Оценивание вещественного параметра (продолжение)

Малов Сергей Васильевич

Санкт-Петербургский государственный электротехнический университет

3 октября 2020 г.

План

1. Подчиненные и достаточные статистики
2. Несмещенное оценивание
3. Асимптотическое оценивание

Подчиненные и достаточные статистики

Определения

- Статистика $T : \mathfrak{X} \to E$ (измеримая) называется подчиненной, если ее распределение не зависит от θ
- Статистика $T : \mathfrak{X} \to E$ (измеримая) называется достаточной, если условное распределение наблюдений при условии T не зависит от θ,
$$\mathbb{P}_\theta(X \in A | T) = f(A), A \in \mathfrak{F}.$$

Свойства

- Подчиненная статистика не несет в себе статистической информации о параметре
- Достаточная статистика содержит всю статистическую информацию о параметре

Достаточные статистики

Некоторые примеры

- Весь набор наблюдений $X \in \mathfrak{X}$ – достаточная статистика
  - такая достаточная статистика называется тривиальной
- Если $X = (X_1, \ldots, X_n)$ – выборка из распределения с функцией распределения F, то набор порядковых статистик $X_{(1)}, \ldots, X_{(n)}$ – достаточная статистика

Упражнение

Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения Бернулли $\mathrm{Bi}(1, p)$. Показать, что $T = \sum_{i=1}^n X_i$ – достаточная статистика.

Решение. По определению условной вероятности и формуле Бернулли

$$\mathbb{P}_p(X_1 = i_1, \ldots, X_n = i_n | T = k) = \frac{\mathbb{P}_p(X_1 = i_1, \ldots, X_n = i_n, T = k)}{\mathbb{P}(T = k)}$$

$$= \frac{p^k(1-p)^{n-k} 1_{\{\sum_{j=1}^n i_j = k\}}}{C_n^k p^k (1-p)^{n-k}} = 1_{\{\sum_{j=1}^n i_j = k\}} / C_n^k$$

не зависит от $p \in (0, 1)$. $\Rightarrow$ T – достаточная статистика. ■

Достаточные статистики

Некоторые примеры

- Весь набор наблюдений $X \in \mathfrak{X}$ – достаточная статистика
  - такая достаточная статистика называется тривиальной
- Если $X = (X_1, \ldots, X_n)$ – выборка из распределения с функцией распределения F, то набор порядковых статистик $X_{(1)}, \ldots, X_{(n)}$ – достаточная статистика

Упражнение

Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения Бернулли $\mathrm{Bi}(1, p)$. Показать, что $T = \sum_{i=1}^n X_i$ – достаточная статистика.

Решение. По определению условной вероятности и формуле Бернулли

$$\mathbb{P}_p(X_1 = i_1, \ldots, X_n = i_n | T = k) = \frac{\mathbb{P}_p(X_1 = i_1, \ldots, X_n = i_n, T = k)}{\mathbb{P}(T = k)}$$

$$= \frac{p^k(1-p)^{n-k} 1_{\{\sum_{j=1}^n i_j = k\}}}{C_n^k p^k (1-p)^{n-k}} = 1_{\{\sum_{j=1}^n i_j = k\}} / C_n^k$$

не зависит от $p \in (0, 1)$. $\Rightarrow$ T – достаточная статистика. ∎

Теорема фактроизации

Функция правдоподобия

Теорема (Нейман–Фишер)

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ – статстический эксперимент; $\mathcal{P} \ll \mu$ и $p_\theta = \frac{d\mathbb{P}_\theta}{d\theta} : \theta \in \Theta$ – соответствующие плотности распределения; $L(x;\theta) = p_\theta(x)$ – функция правдоподобия. Тогда, статистика $T : \mathfrak{X} \to E$ достаточна $\Leftrightarrow$ существуют функции $g_\theta : E \to \mathbb{R}$ и $h : \mathfrak{X} \to \mathbb{R}$, такие что

$$L(X;\theta) = g_\theta(T(X))h(X)$$

с вероятностью $\mathbb{P}_\theta$ равной 1 при каждом $\theta \in \Theta$

- Функция правдоподобия $L_\theta : \mathfrak{X} \to \mathbb{R}_+$, $\theta \in \Theta$ – достаточная статистика

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$; параметр $\theta = (a, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Найти минимальную достаточную статистику.

Решение. Функция правдоподобия

$$\begin{aligned} L(X;\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\Big(-\frac{(x_i - a)^2}{2\sigma^2}\Big) \\ &= \frac{1}{(2\pi)^n \sigma^n} \exp\Big(-\frac{1}{2\sigma^2} \textstyle\sum_{i=1}^{n} (x_i - a)^2\Big) \\ &= \frac{1}{(2\pi)^n \sigma^n} \exp\Big(-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + \frac{a\sum_{i=1}^{n} x_i}{\sigma^2} - \frac{na^2}{2\sigma^2}\Big)\Big) \end{aligned}$$

Выбираем $g_\theta(u, v) \cong L(X;\theta)$ (с точностью до известной постоянной) при $u = \sum_{i=1}^{n} x_i$ и $v = \sum_{i=1}^{n} x_i^2 \Rightarrow (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ – минимальная достаточная статистика. ∎

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$; параметр $\theta = (a, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Найти минимальную достаточную статистику.

Решение. Функция правдоподобия

$$
\begin{aligned}
L(X;\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\Big(-\frac{(x_i-a)^2}{2\sigma^2}\Big) \\
&= \frac{1}{(2\pi)^n\sigma^n} \exp\Big(-\frac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (x_i-a)^2\Big) \\
&= \frac{1}{(2\pi)^n\sigma^n} \exp\Big(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \frac{a\sum_{i=1}^n x_i}{\sigma^2} - \frac{na^2}{2\sigma^2}\Big)\Big)
\end{aligned}
$$

Выбираем $g_\theta(u,v) \cong L(X;\theta)$ (с точностью до известной постоянной) при $u = \sum_{i=1}^n x_i$ и $v = \sum_{i=1}^n x_i^2 \Rightarrow (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ – минимальная достаточная статистика. ∎

Минимальная достаточная статистика

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из равномерного распределения $U(a,b)$; параметр $\theta = (a,b) \in \{(a,b) \in \mathbb{R} : -\infty < a < b < \infty\}$. Найти минимальную достаточную статистику.

Решение. Плотность распределения

$$p_\theta(x) = \frac{1}{b-a} 1_{\{a \le x \le b\}}.$$

Функция правдоподобия

$$L(X;\theta) = \prod_{i=1}^{n} \frac{1}{b-a} 1_{\{a \le X_i \le b\}} = \frac{1}{(b-a)^n} 1_{\{X_{(1)} \ge a\}} 1_{\{X_{(n)} \le a\}}.$$

Очевидно, что $(X_{(1)}, X_{(n)})$ – минимальная достаточная статистика. ∎

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из равномерного распределения $U(a, b)$; параметр $\theta = (a, b) \in \{(a, b) \in \mathbb{R} : -\infty < a < b < \infty\}$. Найти минимальную достаточную статистику.

Решение. Плотность распределения

$$p_\theta(x) = \frac{1}{b-a} 1_{\{a \le x \le b\}}.$$

Функция правдоподобия

$$L(X; \theta) = \prod_{i=1}^{n} \frac{1}{b-a} 1_{\{a \le X_i \le b\}} = \frac{1}{(b-a)^n} 1_{\{X_{(1)} \ge a\}} 1_{\{X_{(n)} \le a\}}.$$

Очевидно, что $(X_{(1)}, X_{(n)})$ – минимальная достаточная статистика. ∎

Полная достаточная статистика

Определение

Достаточная статистика T называется полной, если

$$\mathbb{E}_\theta g(T) = 0,\ \forall \theta \in \Theta \qquad \Rightarrow \qquad \mathbb{P}_\theta(g(T) = 0) = 1,\ \theta \in \Theta.$$

- Никакая функция полной достаточной статистики, для которой существует математическое ожидание, кроме постоянной не является подчиненной статистикой
- Полная достаточная статистика всегда минимальна

Полная достаточная статистика

Упражнение

Пусть $X_1, \dots, X_n$– выборка из равномерного распределения $U(a,b)$; параметр $\theta = (a,b) \in \{(a,b) \in \mathbb{R} : -\infty < a < b < \infty\}$. Показать полноту минимальной достаточной статистики.

Решение (основная идея). Было получено, что $(X_{(1)}, X_{(n)})$ – минимальная достаточная статистика. Совместная плотность распределения вектора $(X_{(1)}, X_{(n)})$:

$$p_\theta(x,y) = n(n-1)\frac{(y-x)^{n-2}}{(b-a)^n} 1_{\{a \le x \le y \le b\}}.$$

Для простоты, g-непрерывна. Тогда

$$\mathbb{E}_\theta g(X_{(1)}, X_{(n)}) \cong \int_a^b dy \int_a^y g(x,y)(y-x)^{n-2} dx / (b-a)^n = 0$$

влечет (дифференцируем по b, затем по a)

$$\int_a^y g(x,b)(b-x)^{n-2} dx = 0, \ \forall b \quad \Rightarrow \quad g(a,b)(b-a)^{n-2} = 0, \ \forall a < b.$$

Следовательно, $g(a,b) = 0$, $\forall a, b : a < b$. ∎

Полная достаточная статистика

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из равномерного распределения $U(a,b)$; параметр $\theta = (a,b) \in \{(a,b) \in \mathbb{R} : -\infty < a < b < \infty\}$. Показать полноту минимальной достаточной статистики.

Решение (основная идея). Было получено, что $(X_{(1)}, X_{(n)})$ – минимальная достаточная статистика. Совместная плотность распределения вектора $(X_{(1)}, X_{(n)})$:

$$p_\theta(x,y) = n(n-1)\frac{(y-x)^{n-2}}{(b-a)^n} 1_{\{a \le x \le y \le b\}}.$$

Для простоты, g-непрерывна. Тогда

$$\mathbb{E}_\theta g(X_{(1)}, X_{(n)}) \cong \int_a^b dy \int_a^y g(x,y)(y-x)^{n-2} dx/(b-a)^n = 0$$

влечет (дифференцируем по b, затем по a)

$$\int_a^y g(x,b)(b-x)^{n-2} dx = 0, \ \forall b \quad \Rightarrow \quad g(a,b)(b-a)^{n-2} = 0, \ \forall a < b.$$

Следовательно, $g(a,b) = 0, \ \forall a, b : a < b.$ ∎

Полная достаточная статистика

Полная достаточная статистика для многопараметрического экспоненциального семейства

- Многопараметрическое экспоненциальное семейство распределениий имеет плотности
$$p_\theta(x) = h(x)\exp\Big(\sum_{j=1}^k a_j(\theta)\delta_j(x) + r(\theta)\Big),\ \theta \in \Theta \subseteq \mathbb{R}^k.$$
- Функция правдоподобия:
$$L(x;\theta) = \prod_{i=1}^n h(x_i)\exp\Big(\sum_{i=1}^n \sum_{j=1}^k a_j(\theta)\delta_j(x_i) + \nu(\theta)\Big).$$
- Достаточная статистика: $\delta = (\delta_1, \dots, \delta_k)$.
  - при известных $(\delta_1, \dots, \delta_k)$ условное распределение наблюдений не зависит от θ.
- Если набор $(\delta_1, \dots, \delta_k)$ несократим, то δ минимальная достаточная статистика
- При выполнении условия $\dim\{(a_1(\theta), \dots, a_k(\theta)), \theta \in \Theta\} = k$ (достаточное условие) минимальная достаточная статистика является полной

План

Оптимальное оценивание

- В классе всех оценок не существует оптимальной, т. е. минимизирующей риск при всех значениях θ
- В дальнейшем, для определения оптимальности оценки используем гауссовский риск

Несмещенное оценивание

Определение

Оценка $\delta(X)$ параметра θ называется несмещенной, если при любом значении параметра $\theta \in \Theta$

$$\mathbb{E}_\theta \delta(X) = \theta.$$

- В асимптотической схеме оценка $\delta(X) = \delta_n(X)$ называется асимптотически несмещенной, если

$$\mathbb{E}_\theta \delta(X) \to \theta \quad \text{при} \quad n \to \infty.$$

- Смещением оценки называется величина

$$b_g(\theta) = \mathbb{E}_\theta \delta(X) - g(\theta).$$

Несмещенное оценивание

Существовует ли несмещенная оценка?

Контрпример. Пусть $X = X_1$ одно наблюдение, распределение которого принадлежит классу распределений Пуассона, имеющее дискретную плотность вида

$$q(k;\theta) = \mathbb{P}\{X = k\} = \frac{\theta^k}{k!} e^{-1/\theta} \qquad \text{(неклассическая параметризация).}$$

Несмещенность означала бы, что при каждом θ имеет место

$$E_\theta \delta(X) = \sum_{x=0}^{\infty} \delta(x) \frac{\theta^x}{x!} e^{-\theta} = \frac{1}{\theta}, \qquad \sum_{x=0}^{\infty} \delta(x) \frac{\theta^{x+1}}{x!} = e^\theta = \sum_{r=0}^{\infty} \frac{\theta^r}{r!},$$

но этого не может быть.

- НРМД оценка является эффективной в классе несмещенных оценок

Несмещенное оценивание

Риск несмещенной оценки совпадает с ее дисперсией

$$R_\delta(\theta) = \mathbb{E}_\theta(\delta(X) - \theta)^2 = \mathbb{E}_\theta(\delta(X) - \mathbb{E}_\theta\delta(X))^2.$$

Определение

Несмещенная оценка $\delta(X)$ параметра θ называется несмещенной с равномерно-минимальной дисперсией (НРМД), если для любой несмещенной оценки $\delta^*(X)$

$$R_\delta(\theta) \le R_{\delta^*}(\theta), \quad \text{при всех} \quad \theta \in \Theta.$$

- НРМД оценка является оптимальной в классе несмещенных оценок

Единственность НРМД-оценки

(!) НРМД оценка не всегда существует.

Теорема

Может существовать не более одной НРМД-оценки параметра $\theta \in \Theta$.

Доказательство. Допустим, что существуют две НРМД-оценки δ_1 и δ_2. Рассмотрим оценку $\delta_3 = (\delta_1 + \delta_2)/2$. Имеем, по определению НРМД-оценки при каждом $\theta \in \Theta$

$$\mathbb{D}_\theta \delta_3 = \frac{\mathbb{D}_\theta \delta_1 + \mathbb{D}_\theta \delta_2 + 2\mathrm{cov}(\delta_1, \delta_2)}{4} \geq \mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2.$$

Отсюда следует, что $\mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2 \leq \mathrm{cov}(\delta_1, \delta_2)$. Тогда $\mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2 = \mathrm{cov}(\delta_1, \delta_2)$ и, следовательно,

$$\mathbb{D}(\delta_1 - \delta_2) = \mathbb{D}_\theta \delta_1 + \mathbb{D}_\theta \delta_2 - 2\mathrm{cov}(\delta_1, \delta_2) = 0.$$

Таким образом, $\delta_1 = \delta_2$ с вероятностью 1. ■

Единственность НРМД-оценки

(!) НРМД оценка не всегда существует.

Теорема

Может существовать не более одной НРМД-оценки параметра $\theta \in \Theta$.

Доказательство. Допустим, что существуют две НРМД-оценки δ_1 и δ_2. Рассмотрим оценку $\delta_3 = (\delta_1 + \delta_2)/2$. Имеем, по определению НРМД-оценки при каждом $\theta \in \Theta$

$$\mathbb{D}_\theta \delta_3 = \frac{\mathbb{D}_\theta \delta_1 + \mathbb{D}_\theta \delta_2 + 2\mathrm{cov}(\delta_1, \delta_2)}{4} \geq \mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2.$$

Отсюда следует, что $\mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2 \leq \mathrm{cov}(\delta_1, \delta_2)$. Тогда $\mathbb{D}_\theta \delta_1 = \mathbb{D}_\theta \delta_2 = \mathrm{cov}(\delta_1, \delta_2)$ и, следовательно,

$$\mathbb{D}(\delta_1 - \delta_2) = \mathbb{D}_\theta \delta_1 + \mathbb{D}_\theta \delta_2 - 2\mathrm{cov}(\delta_1, \delta_2) = 0.$$

Таким образом, $\delta_1 = \delta_2$ с вероятностью 1. ■

Несмещенное оценивание и достаточные статистики

Теорема (Рао–Блэкуэлл–Колмогоров)

Пусть T – достаточная статистика для семейства $\mathcal{P}$; δ – оценка параметра θ; $\eta = \eta(T) = \mathbb{E}(\delta|T)$ –условное математическое ожидание δ при условии T. Тогда:

(i). $R_\delta(\theta) \geq R_\eta(\theta)$ для любого $\theta \in \Theta$

(ii). $R_\delta(\theta) > R_\eta(\theta)$, если $\mathbb{P}_\theta(\delta = \eta) < 1$

- НРМД является функцией от минимальной достаточной статистики
- Утверждение (i) теоремы остается верным для любой выпуклой (вниз) функции потерь
- Утверждение (ii) теоремы остается верным, если функция потерь строго выпуклая (вниз)

Теорема (Леман–Шеффе)

Существует не более одной несмещенной (с фиксированным смещением) оценки параметра θ, являющейся функцией от полной достаточной статистики.

Выводы

- Эффективные оценки в классе оценок с фиксированным смещением следует искать как функции от минимальных достаточных статистик
- Если минимальная достаточная статистика является полной, то при каждом фиксированном значении смещения существует единственная оценка, минимизирующая риск в классе оценок с таким смещением
- Если несмещенная оценка $\delta(T)$ является функцией от полной достаточной статистики T, то δ – НРМД-оценка

Алгоритмы несмещенного оценивания

Алгоритм 1

- Найти оценку $\delta_0(T)$, являющуюся функцией от минимальной достаточной статистики T (н-р, оценку максимального правдоподобия)
- Доказать полноту минимальной достаточной статистики T
- Скорректировать смещение $\delta(T) = f(\delta_0(T), T)$: $\mathbb{E}_\theta \delta_0 = 0$, $\theta \in \Theta$

Алгоритм 2

- Найти произвольную несмещенную оценку δ_0, являющуюся функцией от минимальной достаточной статистики T (например, оценку максимального правдоподобия)
- Найти минимальную достаточную статистику T и доказать ее полноту
- Вычислить условное математическое ожидание $\delta(T) = \mathbb{E}_\theta(\delta_0(X)|T)$.

План

1. Подчиненные и достаточные статистики
2. Несмещенное оценивание
3. Асимптотическое оценивание

Состоятельность

Определение

Оценка $\delta(X)$ параметра θ называется состоятельной, если при каждом значении $\theta \in \Theta$ для любого $\epsilon > 0$

$$\mathbb{P}_\theta(|\delta(X) - \theta| > \epsilon) \to 0,$$

при $n \to \infty$, и сильно состоятельной, если $\delta(X) \to \theta$ при $n \to \infty$ с вероятностью $\mathbb{P}_\theta = 1$ при каждом $\theta \in \Theta$.

- Состоятельность означает сходимость $\delta(X) \to \theta$ при $n \to \infty$ по вероятности $\mathbb{P}_\theta$ при каждом $\theta \in \Theta$.
- Для практических целей состоятельность и сильная состоятельность неразличимы (поскольку рассматриваются конечные наборы наблюдений), однако с точки зрения асимптотической теории условие сильной состоятельности более жесткое.
- В асимптотической статистике обычно ограничиваются рассмотрением только состоятельных оценок

Асимптотическая нормальность

Определение

Оценка δ параметра θ называется асимптотически нормальной, если при каждом $\theta \in \Theta$ существует $\sigma^2(\theta)$, такая что

$$\mathbb{P}_\theta(\sqrt{n}(\delta(X) - \theta) < x) \to \Phi(x/\sigma(\theta)),$$

$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ – функция стандартного нормального распределения.

- Асимптотическая нормальность может быть переписана в терминах сходимости по распределению

$$\sqrt{n}(\delta(X) - \theta) \Longrightarrow N(0, \sigma^2(\theta)).$$

- Асимптотическая нормальность оценки влечет ее состоятельность
- Асимптотическая нормальность – удобное свойство для построения доверительных оценок и статистических критериев

Асимптотическая эффективность

Определение

Оценка $\delta(X) \in \mathcal{C}$ параметра θ называется асимптотически эффективной в классе $\mathcal{C}$, если для любой оценки $\delta^*(X)$

$$\liminf_{n\to\infty}(R_\delta(\theta) - R_{\delta^*}(\theta)) \geq 0, \text{ для люого } \theta \in \Theta.$$

Определение

Асимптотически нормальная оценка $\delta(X)$:

$$\sqrt{n}(\delta(X) - \theta) \Rightarrow \mathcal{N}(0, \sigma^2(\theta))$$

называется асимптотически эффективной по Питмэну, если для любой асимтотически нормальной оценки $\delta^*(X)$:

$$\sqrt{n}(\delta^*(X) - \theta) \Rightarrow \mathcal{N}(0, \sigma_*^2(\theta))$$

при любом $\theta \in \Theta$ выполнено неравенство

$$\sigma^2(\theta) \leq \sigma_*^2(\theta).$$

Оценка дисперсии

Упражнение

Найти эффективную оценку дисперсии в классе оценок вида $s(\lambda) = \lambda s'^2$: $s'^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ (несмещенная оценка дисперсии), $\lambda > 0$, по выборке из нормального распределения $\mathcal{N}(a, \sigma^2)$.

Решение. Вычислим среднеквадратичное отклонение (риск) оценок этого типа:

$$\mathbb{E}_\theta(s(\lambda) - \sigma^2)^2 = \mathbb{E}_\theta(\lambda(s'^2 - \sigma^2) + (\lambda - 1)\sigma^2)^2 = \lambda^2 \mathbb{D}s'^2 + (\lambda - 1)^2\sigma^4.$$

По теореме Фишера $(n-1)s'^2/\sigma^2$ имеет распределение χ^2_{n-1}. Дисперсия распределения χ^2_{n-1} равна $2\,(n-1)$. Тогда $\mathbb{D}s'^2 = = 2\sigma^4/(n-1)$. Следовательно,

$$\mathbb{E}_\theta(s(\lambda) - \sigma^2)^2 = \left(\frac{2\lambda^2}{n-1} + (\lambda - 1)^2\right)\sigma^4.$$

Минимум выражения в скобках достигается при $\lambda = \frac{n-1}{n+1}$. Таким образом, наилучшей в указанном классе оценок дисперсии нормальной выборки является смещенная оценка $\dfrac{1}{n+1}\sum_{i=1}^n (X_i - \overline{X})^2$. ■