Статистичні методи теорії ризику

© Ямненко Р.Є

КНУ імені Тараса Шевченка
механіко-математичний факультет
магістратура "Актуарна та фінансова математика"

I семестр 2012

Зміст

Статистичне оцінювання параметрів розподілу

На практиці точний розподіл величини позовів (чи їх кількості) рідко буває відомим.

Припустимо, що розподіл позовів належить певній параметричній сім'ї розподілів

$$F_{\boldsymbol{\theta}}(x),\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d,\ d \geq 1,$$

але значення параметрів $\theta_1, \ldots, \theta_d$ невідомі. Тоді їх треба оцінити за відомими величинами позовів, використовуючи прийнятний статистичний метод.

Згадаємо коротко такі основні методи оцінювання, як методи моментів, максимальної вірогідності та квантилів.

Точкові оцінки

Вибірка

Випадковий вектор $\mathbf{X} = (X_1, \ldots, X_n)$ зі значеннями в просторі $\mathbb{R}^n$ називають *вибіркою*. Вибірку, утворену послідовністю н.о.р. в.в. $X_1, \ldots, X_n$, кожна з яких має розподіл F, називають *повторною* або *n-кратною вибіркою з розподілу F*

Точно чи наближено визначити значення $\boldsymbol{\theta}$ за спостереженнями $\mathbf{x} = (x_1, \ldots, x_n)$ вибірки $\mathbf{X}$ означає, що реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ треба поставити у відповідність значення $\boldsymbol{\theta}$, тобто задати функцію $T(\cdot)$ зі значеннями в Θ – множині можливих значень параметра $\boldsymbol{\theta}$ – таку, що $T(\mathbf{x})$ точно чи хоча б наближено дорівнює $\boldsymbol{\theta}$. Значення $\hat{\boldsymbol{\theta}} = T(\mathbf{x})$ надалі й будемо використовувати як $\boldsymbol{\theta}$. Треба зазначити, що для кожної реалізації $\mathbf{x}$ вибірки $\mathbf{X}$ значення $\hat{\boldsymbol{\theta}}$ буде своє, а тому є в.в. як функція від вибірки $\mathbf{X}$, тобто $\hat{\boldsymbol{\theta}} = T(\mathbf{X})$.

Оцінка

Борелеву функцію $T\colon \mathbb{R}^n \mapsto \Theta$ називають *статистикою*, а $T(\mathbf{X}) = T(X_1, \dots, X_n)$ – борелеву функцію від вибірки зі значеннями в Θ – *оцінкою*.

Для того самого $\boldsymbol{\theta}$ можна запропонувати багато оцінок.

Питання

Наскільки великою є похибка $|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|$ від заміни справжнього значення параметра $\boldsymbol{\theta}$ на оцінене $\hat{\boldsymbol{\theta}}$?

Кількісно міру похибки від заміни $\boldsymbol{\theta}$ на $\hat{\boldsymbol{\theta}}$ (міру розсіювання $\hat{\boldsymbol{\theta}}$ відносно $\boldsymbol{\theta}$) описують за допомогою величини

$$\mathsf{E}_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|^2,$$

де $\mathsf{E}_{\boldsymbol{\theta}}X$ означає математичне сподівання в.в. X відносно параметричної сім'ї розподілів $F_{\boldsymbol{\theta}}$ та є функцією $\boldsymbol{\theta}$:

$$\mathsf{E}_{\boldsymbol{\theta}}X = \int_{\mathbb{R}} x \mathrm{d}F_{\boldsymbol{\theta}}(x).$$

Серед усіх оцінок з однією й тією самою дисперсією $\mathsf{D}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}$ мінімальну міру розсіювання відносно θ мають оцінки, для яких $\mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} = \theta$. Це випливає з перетворень

$$\mathsf{E}_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|^2 = \mathsf{E}_{\boldsymbol{\theta}}|(\boldsymbol{\theta} - \mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}) + (\mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})|^2$$

$$= \mathsf{E}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} - \mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})^2 + 2\mathsf{E}_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})(\mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \mathsf{E}_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})^2$$

$$= \mathsf{D}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} + \mathsf{E}_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})^2.$$

Незміщеність

Оцінку $\hat{\boldsymbol{\theta}}$ називають *незміщеною оцінкою параметра* $\boldsymbol{\theta}$, якщо

$$\mathsf{E}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}.$$

Часто можна розглядати не одну оцінку $\hat{\boldsymbol{\theta}} = T(\mathbf{X})$, побудовану за вибіркою $\mathbf{X} = (X_1, \dots, X_n)$, а послідовність оцінок $\hat{\boldsymbol{\theta}}_n = T_n(\mathbf{X})$, $n = 1, 2, \dots$ У цьому разі природно говорити про асимптотичну поведінку послідовності оцінок $\hat{\boldsymbol{\theta}}_n$.

Консистентність

Послідовність оцінок $\hat{\boldsymbol{\theta}}_n$ називають *консистентною послідовністю оцінок параметра* $\boldsymbol{\theta}$, якщо для довільного $\varepsilon > 0$

$$\mathrm{P}(|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}| > \varepsilon) \to 0, \quad n \to \infty.$$

Асимптотична незміщеність

Послідовність оцінок $\hat{\boldsymbol{\theta}}_n$ називають *асимптотично незміщеною послідовністю оцінок параметра* $\boldsymbol{\theta}$, якщо для довільного $\varepsilon > 0$

$$\mathrm{E}_{\boldsymbol{\theta}} \hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}, \quad n \to \infty.$$

Приклад

Нехай $\mathbf{X} = (X_1, \ldots, X_n)$ – повторна вибірка з рівномірного розподілу, заданого на відрізку $[a, b]$, де значення параметрів a та b невідомі. Які з оцінок

$$\hat{\theta}_1 = \max\{X_1, \ldots, X_n\}, \quad \hat{\theta}_2 = \min\{X_1, \ldots, X_n\},$$

$$\hat{\theta}_3 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\theta}_4 = \frac{X_{n-1} + X_n}{2}$$

та яких параметрів є

- i) незміщеними,
- ii) консистентними оцінками?

Метод моментів

Першим загальним методом побудови оцінок невідомих параметрів за вибіркою був метод моментів, запропонований Карлом Пірсоном. Згідно з цим методом певну кількість вибіркових моментів

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{1}$$

прирівнюють до відповідних теоретичних моментів

$$m_k(\boldsymbol{\theta}) = m_k(\theta_1, \ldots, \theta_d) = \int_{\mathbb{R}} x^k \mathrm{d}F_{\boldsymbol{\theta}}(x) \tag{2}$$

розподілу $F_{\boldsymbol{\theta}}(x)$, обчислених за значеннями параметрів $\theta_1, \ldots, \theta_d$, що дорівнюють відповідно $\hat{\theta}_1, \ldots, \hat{\theta}_d$.

Перший вибірковий момент

$$\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

прирівнюють до першого теоретичного моменту

$$m_1(\hat{\boldsymbol{\theta}}) = m_1(\hat{\theta}_1, \ldots, \hat{\theta}_d) = \int_{\mathbb{R}} x \mathrm{d}F_{\hat{\boldsymbol{\theta}}}(x);$$

другий вибірковий момент

$$\hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

– до другого теоретичного моменту

$$m_2(\hat{\boldsymbol{\theta}}) = m_2(\hat{\theta}_1, \ldots, \hat{\theta}_d) = \int_{\mathbb{R}} x^2 \mathrm{d}F_{\hat{\boldsymbol{\theta}}}(x)$$

і так далі.

Розглядаючи кількість моментів, що дорівнює кількості невідомих параметрів, які треба оцінити, одержують таку саму кількість рівнянь для визначення невідомих параметрів:

$$\begin{aligned} \hat{m}_1 &= m_1(\hat{\boldsymbol{\theta}}); \\ \hat{m}_2 &= m_2(\hat{\boldsymbol{\theta}}); \\ &\vdots \\ \hat{m}_d &= m_d(\hat{\boldsymbol{\theta}}). \end{aligned}$$

Розв'язуючи ці рівняння відносно $\hat{\theta}_1, \ldots, \hat{\theta}_d$, знаходять шукані оцінки.

Оцінка методу моментів є строго конзистентною.

Вправа

Повторну випадкову вибірку розміру n взяли з розподілу зі щільністю $f(x) = \frac{\alpha}{(1+x)^{\alpha+1}}$, $x > 0$, де $\alpha > 0$ – невідомий параметр.

i) Обчисліть EX, коли $\alpha > 1$.

ii) Визначте оцінку методом моментів параметра α.

Вправа

Повторну випадкову вибірку розміру n взяли з гамма-розподілу $\Gamma(\alpha, \lambda)$. Визначте оцінку методом моментів параметрів α та λ.