Точечные и интервальные оценки

Статистики и оценки

Пусть $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$, $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ — параметрическая вероятностная статистическая модель

$X = (X_1, \dots X_n)$ — выборка из неизв. распр $P \in \mathcal{P}$

Задача оценки параметра θ

Опр. Пусть $(E, \mathcal{E})$ — измеримое пр-во

Тогда измеримое отображение $S: \mathcal{X} \to E$ наз-ся статистикой

Если $E = \Theta$, то будем называть $S(X)$ оценкой

Замечание Можно оценивать и $\tau(\theta)$

Примеры: 1) $\overline{X}, \overline{X^2}, \overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ — выборочный k-й момент

Пусть g — борел ф-ция $\overline{g(x)} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$ — выборочная хар-ка ф-ции g

2) $S^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ выборочная дисперсия

3) $X_{(k)}$ k-ая порядковая статистика

$(X_{(1)} \text{ — } X_{(n)})$ вариационный ряд

Свойства оценок

Пусть $X = (X_1 \dots X_n)$ — выборка из неизв. распр $P \in \{P_\theta \mid \theta \in \Theta \subset \mathbb{R}^d\}$

Замечание для распределения P_θ будем обозначать

E_θ — мат. ож.

D_θ — дисперсия

P_θ п.н. — сходимость п.н.

d_θ — сходимость по распр.

Опр Оценка $\hat{\theta}$ наз-ся несмещённой оценкой $\tau(\theta)$, если

$$E_\theta \hat{\theta}(X_1 \dots X_n) = E_\theta \hat{\theta}(X) = \tau(\theta) \quad \forall \theta \in \Theta$$

Смысл: при многократном повторении эксперимента получим истинное $\tau(\theta)$ (в среднем)

Пример 1) $\hat{\theta}_1 = \overline{X}$ $\hat{\theta}_2 = X_1$ – несмещ оценки для $\tau(\theta) = E_\theta X_1$

2) $\mathcal{P} = \{ Bern(\theta) \mid \theta \in (0,1) \}$

$\overline{X}$, X_1 – несмещ. оценки θ

3) $\mathcal{P} = \{ Exp(\theta) \mid \theta > 0 \}$

$\overline{X}$ – несм. оценка для $1/\theta$

// $S^2 = \overline{X^2} - \overline{X}^2$

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

$= EX_1^2 - \frac{1}{n^2}E\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j\right) =$

$= EX_1^2 - \frac{1}{n}EX_1^2 - \frac{1}{n^2}E\sum_{i\neq j} X_i X_j =$

$= EX_1^2 - \frac{1}{n}EX_1^2 - \frac{n-1}{n}(EX_1)^2 = \frac{n-1}{n}DX_1$

$\Rightarrow$ S^2 – смещённая оценка DX_1, а $\frac{n}{n-1}S^2$ – несмещ.

Асимптотические св-ва

Пусть $X = (X_1, X_2 \dots)$ выборка неогр. размера из неизв $P \in \{P_\theta \mid \theta \in \Theta \subset \mathbb{R}^d\}$

1) Оценка (посл-ть оценок) $\hat{\theta}(x_1 \dots x_n)$ наз-ся состоятельной $=$ слабо состоятельной оценкой $\tau(\theta)$, если $\hat{\theta}(X_1 \dots X_n) \xrightarrow{P_\theta} \tau(\theta) \quad \forall \theta \in \Theta$

2) — " — сильно состоятельной

$$\hat{\theta}(X_1 \dots X_n) \xrightarrow{P_\theta \text{ п.н.}} \tau(\theta)$$

3) Оценка $\hat{\Theta}$ наз-ся асимптотически нормальной оценкой Θ, если

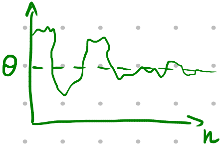
$$\sqrt{n}\,(\hat{\theta} - \Theta) \xrightarrow{d_\theta} \mathcal{N}(0, \Sigma(\theta)) \quad \forall \theta \in \Theta$$

где $\Sigma(\theta)$ – асимптотическая матрица ковариаций

$d = 1$ $\quad \Sigma(\Theta) = \sigma^2(\theta)$ – асимптотическая дисперсия

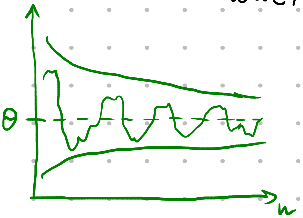
Смысл 1) Состоятельность

С ростом кол-ва наблюдений (n) вероятность большого отклонения $\hat{\Theta}$ от истинного Θ мала



Здесь нет никакой численной хар-ки степени отклонения

2) Асимптотическая нормальность



Даёт численную хар-ку степени отклонения

$$\sqrt{n}\,(\hat{\theta} - \theta) \overset{d_\theta\ \text{прибл}}{\sim} N(0, \underset{\overset{\|}{\sigma^2(\theta)}}{\Sigma(\theta)})$$

$$\hat{\theta} \overset{d_\theta\ \text{приб}}{\sim} N\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$$

При $d > 1$ $\Sigma(\theta)$ "задаёт трубу"

3) Сильная состоятельность $\Rightarrow$ конечное число плохих отклонений

имеет смысл для данных, поступающих последовательно

Пример: $X_1 \dots X_n$ — выборка из распр. Лапласа (θ) со сдвигом θ

$$p_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$$

ЗБЧ: $\overline{X}$ — (сильно) состоят оценка $E_\theta X_1 = \theta$

ЦПТ: $\sqrt{n}(\overline{X} - \theta) \xrightarrow{d_\theta} N(0, D_\theta X_1)$

$D_\theta X_1 = 2 \Rightarrow \overline{X}$ — а.н. оц. θ с а.д. $\sigma^2(\theta) = 2$

$\overline{X} \overset{\text{прибл.}}{\sim} N\left(\theta, \frac{2}{n}\right)$

По св-ву норм распр с вер-ю ≈ 0.95 $\overline{X} \in \left(\theta \pm 2\sqrt{2/n}\right)$

$\overline{X} - 2\sqrt{\frac{2}{n}} < \theta < \overline{X} + 2\sqrt{\frac{2}{n}}$

доверит интервал

Пусть $n = 200 \quad \bar{X} = 1$

Тогда нер-во примет вид $\quad 0{,}8 < \theta < 1{,}2$

реализация дов. интервала

Сильн. сост $\Rightarrow$

сост.

Ас. нормальность $\Rightarrow$

Утв. Пусть $X_1, \dots X_n$ – выборка, $E_\theta X_1^{2k} < +\infty$, тогда

$\overline{X^k}$ – (сильно) состоятельная, асимпт. норм. оценка $E_\theta X_1^k$

Наследование свойств

Пусть оценка $\psi(\theta)$ с какими-то св-вами

Хотим получить оценку на $\varphi(\theta)$ с теми же св-вами

Теорема (о наследовании сходимостей)

1) Пусть $\xi_n \xrightarrow{\text{п.н.}\,(P)} \xi$ — случ. векторы разм. m, $h: \mathbb{R}^m \to \mathbb{R}^k$ непр почти всюду относительно ξ (т.е. $\exists B_{\in \mathcal{B}(\mathbb{R}^m)}$ h непр в B и $P(\xi \in B) = 1$)

Тогда $h(\xi_n) \xrightarrow{\text{п.н.}\,(P)} h(\xi)$

2) Пусть $\xi_n \xrightarrow{d} \xi$ — случ векторы разм m, $h: \mathbb{R}^m \to \mathbb{R}^k$ непр

Тогда $h(\xi_n) \xrightarrow{d} h(\xi)$

Пример $\{\xi_n,\ n \in \mathbb{N}\}$ – н.о.р.с.в. $E\xi_n = a \neq 0$

УЗБЧ: $\frac{S_n}{n} \xrightarrow{п.н.} a = \xi$, где $S_n = \xi_1 + \ldots + \xi_n$

Тогда по т. о насл. сх. для $h(x) = \frac{1}{x}$

$$\frac{n}{S_n} = h\left(\frac{S_n}{n}\right) \xrightarrow{п.н.} h(a) = \frac{1}{a}$$

Утв. Пусть $\hat{\theta}$ – (сильно) сост. оценка θ

$\tau(\theta)$ – непр на Θ ф-ция

$\tau(\hat{\theta})$ – (сильно) сост оценка $\tau(\theta)$

Теорема (лемма Слуцкого)

Пусть $\{\xi_n, n \in \mathbb{N}\}$, $\{\eta_n, n \in \mathbb{N}\}$ посл-ти сл. вел. (независ.)

$$\xi_n \xrightarrow{d} \xi \qquad \eta_n \xrightarrow{d} C = const$$

$$\Rightarrow \quad \xi_n + \eta_n \xrightarrow{d} \xi + C$$

$$\xi_n \eta_n \xrightarrow{d} C\xi$$

Замечание Лемма Слуцкого не следует из т. о насл. сход., т.к. из сходимости компонент вектора по d не следует сходимость вектора

Теорема (о производной)

Пусть ξ_n, ξ — случ. вект. разм d

$\xi_n \xrightarrow{d} \xi$, $h(x): \mathbb{R}^d \to \mathbb{R}^k$ — непр. диф. ф-ция в (·) $a \in \mathbb{R}^d$

числ. посл-ть $b_n \neq 0$, $b_n \to 0$

Тогда $$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} h'(a)\xi$$

$$h'(a) = \left.\frac{\partial h}{\partial x}\right|_a$$

▲ $d = 1$

$$H(x) = \begin{cases} \dfrac{h(a+x) - h(a)}{x}, & x \neq 0 \\ h'(a), & x = 0 \end{cases}$$

$H(x)$ непр в $(\cdot)$ $x = 0$

$\xi_n b_n \xrightarrow{d} 0_\xi = 0$ (по л. Слуцкого)

Применим т. о насл. сх.

$$H(\xi_n b_n) = \frac{h(\xi_n b_n + a) - h(a)}{\xi_n b_n} \xrightarrow{d} H(0) = h'(a) = \text{const}$$

Умножим на ξ_n и применим л. Слуцкого ■

Пример Пусть ξ_n – н.о.р.сл.в. $E\xi_n = a$ $D\xi_n = \sigma^2$

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{d} ? \qquad S_n = \xi_1 + \dots + \xi_n$$

Применим т. о производной $\xi_n = \sqrt{n}\left(\frac{S_n}{n} - a\right) \underset{\text{ЦПТ}}{\xrightarrow{d}} N(0, \sigma^2)$

$$h(x) = \frac{1}{x} \qquad b_n = \frac{1}{\sqrt{n}}$$

$$\Rightarrow \frac{h(a + \xi_n b_n) - h(a)}{b_n} = \sqrt{n}\left(h\left(\frac{S_n}{n}\right) - h(a)\right) = \sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{d} N(0, \sigma^2)\left(-\frac{1}{a^2}\right) =$$

$$= N\left(0, \frac{\sigma^2}{a^4}\right)$$

Следствие $\frac{1}{\overline{X}}$ — ас. норм. оц. $\frac{1}{a}$ с ас. дисп. $\frac{\sigma^2}{a^4}$

Мы использовали:

- т. о производной
- $h(x) = \ldots$

Теорема (Дельта (δ) - метод)

Пусть $\hat{\theta}$ - ас. норм. оц. $\theta \in \Theta \subset \mathbb{R}^d$ с ас. матр. ков. $\Sigma(\theta)$

Пусть $\tau(x): \mathbb{R}^d \to \mathbb{R}^k$ - непр дифф-ма на Θ

Тогда $\tau(\hat{\theta})$ - ас. норм. оц. $\tau(\theta)$ с ас. матр. ков. $D(\theta)\,\Sigma(\theta)\,D(\theta)^T$

где $D(\theta) = \left.\dfrac{\partial \tau}{\partial x}\right|_{x=\theta}$

$\blacktriangle$ В т. о произв. $h(x) = \tau(x)$ $\quad b_n = \dfrac{1}{\sqrt{n}}$

$$\xi_n = \sqrt{n}\,(\hat{\theta} - \theta) \qquad a = \theta$$

$$\frac{h(\theta - \xi_n b_n) - h(\theta)}{b_n} = \sqrt{n}\left(h\left(\theta + \frac{1}{\sqrt{n}}\sqrt{n}\,(\hat{\theta} - \theta)\right) - h(\theta)\right) =$$

$$= \sqrt{n}\left(\tau(\hat{\theta}) - \tau(\theta)\right) \to N(0, \Sigma(\theta)) \cdot \underbrace{\frac{\partial \tau}{\partial x}\Big|_{x=\theta}}_{\overset{\shortparallel}{D(\theta)}} = N(0, D(\theta)\Sigma(\theta)D(\theta)^T)$$

Пример $X_1 \dots X_n \sim \mathcal{E}xp(\theta)$, $\theta > 0$

ЦПТ $\sqrt{n}\left(\overline{X} - \frac{1}{\theta}\right) \xrightarrow{d_\theta} N\left(0, \frac{1}{\theta^2}\right)$

$\frac{1}{\theta} = E_\theta X_1$, $\frac{1}{\theta^2} = D_\theta X_1$

$\Rightarrow$ $\overline{X}$ – а.н.о. $\frac{1}{\theta}$ с ас. д $\frac{1}{\theta^2}$

δ– метод для $\tau(x) = \frac{1}{x}$

$\tau(\overline{X}) = \frac{1}{\overline{X}}$ а.н.о $\theta = \tau\left(\frac{1}{\theta}\right)$ с ас. д. $\frac{1}{\theta^2}\left[\left(\frac{1}{x}\right)'_{x=\frac{1}{\theta}}\right]^2 = \frac{1}{\theta^2}\theta^4 = \theta^2$

Доверительный интервал

Опр Пусть $X = (X_1, \dots, X_n)$ — выборка из неизв расп.

$$P \in \mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$$

- Если $\Theta \subset \mathbb{R}$, то пара статистик $T_1(X)$, $T_2(X)$ наз-ся дов. интервалом для θ уровня доверия α, если

$$P(T_1(X) \leq \theta \leq T_2(X)) \geq \alpha \quad \forall \theta \in \Theta$$

$\uparrow$

если точное, то точный дов. интервал

- Если $\Theta \subset \mathbb{R}^d$, то $S(X) \subset \Theta$ наз-ся дов. областью для θ уровня доверия α

$$P_\theta(\theta \in S(x)) \geq \alpha \quad \forall \theta \in \Theta$$