Теорія ймовірностей, статистика та ймовірнісні процеси

Лекція 12.

Розора Ірина Василівна

Київ, 2020

Зміст

Нехай $X : \Omega \to \mathbf{R}^n$ – n-кратна вибірка, тобто $X = (\xi_1, \xi_2, \ldots, \xi_n)$ і ξ_k — незалежні однаково розподілені в.в. (н.о.р.в.в.) із щільністю $f(x, \theta)$, де параметр $\theta \in \Theta$.

Зауваження

Якщо елементи вибірки ξ_k дискретно розподілені, то будемо розглядати їх точкові ймовірності $P(x, \theta)$.
Надалі, всі твердження будемо записувати саме для неперервної вибірки.

Функція вірогідності

Функцією вірогідності (або функцією правдоподібності) кратної вибірки називається сумісна щільність розподілу (ссумісна ймовірність для д.в.в.) вибірки:

$$L(x, \theta) \equiv f(x, \theta) = \prod_{k=1}^{n} f(x_k, \theta), \quad x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n.$$

Вибіркова функція вірогідності

Вибірковою функцією вірогідності (емпіричною функцією вірогідності) називається в.в., що отримується в результаті підстановки у функцію вірогідності замість аргумента $x \in \mathbf{R}^n$ значення вибірки як випадкового вектора

$$L(X,\theta) \equiv L(x,\theta)|_{x=X}.$$

Функція вірогідності спостереження

Функцією вірогідності спостереження для кратної вибірки називається щільність розподілу спостереження ξ_1

$$f(y, \theta),$$

тобто така вимірна функція f, що

$$\mathbf{P}_\theta(\xi_k \in B) = \int_B f(y, \theta)\,\mathrm{d}y, \quad \forall B \in \mathcal{B}, \forall \theta \in \Theta.$$

Зміст

Функція впливу

У даному розділі припускатимемо, що параметричний простір евклідів: $\theta \subset \mathbf{R}^d$, а функція вірогідності $L(x, \theta)$ – диференційовна за θ.

Функція впливу

Функцією впливу, або функцією внеску, вибірки X називається частинна похідна за параметром θ від логарифма вибіркової функції вірогідності :

$$U(X,\theta) \equiv \frac{\partial}{\partial\theta} \ln L(X,\theta).$$

У випадку кратної вибірки функцією впливу спостереження ξ називається похідна за θ від логарифма функції вірогідності спостереження:

$$u(\xi,\theta) \equiv \frac{\partial}{\partial\theta} \ln f(\xi,\theta).$$

Теорема (про функцію впливу кратної вибірки)

Для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ вибіркова функція впливу дорівнює сумі функцій впливу спостережень, які її утворюють:

$$U(X, \theta) = \sum_{k=1}^{n} u(\xi_k, \theta).$$

Доведення.

Доведення випливає з означення функції вірогідності кратної вибірки та лінійності частинної похідної:

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)=\frac{\partial}{\partial\theta}\prod_{k=1}^{n}f(\xi_k,\theta)$$

$$=\frac{\partial}{\partial\theta}\sum_{k=1}^{n}\ln f(\xi_k,\theta)=\sum_{k=1}^{n}\frac{\partial}{\partial\theta}\ln f(\xi_k,\theta)$$

$$=\sum_{k=1}^{n}u(\xi_k,\theta)$$

□

Приклад. Пуассонівська вибірка

Для кратної вибірки з розподілом Пуассона $\Pi(\lambda)$ та невідомим параметром $\theta = \lambda$

$$u(y,\theta) = \frac{\partial}{\partial\theta} \ln\left(\exp(-\theta)\frac{\theta^y}{y!}\right) = y/\theta - 1, \quad y \in \mathbb{Z}_+,$$

$$U(X,\theta) = \sum_{k=1}^{n}(\xi_k/\theta - 1) = n(\overline{X}/\theta - 1),$$

$\overline{X} = \frac{1}{n}\sum_{k=1}^{n}\xi_k$—вибіркове середнє.

Приклад. Вибірка з гама-розподілу

Для кратної вибірки з гама-розподілом $\Gamma(\lambda, \alpha)$ та невідомим параметром $\theta = (\lambda, \alpha)$

$$\ln f(y, \theta) = \alpha \ln \lambda + (\alpha - 1) \ln y - \lambda y - \ln \Gamma(\alpha),$$

$$u_1(y, \theta) = \alpha/\lambda - y,$$

$$u_2(y, \theta) = \ln \lambda + \ln y - \psi(\alpha), \quad \psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha),$$

$$U_1(X, \theta) = n(\alpha/\lambda - \hat{\mu}_n),$$

$$U_2(X, \theta) = n \ln \lambda + \ln \prod_{k=1}^{n} \xi_k - n\psi(\alpha).$$

Приклад. Вибірка з гама-розподілу

Для кратної вибірки з гама-розподілом $\Gamma(\lambda, \alpha)$ та невідомим параметром $\theta = (\lambda, \alpha)$

$$\ln f(y, \theta) = \alpha \ln \lambda + (\alpha - 1) \ln y - \lambda y - \ln \Gamma(\alpha),$$

$$u_1(y, \theta) = \alpha/\lambda - y,$$

$$u_2(y, \theta) = \ln \lambda + \ln y - \psi(\alpha), \quad \psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha),$$

$$U_1(X, \theta) = n(\alpha/\lambda - \hat{\mu}_n),$$

$$U_2(X, \theta) = n \ln \lambda + \ln \prod_{k=1}^{n} \xi_k - n\psi(\alpha).$$

Приклад. Вибірка з гама-розподілу

Для кратної вибірки з гама-розподілом $\Gamma(\lambda, \alpha)$ та невідомим параметром $\theta = (\lambda, \alpha)$

$$\ln f(y, \theta) = \alpha \ln \lambda + (\alpha - 1) \ln y - \lambda y - \ln \Gamma(\alpha),$$

$$u_1(y, \theta) = \alpha/\lambda - y,$$

$$u_2(y, \theta) = \ln \lambda + \ln y - \psi(\alpha), \quad \psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha),$$

$$U_1(X, \theta) = n(\alpha/\lambda - \hat{\mu}_n),$$

$$U_2(X, \theta) = n \ln \lambda + \ln \prod_{k=1}^{n} \xi_k - n\psi(\alpha).$$

Умови регулярності

1. Множина тих значень вибірки X, для яких функція вірогідності $L(X, \theta)$ додатна, не залежить від θ.
2. $L(X, \theta)$ двічі неперервно диференційовна за θ.
3. $U(X, \theta)$ – ненульова та інтегровна у квадраті, тобто:

$$0 < \mathbf{M}_\theta U^2(X, \theta) < \infty, \quad \forall \theta \in \Theta.$$

4. Знак похідної за параметром θ можна внести під знак інтегралів вигляду $\int_S g(x, \theta) L(x, \theta) \lambda(\,\mathrm{d}x)$ з функцією вірогідності $L(x, \theta)$ для певних функцій g, що спричиняється умовою 1 і збіжністю інтеграла від похідної.

Зміст

Властивості функції впливу, інформація за Фішером

Теорема (про центрованість функції впливу)

За умов регулярності функція впливу центрована:

$$\mathbf{M}_\theta U(X, \theta) = 0, \quad \forall \theta \in \Theta.$$

Доведення.

Оскільки функція вірогідності $L(x,\theta)$— це сумісна щільність всієї вибірки, то

$$\int_{\mathbf{R}^n} L(x,\theta)dx = 1.$$

Візьмемо похідну від лівої і правої частин рівності за θ, міняючи диференціювання та інтегрування місцями:

$$0 = \int_{\mathbf{R}^n} \frac{\partial L(x,\theta)}{\partial\theta}dx = \int_{\mathbf{R}^n} \frac{\partial L(x,\theta)}{\partial\theta}\frac{L(x,\theta)}{L(x,\theta)}dx =$$

$$\int_{\mathbf{R}^n} \frac{\partial \ln(L(x,\theta))}{\partial\theta}L(x,\theta)dx = \mathbf{M}_\theta \ln(L(X,\theta)) = \mathbf{M}_\theta U(X,\theta)$$

$\square$

Інформація за Фішером

Нехай параметр $\theta \in \Theta \subset R$ – скалярний. Інформацією за Фішером у вибірці X називається функція

$$I(\theta) \equiv \mathbf{D}_\theta U(X,\theta) = \mathbf{M}_\theta U^2(X,\theta).$$

Друга рівність в означенні випливає з теореми про центрованість функції впливу та з властивостей дисперсії.

Інформація за Фішером

Нехай параметр $\theta \in \Theta \subset R$ – скалярний. Інформацією за Фішером у вибірці X називається функція

$$I(\theta) \equiv \mathbf{D}_\theta U(X,\theta) = \mathbf{M}_\theta U^2(X,\theta).$$

Друга рівність в означенні випливає з теореми про центрованість функції впливу та з властивостей дисперсії.

Теорема (про обчислення інформації за Фішером)

За умов регулярності справедлива тотожність

$$I(\theta) = -\mathbf{M}_\theta \frac{\partial^2}{\partial\theta^2} \ln L(X,\theta)$$

$$= -\mathbf{M}_\theta \frac{\partial}{\partial\theta} U(X,\theta), \quad \forall\theta \in \Theta.$$

Теорема (про обчислення інформації за Фішером)

За умов регулярності справедлива тотожність

$$I(\theta) = -\mathbf{M}_\theta \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta)$$

$$= -\mathbf{M}_\theta \frac{\partial}{\partial \theta} U(X, \theta), \quad \forall \theta \in \Theta.$$

Доведення.

З теореми про центрованість функції впливу $\mathbf{M}_\theta U(X,\theta) = 0$ або

$$0 = \int_{\mathbf{R}^n} U(x,\theta) L(x,\theta) dx.$$

Продиференцюємо рівність за θ:

$$0 = \int_{\mathbf{R}^n} \left(\frac{\partial U(x,\theta)}{\partial \theta} L(x,\theta) + U(x,\theta) \frac{\partial L(x,\theta)}{\partial \theta} \right) dx =$$

$$= \mathbf{M}_\theta \frac{\partial U(X,\theta)}{\partial \theta} + \int_{\mathbf{R}^n} U(x,\theta) \frac{\partial L(x,\theta)}{\partial \theta} \frac{L(x,\theta)}{L(x,\theta)} dx =$$

$$= \mathbf{M}_\theta \frac{\partial U(X,\theta)}{\partial \theta} + \int_{\mathbf{R}^n} U(x,\theta) \frac{\partial \ln(L(x,\theta))}{\partial \theta} L(x,\theta) dx =$$

$$= \mathbf{M}_\theta \frac{\partial U(X,\theta)}{\partial \theta} + \mathbf{M}_\theta U^2(X,\theta)$$

Теорема (про адитивність інформації за Фішером)

Нехай для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ функція

$$i(\theta) \equiv \mathbf{D}_\theta u(\xi_1, \theta)$$

задає інформацію за Фішером в одному спостереженні. Тоді повна інформація за Фішером дорівнює:

$$I(\theta) = ni(\theta).$$

Приклад. Схема Бернуллі

Нехай вибірка $X = (\chi_1, \ldots, \chi_n)$ містить результати випробувань Бернуллі з невідомою ймовірністю успіху θ, де χ_k – індикатор успіху в k-му випробуванні. Функція вірогідності одного спостереження є щільністю відносно точкової міри і має вигляд

$$f(y, \theta) = \theta \mathbf{I}_{y=1} + (1-\theta)\mathbf{I}_{y=0} = \theta^y (1-\theta)^{1-y}, \quad y \in \{0, 1\},$$

теоретична функція впливу дорівнює

$$u(y, \theta) = \frac{\partial}{\partial \theta} \ln f(y, \theta) = \frac{y}{\theta} - \frac{1-y}{1-\theta} = \frac{y-\theta}{\theta(1-\theta)},$$

Приклад. Схема Бернуллі

Нехай вибірка $X = (\chi_1, \ldots, \chi_n)$ містить результати випробувань Бернуллі з невідомою ймовірністю успіху θ, де χ_k – індикатор успіху в k-му випробуванні. Функція вірогідності одного спостереження є щільністю відносно точкової міри і має вигляд

$$f(y,\theta) = \theta \mathbf{I}_{y=1} + (1-\theta)\mathbf{I}_{y=0} = \theta^y (1-\theta)^{1-y}, \quad y \in \{0,1\},$$

теоретична функція впливу дорівнює

$$u(y,\theta) = \frac{\partial}{\partial\theta} \ln f(y,\theta) = \frac{y}{\theta} - \frac{1-y}{1-\theta} = \frac{y-\theta}{\theta(1-\theta)},$$

Приклад. Схема Бернуллі (продовження)

функція вірогідності та впливу всієї вибірки

$$L(X,\theta) = \prod_{k=1}^{n} \theta^{\chi_k}(1-\theta)^{1-\chi_k} = \theta^{\nu_n}(1-\theta)^{n-\nu_n}, \quad \nu_n = \sum_{k=1}^{n} \chi_k,$$

$$U(X,\theta) = \sum_{k=1}^{n} u(\chi_k,\theta) = \frac{n(\hat{\theta}_n - \theta)}{\theta(1-\theta)}, \quad \hat{\theta}_n = \frac{\nu_n}{n},$$

функції інформації за Фішером мають вигляд

$$i(\theta) = \mathbf{M}_\theta \left(\frac{\chi_1 - \theta}{\theta(1-\theta)} \right)^2 = \frac{\mathbf{D}_\theta \chi_1}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)},$$

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$

Приклад. Показниковий розподіл.

Нехай $X = (\xi_1, \ldots, \xi_n)$ кратна вибірка з показниковим розподілом: $\xi_k \cong Exp(\theta)$. Функції впливу дорівнюють:

$$u(y, \theta) = \frac{\partial}{\partial \theta}(\ln \theta - \theta y) = \frac{1}{\theta} - y,$$

$$U(X, \theta) = \sum_{k=1}^{n} u(\xi_k, \theta) = \frac{n}{\theta} - n\hat{\mu}_n,$$

$\hat{\mu}_n = \overline{X}$, а інформація за Фішером у спостереженні та у вибірці

$$i(\theta) = \mathbf{M}_\theta \left(\frac{1}{\theta} - \xi_1 \right)^2 = \mathbf{D}_\theta \xi_1 = \frac{1}{\theta^2},$$

$$I(\theta) = \frac{n}{\theta^2}$$

Приклад. Показниковий розподіл.

Нехай $X = (\xi_1, \ldots, \xi_n)$ кратна вибірка з показниковим розподілом: $\xi_k \cong Exp(\theta)$. Функції впливу дорівнюють:

$$u(y, \theta) = \frac{\partial}{\partial \theta}(\ln \theta - \theta y) = \frac{1}{\theta} - y,$$

$$U(X, \theta) = \sum_{k=1}^{n} u(\xi_k, \theta) = \frac{n}{\theta} - n\hat{\mu}_n,$$

$\hat{\mu}_n = \overline{X}$, а інформація за Фішером у спостереженні та у вибірці

$$i(\theta) = \mathbf{M}_\theta \left(\frac{1}{\theta} - \xi_1\right)^2 = \mathbf{D}_\theta \xi_1 = \frac{1}{\theta^2},$$

$$I(\theta) = \frac{n}{\theta^2}$$

Аглоритм знаходження інформації за Фішером вибірки

1. Запиати щільність (ймовірність) розподілу — функцію вірогідності $L(X,\theta)=\prod_{k=1}^{n} f(\xi_k,\theta)$;
2. $\ln(L(X,\theta))$;
3. $U(X,\theta)=\frac{\partial L(X,\theta)}{\partial\theta} \quad \Rightarrow \quad I(\theta)=D_\theta U(X,\theta)$ або
4. $\frac{\partial U(X,\theta)}{\partial\theta} \quad \Rightarrow \quad I(\theta)=-M_\theta \frac{\partial U(X,\theta)}{\partial\theta}$

Зауваження

Інформацію за Фішером можна шукати для одного спостереження, а потім потрібно використати теорему про адитивність інформації за Фішером.

Нерівність Крамера – Рао

Розглянемо задачу оцінювання значення дійсної параметричної функції $\tau(\theta)$ у класі Γ_τ незсунутих її оцінок.

Теорема (про нерівність та критерій Крамера – Рао)

Нехай параметр θ є скалярним: $\theta \in \mathbf{R}$.
(a) Якщо $T = T(X) \in \Gamma_\tau$ – довільна незсунута оцінка $\tau(\theta)$, і виконуються умови регулярності, то $\forall \theta \in \Theta$ має місце нерівність Крамера – Рао

$$\mathbf{M}_\theta (T - \tau)^2 \equiv \mathbf{D}_\theta T \geq \frac{\tau_\theta^2(\theta)}{I(\theta)},$$

де

$$\tau_\theta(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \tau(\theta),$$

$I(\theta)$ – інформація за Фішером у вибірці X.

(б) Рівність у нерівності (а) виконується тоді й тільки тоді, коли оцінка T є лінійною функцією від функції впливу вибірки:

$$T(X) - \tau(\theta) = c(\theta)U(X,\theta) \text{ м.н.}, \quad \forall\theta \in \Theta,$$

для деякої дійсної $c(\theta)$. Ця стала дорівнює

$$c(\theta) = \frac{\tau_\theta(\theta)}{I(\theta)}.$$

Зауваження

Якщо функція $\tau(\theta) = \theta$, то нерівність Крамера-Рао має вигляд

$$\mathbf{M}_\theta (T - \theta)^2 \equiv \mathbf{D}_\theta T \geq \frac{1}{I(\theta)},$$

Ефективна оцінка

Оцінка $T \in \Gamma_\tau$ називається ефективною оцінкою параметричної функції $\tau(\theta)$, якщо нерівність Крамера – Рао для неї є рівністю, тобто у випадку, коли ця оцінка має найменше можливе значення дисперсії у класі Γ_τ всіх незсунутих оцінок.

Твердження (б) дає критерій ефективності Крамера – Рао.

Ефективна оцінка

Оцінка $T \in \Gamma_\tau$ називається ефективною оцінкою параметричної функції $\tau(\theta)$, якщо нерівність Крамера – Рао для неї є рівністю, тобто у випадку, коли ця оцінка має найменше можливе значення дисперсії у класі Γ_τ всіх незсунутих оцінок.

Твердження (6) дає критерій ефективності Крамера – Рао.

I спосіб перевірки оцінки на ефективність

1. Знайти дисперсію оцінки $D_\theta T(X)$;
2. Обчислит інформацію за Фішером $I(\theta) = D_\theta U(X, \theta)$ або $I(\theta) = -M_\theta \frac{\partial U(X,\theta)}{\partial \theta}$;
3. Якщо $D_\theta T(X) = \frac{1}{I(\theta)}$, то оцінка $T(X)$ є ефективною.

II спосіб перевірки оцінки на ефективність

Використати критерій ефектривності:

1. Запиати щільність (ймовірність) розподілу — функцію вірогідності $L(X,\theta) = \prod_{k=1}^{n} f(\xi_k,\theta)$;
2. $\ln(L(X,\theta))$;
3. $U(X,\theta) = \frac{\partial L(X,\theta)}{\partial \theta}$;
4. Якщо можна подати $U(X,\theta) = k(\theta)(T(X) - \theta)$, де $k(\theta)$— деяка функція, що не залежить від вибірки,то оцінка $T(X)$ є ефективною для параметра θ.

Ефективна оцінка у схемі Бернуллі

Розглянемо задачу оцінювання, в якій проводяться n випробувань Бернуллі з невідомою ймовірністю $p = \theta$ успіху в окремому випробуванні. Припустимо, що спостерігається вибірка $X = (\chi_1, \ldots, \chi_n)$, де χ_k – індикатор k-го успіху. Логарифмічна функція вірогідності має вигляд

$$\ln L(X, \theta) = \nu_n(X) \ln \theta + (n - \nu_n(X)) \ln(1 - \theta),$$

де $\nu_n(X) = \sum_{k=1}^{n} \chi_k$ – загальна кількість успіхів. Звідси

$$U(X, \theta) = \frac{\nu_n(X)}{\theta} - \frac{n - \nu_n(X)}{1 - \theta} = \frac{\nu_n(X)}{\theta(1 - \theta)} - \frac{n}{1 - \theta},$$

функція інформації за Фішером дорівнює

$$I(\theta) = \frac{\mathbf{D}_\theta \nu_n(X)}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.$$

Розглянемо оцінку для $\theta = p$ як вибіркове середнє:

$$T(X) = \hat{\theta} = \frac{\nu_n(X)}{n} = \frac{1}{n}\sum_{k=1}^{n} \chi_k.$$

Знайдемо дисперсію цієї оцінки

$$D(\hat{\theta}) = D(\frac{1}{n}\sum_{k=1}^{n}\chi_k) = \frac{1}{n^2}\sum^{n} D\chi_k =$$

$$= \frac{1}{n^2}\sum^{n}\theta(1-\theta) = \frac{\theta(1-\theta)}{n}.$$

Оскільки $D(\hat{\theta}) = \frac{1}{I(\theta)}$, то оцінка $\hat{\theta}$ ефективна для параметра θ.

Приклад

Розглянемо кратну вибірку $X = (\xi_1, \ldots, \xi_n)$ з показниковим розподілом, $\xi_k \sim Exp(\frac{1}{\theta})$. Покажемо, що $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\xi_k$ є ефективною оцінкою для θ.
Використаємо другий спосіб (критерій ефективності Крамера-Рао).
Оскільки $f(x, \theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x > 0$, то

1. функція вірогідності
$$L(X, \theta) = \prod_{k=1}^{n} f(\xi_k, \theta) = \frac{1}{\theta^n}e^{-\sum_{k=1}^{n}\frac{\xi_k}{\theta}};$$
2. $$\ln(L(X, \theta)) = -n\ln(\theta) - \sum_{k=1}^{n}\frac{\xi_k}{\theta};$$

1. Знаходимо функцію впливу

$$U(X,\theta) = \frac{\partial \ln(L(X,\theta))}{\partial \theta} = -\frac{n}{\theta} + \sum_{k=1}^{n} \frac{\xi_k}{\theta^2} =$$

$$= \frac{n}{\theta^2}\left(-\theta + \frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{n}{\theta^2}\left(\hat{\theta} - \theta\right).$$

Отже, $\hat{\theta}$є ефективною оцінкою параметра θ.

Приклад

Розглянемо кратну вибірку $X = (\xi_1, \ldots, \xi_n)$ з нормальним розподілом, $\xi_k \sim N(0, \sigma^2)$. Дослідити оцінку $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \xi_k^2$ параметра σ^2 на ефективність двома способами.

Приклад

Розглянемо кратну вибірку $X = (\xi_1, \ldots, \xi_n)$ із розподілом Релея, тобто ξ_k мають щільність

$$f(x, \theta) = \frac{x}{\theta} \exp\{-\frac{x^2}{2\theta}\}, \quad x > 0.$$

Чи є оцінка $\hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} \xi_k^2$ ефективною параметра θ?

ПИТАННЯ?