1\bullet}$, $p_{\bullet 1}, p_{\bullet 2}, \ldots, p_{\bullet r-1}$, отримуємо систему рівнянь

$$\frac{\partial \ln L(\theta)}{\partial p_{i\bullet}} = \frac{\nu_{i\bullet}}{p_{i\bullet}} - \frac{n - \sum\limits_{i=1}^{s-1} \nu_{i\bullet}}{1 - \sum\limits_{i=1}^{s-1} p_{i\bullet}} = 0, \quad i = \overline{1, s-1},$$

$$\frac{\partial \ln L(\theta)}{\partial p_{\bullet j}} = \frac{\nu_{\bullet j}}{p_{\bullet j}} - \frac{n - \sum\limits_{j=1}^{r-1} \nu_{\bullet j}}{1 - \sum\limits_{j=1}^{r-1} p_{\bullet j}} = 0, \quad j = \overline{1, r-1}.$$

Розв'язавши цю систему, встановлюємо, що оцінка максимальної правдоподібності параметра θ має вигляд

$$\hat{p}_{i\bullet} = \frac{\nu_{i\bullet}}{n}, \quad \hat{p}_{\bullet j} = \frac{\nu_{\bullet j}}{n},$$

тобто

$$\hat{\theta} = \frac{1}{n}(\nu_{1\bullet}, \nu_{2\bullet}, \ldots, \nu_{s-1\bullet}, \nu_{\bullet 1}, \nu_{\bullet 2}, \ldots, \nu_{\bullet r-1}).$$

Отже, у випадку, коли гіпотеза H_0 справджується, розподіл відхилення

$$D(\hat{F}_n, G) = \sum_{i=1}^{s} \sum_{i=1}^{r} \frac{(\nu_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}} =$$
$$= n\sum_{i=1}^{s} \sum_{i=1}^{r} \frac{(\nu_{ij} - \hat{\nu}_{i\bullet}\hat{\nu}_{\bullet j}/n)^2}{\hat{\nu}_{i\bullet}\hat{\nu}_{\bullet j}}$$

прямує до χ^2-розподілу з $rs-1-(s+r-2) = (s-1)(r-1)$ ступенями свободи. Відповідно до критерію χ^2, гіпотеза H_0 відхиляється, якщо значення $D(\hat{F}_n, G)$ потрапляє у критичну область $[\chi^2_{1-\alpha}((s-1)(r-1)), +\infty)$, і не відхиляється у протилежному разі. При цьому, ймовірність помилки першого роду дорівнює α.

Розділ 4

Нормальна лінійна регресія

4.1 Метод найменших квадратів

У процесі практичного застосування методів математичної статистики часто виникає задача, найпростіший варіант якої сформульовано нижче.

Припустимо, що між величинами y та x існує лінійна залежність виду

$$y = \alpha + \beta x, \tag{4.1}$$

де x — деяка невипадкова змінна, α, β — невідомі коефіцієнти, які необхідно визначити експериментально. Описану залежність величини y від x називають простою лінійною регресією, а коефіцієнти α, β — коефіцієнтами регресії.

Нехай відбувається n незалежних експериментів. У кожному з них величина x набуває деяких цілком визначених значень x_i, що є відомими і, взагалі кажучи, різними для різних експериментів. Припустимо, що в кожному експерименті спостерігається нормальна випадкова величина Y_i, $i = \overline{1, n}$, з математичним сподіванням

$$\mathrm{M}Y_i = \alpha + \beta x_i, \quad i = \overline{1, n}, \tag{4.2}$$

і невідомою дисперсією σ^2. Таким чином, розглядається вибірка $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, утворена послідовністю незалежних нормально розподілених випадкових величин з однаковими дисперсіями σ^2 і математичними сподіваннями вигляду (4.2).

Запишемо рівняння регресії у вигляді

$$y = a + b(x - \bar{x}),$$

де

$$a = \alpha + b\bar{x}, \quad b = \beta, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Оцінимо коефіцієнти a, b.

Функція правдоподібності вибірки $\mathbf{Y}$ для довільної її реалізації $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ має вигляд

$$L(a, b, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - \mathrm{M}Y_i)^2}{2\sigma^2}\right\} =$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - b(x_i - \bar{x}))^2\right\}.$$

Оцінки трьох невідомих параметрів a, b, σ^2 побудуємо методом максимальної правдоподібності. Для того, щоб відшукати максимум функції правдоподібності, спочатку мінімізуємо квадратичну форму

$$Q(a, b) = \sum_{i=1}^{n}(y_i - a - b(x_i - \bar{x}))^2,$$

а потім, знайдемо максимум за θ функції

$$\varphi(\theta) = (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{Q^*}{2\theta}\right\},$$

де $Q^* = \min\limits_{a,b} Q(a,b)$. Оскільки

$$\begin{aligned}Q(a,b) &= \sum_{i=1}^{n} \left((\bar{x}-a) + (x_i - \bar{x}) - b(x_i - \bar{x})\right)^2 = \\ &= n\left((\bar{x}-a)^2 + b^2 s_1^2 + s_2^2 - 2bk_{12}\right) = \\ &= n\left((\bar{x}-a)^2 + \left(s_1 b - \frac{k_{12}}{s_1}\right)^2 + s_2^2 - \frac{k_{12}^2}{s_1^2}\right),\end{aligned}$$

де

$$s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$k_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

то мінімум Q^* функції $Q(a,b)$ досягається при

$$a = \hat{a} = \bar{x}, \quad b = \hat{b} = \frac{k_{12}}{s_1^2}$$

і дорівнює

$$Q^* = \min_{a,b} Q(a,b) = \sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2 = ns_2^2(1 - \hat{r}^2),$$

де

$$\hat{r} = \frac{k_{12}}{s_1 s_2} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Знайдемо тепер значення θ, при якому функція $\varphi(\theta)$ досягає максимуму. З огляду на те, що функції $\varphi(\theta)$ і $\ln\varphi(\theta)$ досягають максимуму в одній і тій самій точці, шукане значення θ є розв'язком рівняння (необхідна умова екстремуму):

$$(\ln\varphi(\theta))' = 0.$$

Спростивши похідну у правій частині цього рівняння, отримуємо:

$$(\ln\varphi(\theta))' = \left(-\frac{n}{2}\ln(2\pi\theta) - \frac{Q^*}{2\theta}\right)' = -\frac{n}{2\theta} + \frac{Q^*}{2\theta^2} = 0.$$

Звідси

$$\hat{\theta} = \hat{\sigma^2} = \frac{Q^*}{n} = s_2^2(1 - \hat{r}^2).$$

Отже, оцінками невідомих параметрів a, b, σ^2 є

$$\hat{a} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{b} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}, \tag{4.3}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2.$$

Зауваження. Метод оцінювання параметрів, що полягає у мінімізації квадратичної форми $Q(a, b)$, називається методом найменших квадратів, а відповідні оцінки — оцінками методу найменших квадратів.

4.2 Властивості оцінок методу найменших квадратів

Дослідимо властивості побудованих вище оцінок $\hat{a}$, $\hat{b}$, $\hat{\sigma^2}$. Відзначимо, що

$$\mathrm{M}\hat{a} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}Y_i = \frac{1}{n}\sum_{i=1}^{n}(a + b(x_i - \bar{x})) = a,$$

$$\mathrm{M}\hat{b} = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})\mathrm{M}(Y_i - \bar{Y}) =$$

$$= \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})\mathrm{M}(Y_i - \hat{a}) =$$

$$= \frac{b}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{b}{ns_1^2}ns_1^2 = b,$$

а тому оцінки $\hat{a}$ та $\hat{b}$ є незміщеними. Оскільки величини Y_i незалежні,

$$\mathrm{D}\hat{a} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}Y_i = \frac{\sigma^2}{n}.$$

Враховуючи те, що

$$\hat{b} = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}) = \frac{1}{ns_1^2}\sum_{i=1}^{n}(x_i - \bar{x})Y_i,$$

отримуємо

$$\mathrm{D}\hat{b} = \frac{1}{n^2s_1^4}\sum_{i=1}^{n}(x_i - \bar{x})^2\mathrm{D}Y_i = \frac{1}{n^2s_1^4}\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2 = \frac{\sigma^2}{ns_1^2}.$$

З отриманих формул для дисперсій оцінкок $\hat{a}, \hat{b}$ і з другої нерівності Чебишева випливає, що ці оцінки є консистентними для параметрів a, b відповідно.

Далі, розглянемо квадратичну форму $Q(a,b)$. Оскільки випадкові величини $U_i = Y_i - a - b(x_i - \bar{x})$ незалежні і мають розподіл $N(0,\sigma^2)$, то величина $\frac{Q(a,b)}{\sigma^2}$ має χ^2-розподіл з n ступенями свободи. Записавши $Q(a,b)$ у вигляді

$$Q(a,b) = \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}(x_i - \bar{x}) + (\hat{a} - a) + (\hat{b} - b)(x_i - \bar{x}))^2,$$

нескладно перевірити, що

$$Q(a,b) = \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2 + n(\hat{a} - a)^2 + ns_1^2(\hat{b} - b)^2 =$$
$$= n\hat{\sigma^2} + n(\hat{a} - a)^2 + ns_1^2(\hat{b} - b)^2.$$

Тому величини

$$\chi^2_{n-2} = \frac{n\hat{\sigma^2}}{\sigma^2}, \quad Z_1 = \sqrt{n}\frac{\hat{a} - a}{\sigma}, \quad Z_2 = \sqrt{n}\frac{s_1(\hat{b} - b)}{\sigma}$$

незалежні, причому величина χ^2_{n-2} має χ^2-розподіл з $n-2$ ступенями свободи, а величини Z_1 та Z_2 мають нормальний розподіл, як лінійні комбінації незалежних нормальних величин.

Тепер, спираючись на властивості χ^2-розподілу (див. 2.3), знайдемо математичне сподівання і дисперсію оцінки $\hat{\sigma^2}$:

$$\mathrm{M}\hat{\sigma^2} = \frac{\sigma^2 \mathrm{M}\chi^2_{n-2}}{n} = \frac{(n-2)\sigma^2}{n},$$
$$\mathrm{D}\hat{\sigma^2} = \frac{\sigma^4 \mathrm{D}\chi^2_{n-2}}{n^2} = \frac{2(n-2)\sigma^4}{n^2}.$$

З отриманих формул випливає, що оцінка $\hat{\sigma^2}$ є консистентною, але зміщеною, і тому замість неї для оцінювання невідомого параметра σ^2 використовують її незміщений варіант

$$\tilde{\sigma^2} = \frac{n}{n-2}\hat{\sigma^2}. \tag{4.4}$$

Розглянемо випадкові величини Z_1, Z_2. З наведених вище формул для математичного сподівання і дисперсії оцінок $\hat{a}$ та $\hat{b}$ випливає, що Z_1 та Z_2 є стандартними нормальними величинами. Тоді величини

$$T_1 = \frac{Z_1}{\sqrt{\chi^2_{n-2}/(n-2)}} = \sqrt{n}\frac{\hat{a}-a}{\tilde{\sigma}},$$
$$T_2 = \frac{Z_2}{\sqrt{\chi^2_{n-2}/(n-2)}} = s_1\sqrt{n}\frac{\hat{b}-b}{\tilde{\sigma}}$$

за означенням мають розподіл Стьюдента з $n-2$ ступенями свободи і, отже,

$$\mathrm{P}\left\{|T_i| < t_{1-\frac{\alpha}{2}}\right\} = 1-\alpha, \quad i = 1, 2.$$

Останнє співвідношення дозволяє як перевірити гіпотезу щодо значень коефіцієнтів регресії a, b, так і побудувати надійні інтервали для них.

Отримані результати сформулюємо у вигляді теореми:

Теорема (про властивості оцінок методу найменших квадратів). Оцінки $\hat{a}, \hat{b}, \tilde{\sigma^2}$ з (4.3),(4.4) параметрів a, b, σ^2 є незміщеними, конзистентними та незалежними між собою. До того ж, надійнійні інтервали для коефіцієнтів a, b з рівнем надійності $1-\alpha$, мають вигляд

$$\hat{a} - \frac{\tilde{\sigma}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2) < a < \hat{a} + \frac{\tilde{\sigma}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2),$$

$$\hat{b} - \frac{\tilde{\sigma}}{s_1\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2) < b < \hat{b} + \frac{\tilde{\sigma}}{s_1\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-2),$$

де $t_{1-\frac{\alpha}{2}}(n-2)$ — квантиль рівня $1-\alpha$ розподілу Стьюдента з $n-2$ ступенями свободи.

Вправа. Побудувати надійний інтервал для невідомої дисперсії σ^2.

Література

[1] *Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.* – М.: Наука, 1983.

[2] *Боровков А.А. Математическая статистика. Оценка параметров, проверка гипотез.* – М.: Наука, 1984.

[3] *Гихман И.И., Скороход А.В., Ядренко М.Й. Теория вероятностей и математическая статистика.* – К.: Вища школа, 1979.

[4] *Гмурман В.Е. Теория вероятностей и математическая статистика.* – М.: Высшая школа, 2003.

[5] *Горяинов В.Б., Павлов И.В., Цветкова Г.М. и др. Математическая статистика.* – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001.

[6] *Ивченко Г.И., Медведев Ю.И. Математическая статистика.* – М.: Высшая школа, 1992.

[7] *Карташов М.В. Імовірність, процеси, статистика.* – К.: Видавничо-поліграфічний центр «Київський університет», 2008.

[8] *Крамер Г. Математические методы статистики.* – М.: Мир, 1975.

[9] *Леман Э. Проверка статистических гипотез.* – М.: Наука, 1979.

[10] *Леман Э. Теория точечного оценивания.* – М.: Наука, 1991.

[11] *Турчин В.М. Математична статистика.* – К.: Академія, 1999.

Зміст