Теория вероятностей и математическая статистика

Лекция 15

Точечные оценки

Параметрическая статистическая модель

Статистической моделью называется выборочное пространство, на котором задано некоторое семейство распределений.

Статистическая модель называется ***параметрической***, если функции распределения заданного семейства $F\left(x,\vec{\theta}\right)$ зависят от некоторого вектора параметров $\vec{\theta} = (\theta_1, \dots, \theta_r) \in \Theta$.

Множество Θ всевозможных значений параметров называется ***параметрическим множеством***.

Задача нахождения параметров $\vec{\theta} = (\theta_1, \dots, \theta_r) \in \Theta$ для теоретической функции $F\left(x,\vec{\theta}\right)$ по заданной выборке.

Пример: масса распределена по нормальному закону с неизвестными $\theta_1 = m$ и $\theta_2 = \sigma$.

Точечная оценка параметра

Точечной оценкой параметра θ называется статистика $\hat{\theta}(\vec{X}_n)$, т. е. функция от наблюдений.

Примеры точечных оценок для $MX = \mu$:

$$\hat{\theta}_1(\vec{X}_n) = \bar{X},$$

$$\hat{\theta}_2(\vec{X}_n) = \frac{X_{(1)} + X_{(n)}}{2},$$

$$\hat{\theta}_3(\vec{X}_n) = \begin{cases} \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right), & n - \text{чётное}, \\ X_{\left(\frac{n+1}{2}\right)}, & n - \text{нечётное}, \end{cases}$$

$$\hat{\theta}_4(\vec{X}_n) = X_1.$$

Характеристики точечных оценок

Точечную оценку $\hat{\theta}(\vec{X}_n)$ параметра θ можно рассматривать как последовательность случайных величин для $n \in \mathbb{N}$.

Точечная оценка параметра θ называется ***несмещённой***, если $\forall n \in \mathbb{N}$ её математическое ожидание совпадает с оцениваемым параметром θ:

$$\forall n \in \mathbb{N}, \qquad M\hat{\theta}(\vec{X}_n) = \theta.$$

Точечная оценка параметра θ называется ***состоятельной***, если с ростом объёма выборки n она сходится по вероятности к оцениваемому параметру:

$$\hat{\theta}(\vec{X}_n) \xrightarrow{P} \theta \quad \text{при} \quad n \to +\infty.$$

Теорема: $M\hat{\theta} \to \theta$ и $D\hat{\theta} \to 0$ при $n \to +\infty \;\Rightarrow\; \hat{\theta}$ состоятельна.

Несмещённая оценка называется ***эффективной*** в некотором классе несмещённых оценок, если её дисперсия наименьшая в этом классе.

Выборочное среднее

Рассмотрим класс линейных точечных оценок с единичным суммарным коэффициентом:

$$\tilde{\theta}(\vec{X}_n) = \sum_{i=1}^{n} \alpha_i X_i, \qquad \sum_{i=1}^{n} \alpha_i = 1.$$

Теорема: выборочное среднее $\bar{X}$ является несмещённой, состоятельной и эффективной (в классе линейных оценок с единичным суммарным коэффициентом) оценкой математического ожидания $MX = \mu$.

Доказательство: $M\bar{X} = M\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum MX_i = \mu \ \Rightarrow \ \bar{X}$ несмещённая;

ЗБЧ в форме Чебышёва $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu \ \Rightarrow \ \bar{X}$ состоятельная;

для функции $D\tilde{\theta}(\vec{X}_n) = \sigma^2 \sum_{i=1}^{n} \alpha_i^2$ найдём усл. минимум при $\sum_{i=1}^{n} \alpha_i = 1$,

функция Лагранжа $L(\alpha_1, \dots, \alpha_n, \lambda) = \sigma^2 \sum_{i=1}^{n} \alpha_i^2 + \lambda(\sum_{i=1}^{n} \alpha_i - 1)$, тогда

$\forall i \in \overline{1,n}$, $\frac{\partial L}{\partial \alpha_i} = 2\sigma^2 \alpha_i + \lambda = 0$ и минимум достигается при $\alpha_i = \frac{1}{n}$ $\square$

Выборочная и исправленная дисперсии

Теорема: выборочная дисперсия $\hat{\sigma}^2(\vec{X}_n)$ является смещённой и состоятельной оценкой дисперсии $DX = \sigma^2$.

Доказательство смещённости:

$$M\hat{\sigma}^2(\vec{X}_n) = M\left(\frac{1}{n}\sum\left((X_i-\mu)-(\bar{X}-\mu)\right)^2\right) = M\left(\frac{1}{n}\sum(X_i-\mu)^2-(\bar{X}-\mu)^2\right) =$$

$$= \frac{1}{n}\sum DX_i - D\bar{X} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \quad \square$$

Исправленная дисперсия: $$s^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

Исправленное отклонение: $$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}.$$

Метод моментов

Для оценки r-мерного вектора параметров $\vec{\theta} = (\theta_1, \dots, \theta_r)$ выбирают r моментов и приравнивают выборочные моменты к теоретическим:

$$\begin{cases} \hat{\mu}_{i_\alpha}(\vec{X}_n) = \mu_{i_\alpha}(\vec{\theta}) \\ \hat{\nu}_{j_\beta}(\vec{X}_n) = \nu_{j_\beta}(\vec{\theta}) \end{cases} \qquad \alpha \in \overline{1, k}, \qquad \beta \in \overline{1, l}, \qquad k + l = r.$$

Если решение непрерывно зависит от выборочных моментов, то полученная точечная оценка будет состоятельной.

Пример: найдём точечные оценки границ промежутка $[a, b]$, на котором случайная величина равномерно распределена, выберем начальный момент первого порядка и центральный момент второго порядка

$$\begin{cases} \hat{\mu}_1(\vec{X}_n) = \mu_1(a, b) \\ \hat{\nu}_2(\vec{X}_n) = \nu_2(a, b) \end{cases} \qquad \begin{cases} \bar{x} = \dfrac{a + b}{2} \\ \hat{\sigma}^2(\vec{x}_n) = \dfrac{(b - a)^2}{12} \end{cases} \qquad \begin{array}{l} \hat{a} = \bar{x} - \sqrt{3}\,\hat{\sigma}(\vec{x}_n) \\ \hat{b} = \bar{x} + \sqrt{3}\,\hat{\sigma}(\vec{x}_n) \end{array}$$

Метод максимального правдоподобия

Функцией правдоподобия называется функция от случайной выборки $\vec{X}_n$ и вектора параметров $\vec{\theta}$

$$L(\vec{X}_n, \vec{\theta}) = \prod_{i=1}^{n} p(X_i, \vec{\theta}),$$

где для дискретного случая $p(X_i, \vec{\theta})$ является случайной величиной, которая в случае если $X_i = a$ принимает значение, равное вероятности $P(X_i = a)$, и поэтому $L(\vec{x}_n, \vec{\theta}) = P(X_1 = x_1, \dots, X_n = x_n)$;

а для непр. случая $p(X_i, \vec{\theta})$ является плотностью распределения X_i, и поэтому $L(\vec{X}_n, \vec{\theta}) = p_{\vec{X}_n}(\vec{X}_n, \vec{\theta})$.

Для фиксированной реализации $\vec{x}_n$ функция $L(\vec{x}_n, \vec{\theta})$ является функцией от r переменных $\theta_1, \dots, \theta_r$.

Оценкой максимального правдоподобия $\hat{\vec{\theta}}$, соответствующей выборке $\vec{x}_n$, называется точка максимума функции правдоподобия: $L\left(\vec{x}_n, \hat{\vec{\theta}}\right) = \max\limits_{\vec{\theta} \in \Theta} L(\vec{x}_n, \vec{\theta})$.

Уравнениями правдоподобия называются необх. условия точки максимума: $\forall k \in \overline{1, r}, \ \dfrac{\partial L}{\partial \theta_k} = 0$.

Часто удобно рассматривать ***логарифмическую функцию правдоподобия***:

$$\ln L(\vec{X}_n, \vec{\theta}) = \sum_{i=1}^{n} \ln p(X_i, \vec{\theta}),$$

тогда ***уравнения правдоподобия*** принимают вид: $\dfrac{\partial \ln L}{\partial \theta_k} = 0$.

Пример

Найдём точечные оценки максимального правдоподобия для параметров m и σ нормального распределения:

$$L(\vec{X}_n, \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2\sqrt{2\pi}} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2^2}} = \frac{1}{\left(\theta_2\sqrt{2\pi}\right)^n} e^{-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2},$$

$$\ln L(\vec{X}_n, \theta_1, \theta_2) = -n\ln\left(\theta_2\sqrt{2\pi}\right) - \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2,$$

$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1} = \dfrac{1}{\theta_2^2}\displaystyle\sum_{i=1}^{n}(x_i-\theta_1) = 0 \\ \dfrac{\partial \ln L}{\partial \theta_2} = -\dfrac{n}{\theta_2} + \dfrac{1}{\theta_2^3}\displaystyle\sum_{i=1}^{n}(x_i-\theta_1)^2 = 0 \end{cases}$$

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\theta}_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \hat{\sigma}$$

Регулярная статистическая модель

Дискретная стат. модель со скалярным параметром θ называется ***регулярной***, если

$$\forall f(\vec{x}_n), \qquad \frac{\partial}{\partial\theta}\sum_{\vec{x}_n} f(\vec{x}_n)\, L(\vec{x}_n,\theta) = \sum_{\vec{x}_n} f(\vec{x}_n)\,\frac{\partial L(\vec{x}_n,\theta)}{\partial\theta}.$$

Непрерывная стат. модель со скалярным параметром θ называется ***регулярной***, если

$$\forall f(\vec{x}_n), \qquad \frac{\partial}{\partial\theta}\int_{\mathbb{R}^n} f(\vec{x}_n)\, p_{\vec{X}_n}(\vec{x}_n,\theta)\, d\vec{x}_n = \int_{\mathbb{R}^n} f(\vec{x}_n)\,\frac{\partial p_{\vec{X}_n}(\vec{x}_n,\theta)}{\partial\theta}\, d\vec{x}_n,$$

где $d\vec{x}_n = dx_1 \dots dx_n$.

Для регулярной статистической модели

$$M\left(\frac{\partial \ln L(\vec{X}_n,\theta)}{\partial\theta}\right) = \int_{\mathbb{R}^n} \frac{\partial \ln p_{\vec{X}_n}(\vec{x}_n,\theta)}{\partial\theta}\, p_{\vec{X}_n}(\vec{x}_n,\theta)\, d\vec{x}_n = \int_{\mathbb{R}^n} \frac{\partial p_{\vec{X}_n}(\vec{x}_n,\theta)}{\partial\theta}\, d\vec{x}_n =$$

$$= \frac{\partial}{\partial\theta}\int_{\mathbb{R}^n} p_{\vec{X}_n}(\vec{x}_n,\theta)\, d\vec{x}_n = 0,$$

$$M\left(\frac{\partial \ln L(\vec{X}_n,\theta)}{\partial\theta}\right) = M\left(\sum_{i}^{n} \frac{\partial \ln p(X_i,\theta)}{\partial\theta}\right) = n\cdot M\left(\frac{\partial \ln p(X,\theta)}{\partial\theta}\right) = 0.$$

Информация Фишера

Пусть дана статистическая модель со скалярным параметром θ.

Информацией Фишера в n ***наблюдениях*** называется дисперсия логарифмической производной функции правдоподобия:

$$I_n(\theta) = D\left(\frac{\partial \ln L\left(\vec{X}_n, \theta\right)}{\partial \theta}\right) = M\left(\frac{\partial \ln L\left(\vec{X}_n, \theta\right)}{\partial \theta}\right)^2.$$

Информацией Фишера в одном наблюдении называется дисперсия логарифмической производной плотности распределения генеральной совокупности:

$$I(\theta) = D\left(\frac{\partial \ln p(X, \theta)}{\partial \theta}\right) = M\left(\frac{\partial \ln p(X, \theta)}{\partial \theta}\right)^2.$$

Для регулярной статистической модели

$$I_n(\theta) = D\left(\frac{\partial \ln L\left(\vec{X}_n, \theta\right)}{\partial \theta}\right) = n \cdot D\left(\frac{\partial \ln p(X, \theta)}{\partial \theta}\right) = n \cdot I(\theta).$$

Неравенство Рао-Крамера

Теорема (неравенство Рао-Крамера): для несмещённой оценки $\hat{\theta}(\vec{X}_n)$ параметра θ в регулярной статистической модели $D\hat{\theta}(\vec{X}_n) \geq \frac{1}{I_n(\theta)}$.

Доказательство: для $f(\vec{x}_n)$ и $g(\vec{x}_n)$ определим их скалярное произведение

$$(f, g) = M\big(f(\vec{x}_n)\, g(\vec{x}_n)\big) = \int_{\mathbb{R}^n} f(\vec{x}_n)\, g(\vec{x}_n)\, p_{\vec{X}_n}(\vec{x}_n)\, d\vec{x}_n$$

и используем неравенство Коши-Буняковского $(f, g)^2 \leq (f, f)(g, g)$,

для непрерывного случая

$$D\hat{\theta}(\vec{X}_n) \cdot I_n(\theta) = \int_{\mathbb{R}^n} \big(\hat{\theta}(\vec{x}_n) - \theta\big)^2 p_{\vec{X}_n}(\vec{x}_n, \theta)\, d\vec{x}_n \cdot \int_{\mathbb{R}^n} \left(\frac{\partial \ln L(\vec{x}_n, \theta)}{\partial \theta}\right)^2 p_{\vec{X}_n}(\vec{x}_n, \theta)\, d\vec{x}_n \geq$$

$$\geq \left(\int_{\mathbb{R}^n} \big(\hat{\theta}(\vec{x}_n) - \theta\big) \frac{\partial \ln L(\vec{x}_n, \theta)}{\partial \theta}\, p_{\vec{X}_n}(\vec{x}_n, \theta)\, d\vec{x}_n\right)^2 = \left(\int_{\mathbb{R}^n} \big(\hat{\theta}(\vec{x}_n) - \theta\big) \frac{\partial p_{\vec{X}_n}(\vec{x}_n, \theta)}{\partial \theta}\, d\vec{x}_n\right)^2 =$$

$$= \left(\frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\theta}(\vec{x}_n) p_{\vec{X}_n}(\vec{x}_n, \theta)\, d\vec{x}_n - \theta \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} p_{\vec{X}_n}(\vec{x}_n, \theta)\, d\vec{x}_n\right)^2 = \left(\frac{\partial}{\partial \theta}\Big(M\hat{\theta}(\vec{X}_n)\Big) - \theta \frac{\partial}{\partial \theta}(1)\right)^2 = 1$$

$\square$

Критерий эффективности регулярных статистических моделей

Следствие: для несмещённой оценки $\hat{\theta}(\vec{X}_n)$ параметра θ в регулярной статистической модели $D\hat{\theta}(\vec{X}_n) = \frac{1}{I_n(\theta)} \Leftrightarrow \frac{\partial \ln L(\vec{X}_n,\theta)}{\partial\theta} = a(\theta)\left(\hat{\theta}(\vec{X}_n) - \theta\right)$, при этом $a(\theta) = \frac{1}{D\hat{\theta}(\vec{X}_n)}$.

Доказательство: из доказательства теоремы видим, что для функций $f = \frac{\partial \ln L(\vec{X}_n,\theta)}{\partial\theta}$ и $g = \hat{\theta}(\vec{X}_n) - \theta$ равенство $D\hat{\theta}(\vec{X}_n) = \frac{1}{I_n(\theta)}$ эквивалентно равенству в неравенстве Коши-Буняковского $(f,g)^2 = (f,f)(g,g)$, что, как известно из линейной алгебры, эквивалентно равенству $\frac{f}{\|f\|} = \frac{g}{\|g\|}$; в данном случае $1 = D\hat{\theta}(\vec{X}_n) \cdot I_n(\theta) = (f,g)^2 = (f,f)(g,g) = \|f\|^2 \cdot \|g\|^2$, поэтому $g = \frac{f}{\|f\|^2}$, где $\|f\|^2 = (f,f) = \int_{\mathbb{R}^n}\left(\hat{\theta}(\vec{x}_n) - \theta\right)^2 p_{\vec{X}_n}(\vec{x}_n,\theta)\, d\vec{x}_n = D\hat{\theta}(\vec{X}_n)$ □