Ефективність оцінок

Лекція № 4

4.1. Порівняння оцінок, поняття ефективності
4.2. Нерівність Крамера-Рао

4.1. Порівняння оцінок, поняття ефективності

Нехай $\xi_1, \ldots, \xi_n$ є вибіркою об'єма n із параметричної сім'ї розподілів $\mathcal{F}_\theta$, $\theta \in \Theta$.

Означення 1

Кажуть, що оцінка θ_1^* є кращою за оцінку θ_2^* в середньоквадратичному, якщо для довільного $\theta \in \Theta$

$$E\left(\theta_1^* - \theta\right)^2 \leq E\left(\theta_2^* - \theta\right)^2,$$

причому хоча б при одному значенні θ нерівність є строга.

4.1. Порівняння оцінок, поняття ефективності

Нехай генеральна сукупність має рівномірний розподіл $\xi_j \sim U(0, \theta)$.
Порівняємо ОММ $\theta^* = 2\overline{\xi}$ та ОМВ $\theta^{**} = \xi_{(n)}$.
Знайдемо середньоквадратичні відхилення цих оцінок від θ
Враховуючи, що θ^* є незміщеною оцінкою

$$E\left(\theta^* - \theta\right)^2 = D\theta^* = D\left(2\overline{\xi}\right) = 4D\left(\overline{\xi}\right) = \frac{4}{n}D\xi_1 = \frac{\theta^2}{3n}.$$

Оскільки

$$E\xi_{(n)} = \frac{n}{n+1}\theta, \quad E\xi_{(n)}^2 = \frac{n}{n+2}\theta^2,$$

то

$$E\left(\theta^{**} - \theta\right)^2 = E\left(\theta^{**}\right)^2 - 2\theta E\theta^{**} + \theta^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$

Помітимо, що при $n = 1, 2$ – відхилення будуть рівними, але при $n > 2$

$$E\left(\theta^{**} - \theta\right)^2 = \frac{2\theta^2}{(n+1)(n+2)} < \frac{\theta^2}{3n} = E\left(\theta^{*} - \theta\right)^2,$$

тобто оцінка

$$\theta^{**} = \xi_{(n)}$$

є кращою за

$$\theta^{*} = 2\overline{\xi}$$

в середньо квадратичному при $n > 2$.

Виникає питання:

Чи існує найкраща оцінка в середньоквадратичному?

Нехай маємо невироджену задачу, тобто для жодної з оцінок θ^* неможливо, щоб

$$\theta^* = \theta \text{ м.н. на } \Theta.$$

Теорема 1

У класі всеможливих оцінок найкращої оцінки в середньоквадратичному не існує.

4.1. Порівняння оцінок, поняття ефективності

Доведемо від супротивного. Нехай $\widehat{\theta}$ є найкращою оцінкою. Це означає, що для довільної оцінки θ^* та для довільних $\theta \in \Theta$

$$E\left(\widehat{\theta} - \theta\right)^2 \leq E\left(\theta^* - \theta\right)^2. \tag{1}$$

Зафіксуємо $\theta_1 \in \Theta$ і розглянемо оцінку $\theta_1^* = \theta_1$. Тоді

$$E\left(\widehat{\theta} - \theta_1\right)^2 \leq E\left(\theta_1^* - \theta_1\right)^2 = E\left(\theta_1 - \theta_1\right)^2 = 0.$$

В силу довільності вибору θ_1, будемо мати, що для всіх $\theta \in \Theta$

$$E\left(\widehat{\theta} - \theta\right)^2 = 0.$$

Останнє означає, що $\widehat{\theta} = \theta$ м.н. на Θ, що суперечить невиродженості задачі.

4.1. Порівняння оцінок, поняття ефективності

Доведена теорема означає, що ми не можемо шукати найкращу оцінку серед класу всеможливих оцінок і, відповідно, треба розглядати певні підкласи в кожному з яких вже шукати найкращу оцінку.

Часто розглядають оцінки, які мають однакове зміщення, у сенсі, що

$$b(\theta) = E\theta^{*} - \theta$$

для деякої заданої функції $b(\theta)$.

4.1. Порівняння оцінок, поняття ефективності

Позначимо через $K_b = K_{b(\theta)}$ – клас всіх оцінок зі зміщенням рівним заданій функції $b(\theta)$:

$$K_b = \{\theta^* \mid E\theta^* = \theta + b(\theta)\};$$

$K_0 = \{\theta^* \mid E\theta^* = \theta\}$ – клас всіх незміщених оцінок.

Означення 2

Оцінка $\widetilde{\theta} \in K_b$ називається ефективною оцінкою в класі K_b, якщо вона не гірша за всіх інших оцінок класу K_b в сенсі середньоквадратичного підходу, тобто для довільних $\theta^* \in K_b$ та $\theta \in \Theta$

$$E\left(\widetilde{\theta} - \theta\right)^2 \le E\left(\theta^* - \theta\right)^2 .$$

4.1. Порівняння оцінок, поняття ефективності

Зауваження 1

Помітимо, що для $\theta^* \in K_b$

$$E(\theta^* - \theta)^2 = E(\theta^* - E\theta^*)^2 + (E\theta^* - \theta)^2 = D\theta^* + b(\theta).$$

Таким чином, порівняння оцінок з однаковим зміщенням також зводиться до порівняння дисперсій.

Теорема 2

Якщо $\widetilde{\theta}_1 \in K_b$ і $\widetilde{\theta}_2 \in K_b$ – дві ефективні оцінки в K_b, тоді $\widetilde{\theta}_1 = \widetilde{\theta}_2$ м.н.

Нерівність Крамера-Рао

Розглянемо питання існування ефективної оцінки.
Нехай $\mathcal{F}_\theta$, $\theta \in \Theta$, де Θ – обмежений інтервал числової осі, є абсолютно неперервним розподілом, а

$$f(\vec{x};\theta) = f(x_1, \ldots, x_n;\theta),$$

$\theta \in \Theta$ – функція вірогідності реалізації вибірки

$$\vec{\xi} = (\xi_1, \ldots, \xi_n) \quad \text{з} \quad \mathcal{F}_\theta.$$

Помітимо, що

$$f(\vec{x};\theta) = f(x_1, \ldots, x_n;\theta)$$

буде також і щільністю розподілу $\vec{\xi}$.

Лема 1

Нехай для майже всіх $\vec{x} \in \mathbb{R}^n$ існують похідні

$$\frac{\partial}{\partial\theta} f(\vec{x};\theta), \;\; \frac{\partial^2}{\partial\theta^2} f(\vec{x};\theta),$$

для яких виконуються умови

(i) $\left|\frac{\partial^i}{\partial\theta^i} f(\vec{x};\theta)\right| \le u_i(\vec{x})$, $i = 1, 2$, де функції u_i такі, що

$$\int_{\mathbb{R}^n} u_i(\vec{x}) d\vec{x} < \infty, \;\; i = 1, 2;$$

(ii) $E\left(\frac{\partial}{\partial\theta} \ln f(\vec{\xi};\theta)\right)^2 < \infty$, $E\left|\frac{\partial^2}{\partial\theta^2} \ln f(\vec{\xi};\theta)\right| < \infty$.

Тоді

1) $E\frac{\partial}{\partial\theta} \ln f(\vec{\xi};\theta) = 0$;

2) $D\frac{\partial}{\partial\theta} \ln f(\vec{\xi};\theta) = E\left(\frac{\partial}{\partial\theta} \ln f(\vec{\xi};\theta)\right)^2 = -E\frac{\partial^2}{\partial\theta^2} \ln f(\vec{\xi};\theta)$.

Доведення.

На основі вказаних вище характеристик можна побудувати статистичний розподіл г.с.

1) Беручи до уваги те, що $\int_{\mathbb{R}^n} f(\vec{x};\theta)d\vec{x} = 1$ та властивості диференціювання інтегралу за параметром, будемо мати

$$E\frac{\partial}{\partial\theta}\ln f(\vec{\xi};\theta) = \int_{\mathbb{R}^n} \frac{\frac{\partial}{\partial\theta} f(\vec{x};\theta)}{f(\vec{x};\theta)} f(\vec{x};\theta)d\vec{x} = \\ = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(\vec{x};\theta)d\vec{x} = \frac{\partial}{\partial\theta}\left(\int_{\mathbb{R}^n} f(\vec{x};\theta)d\vec{x}\right) = 0. \quad (2)$$

Крім того, з умов Леми випливає, що останню рівність можна продиференціювати, звідки ми отримуємо наступну рівність

$$\int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} f(\vec{x};\theta)d\vec{x} = 0.$$

2) З одного боку,

$$E\left(\frac{\partial}{\partial\theta}\ln f(\vec{\xi};\theta)\right)^2 = \int_{\mathbb{R}^n}\left(\frac{\partial}{\partial\theta}\ln f(\vec{x};\theta)\right)^2 f(\vec{x};\theta)d\vec{x} = \int_{\mathbb{R}^n}\frac{\left(\frac{\partial}{\partial\theta}f(\vec{x};\theta)\right)^2}{f(\vec{x};\theta)}d\vec{x}.$$

З іншого боку,

$$E\frac{\partial^2}{\partial\theta^2}\ln f(\vec{\xi};\theta) = E\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(\vec{\xi};\theta)}{f(\vec{\xi};\theta)}\right) = E\,\frac{\frac{\partial^2}{\partial\theta^2}f(\vec{\xi};\theta)\cdot f(\vec{\xi};\theta) - \left(\frac{\partial}{\partial\theta}f(\vec{\xi};\theta)\right)^2}{f^2(\vec{\xi};\theta)} =$$

$$= E\,\frac{\frac{\partial^2}{\partial\theta^2}f(\vec{\xi};\theta)}{f(\vec{\xi};\theta)} - E\,\frac{\left(\frac{\partial}{\partial\theta}f(\vec{\xi};\theta)\right)^2}{f^2(\vec{\xi};\theta)} = \int_{\mathbb{R}^n}\frac{\partial^2}{\partial\theta^2}f(\vec{x};\theta)d\vec{x} - \int_{\mathbb{R}^n}\frac{\left(\frac{\partial}{\partial\theta}f(\vec{x};\theta)\right)^2}{f(\vec{x};\theta)}d\vec{x} =$$

$$= -\int_{\mathbb{R}^n}\frac{\left(\frac{\partial}{\partial\theta}f(\vec{x};\theta)\right)^2}{f(\vec{x};\theta)}d\vec{x},$$

і Лему доведено.

Означення 3

$$I_n(\theta) = E\left(\frac{\partial}{\partial\theta}\ln f(\vec{\xi};\theta)\right)^2 = -E\frac{\partial^2}{\partial\theta^2}\ln f(\vec{\xi};\theta)$$

називається кількістю інформації за Фішером.

Теорема 3

Якщо виконано умови Леми 1 та для незміщеної оцінки $\widehat{\theta}$
(iii) існує така невід'ємна функція U, інтегрована у сенсі $\int_{\mathbb{R}^n} U(\vec{x})d\vec{x} < \infty$, що для довільного $\vec{x} \in \mathbb{R}^n$

$$\left|\widehat{\theta}(\vec{x}) \frac{\partial}{\partial\theta} f(\vec{x}, \theta)\right| \leq U(\vec{x}), \tag{3}$$

тоді

$$D\widehat{\theta} \geq \frac{1}{I_n(\theta)}, \tag{4}$$

причому рівність досягається тоді і тільки тоді, коли для деякого $C = C_{n,\theta}$

$$\frac{\partial}{\partial\theta} \ln f(\vec{x}, \theta) = C_{n,\theta}\left(\widehat{\theta}(\vec{x}) - \theta\right). \tag{5}$$

Доведення.
Нехай $\widehat{\theta} \in K_0$ – деяка незміщена оцінка, що задовольняє (3). Тоді, враховуючи, що $\widehat{\theta} = \widehat{\theta}(\xi_1, \ldots, \xi_n) = \widehat{\theta}\left(\vec{\xi}\right)$,

$$\theta = E\widehat{\theta} = \int_{\mathbb{R}^n} \widehat{\theta}(\vec{x})\, f(\vec{x}, \theta)\, d\vec{x}.$$

Продиференціюємо ліву та праву частини останньої рівності за θ. Тоді, з (2) та нерівності Коші-Буняковського,

$$1 = \int_{\mathbb{R}^n} \widehat{\theta}(\vec{x}) \frac{\partial}{\partial\theta} f(\vec{x}, \theta)\, d\vec{x} = \int_{\mathbb{R}^n} \left(\widehat{\theta}(\vec{x}) - \theta\right) \frac{\partial}{\partial\theta} f(\vec{x}, \theta)\, d\vec{x} + \int_{\mathbb{R}^n} \theta \frac{\partial}{\partial\theta} f(\vec{x}, \theta)\, d\vec{x} =$$

$$= \int_{\mathbb{R}^n} \left[\left(\widehat{\theta}(\vec{x}) - \theta\right) \sqrt{f(\vec{x}, \theta)}\right] \cdot \frac{\frac{\partial}{\partial\theta} f(\vec{x}, \theta)}{\sqrt{f(\vec{x}, \theta)}} d\vec{x} \le$$

$$\le \sqrt{\int_{\mathbb{R}^n} \left(\widehat{\theta}(\vec{x}) - \theta\right)^2 f(\vec{x}, \theta)\, d\vec{x}} \cdot \sqrt{\int_{\mathbb{R}^n} \frac{\left(\frac{\partial}{\partial\theta} f(\vec{x}, \theta)\right)^2}{f(\vec{x}, \theta)} d\vec{x}},$$

або, враховуючи, що

$$D\widehat{\theta} = \int_{\mathbb{R}^n} \left(\widehat{\theta}(\vec{x}) - \theta\right)^2 f(\vec{x}, \theta)\, d\vec{x}$$

та

$$I_n(\theta) = \int_{\mathbb{R}^n} \frac{\left(\frac{\partial}{\partial\theta} f(\vec{x}, \theta)\right)^2}{f(\vec{x}, \theta)} d\vec{x}$$

(див. доведення Леми 1) отримуємо (4).
Оскільки знак рівності в (4) еквівалентно знаку рівності у нерівності Коші-Буняковського, отримуємо, що

$$\left(\widehat{\theta}(\vec{x}) - \theta\right) \sqrt{f(\vec{x}, \theta)}$$

та

$$\frac{\frac{\partial}{\partial\theta} f(\vec{x}, \theta)}{\sqrt{f(\vec{x}, \theta)}}$$

мають бути лінійно залежними або (5) і Теорему доведено.

Наслідок 1

Якщо $D\widehat{\theta} = \frac{1}{I_n(\theta)}$ або виконується рівність (5), тоді $\widehat{\theta}$ є ефективною оцінкою параметра θ.

Наслідок 2

Якщо $\xi_1, \ldots, \xi_n$ є н.о.р.в.в. із щільністю $g(x, \theta)$, тоді

$$I_n(\theta) = nI(\theta),$$

де $I(\theta) = E\left(\frac{\partial}{\partial\theta} \ln g(\xi_1; \theta)\right)^2 = -E\frac{\partial^2}{\partial\theta^2} \ln g(\xi_1; \theta)$.

Доведення випливає беспосередньо з того, що функція вірогідності в цьому випадку має вигляд

$$f(x_1, \ldots, x_n; \theta) = \prod_{j=1}^{n} g(x_j; \theta).$$

4.2. Нерівність Крамера-Рао

Зауваження 2

Якщо $\xi_1, \ldots, \xi_n$ є вибіркою із дискретного розподілу, то всі попередні міркування залишаються вірними, якщо щільність розподілу замінити ймовірностями, а замість інтегрування за мірою Лебега розглядати підсумовування (інтегрування за рахуючою мірою).

Зауваження 3

Нерівність Крамера-Рао можна перенести на багатовимірний випадок:

$$E\left(\widehat{\theta}-\theta\right)\left(\widehat{\theta}-\theta\right)^T \geq I_n^{-1}(\theta),$$

де кількість інформації за Фішером

$$I_n(\theta) = -\left(E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f(\vec{\xi};\theta)\right)_{i,j=1}^q$$

Дякуємо за увагу!