

Семинар №5

Сравнение оценок. Эффективные оценки.

Сравнение оценок.

Пусть $g(x, y) > 0$ – некоторая функция (называемая функцией потерь, как правило, симметричная) на \mathbb{R}^2 . Например, $g(x, y) = |x - y|$ или $g(x, y) = (x - y)^2$ (так называемая квадратичная функция потерь).

Определение. Пусть $\theta^*(X)$ – оценка параметра θ . Тогда функция $R(\theta^*, \theta) = E_{\theta}g(\theta^*(X), \theta)$ называется функцией риска оценки θ^* .

Равномерный подход: Оценка θ^* лучше $\hat{\theta}$, если $\forall \theta \in \Theta$ выполнено $R(\theta, \theta^*) \leq R(\theta, \hat{\theta})$ и для некоторого θ неравенство строгое.

Если $\theta^*(X)$ лучше всех других оценок $\hat{\theta} \in \mathcal{K}$ (\mathcal{K} – некоторый класс оценок, например, все несмещённые оценки или оценки, являющиеся функциями от первого члена вариационного ряда) $\implies \theta^*$ – наилучшая в классе \mathcal{K} .

Определение. Равномерный подход с квадратичной функцией потерь называется среднеквадратичным.

Задача. Пусть X_1, \dots, X_n – выборка из $R(0, \theta)$. Класс оценок $\mathcal{K} = \{cX_{(1)}, \text{ где } c \in \mathbb{R}\}$. Найдите наилучшую оценку в \mathcal{K} в среднеквадратичном подходе.

Решение.

$$E_{\theta}X_{(1)} = \frac{\theta}{n+1},$$

$$E_{\theta}X_{(1)}^2 = \int_0^{\theta} x^2 \frac{(\theta - x)^{n-1}n}{\theta^n} dx = n\theta^2 \mathbf{B}(3, n) = n\theta^2 \frac{\Gamma(n)\Gamma(3)}{\Gamma(n+3)} = \frac{n\theta^2 2(n-1)!}{(n+2)!} = \frac{2\theta^2}{(n+1)(n+2)},$$

поскольку плотность $X_{(1)}$ равна

$$p_{X_{(1)}}(x) = \frac{d}{dx}P(X_{(1)} > x) = \frac{d}{dx}P(\forall i X_i > x) = \frac{d}{dx} \left(\frac{\theta - x}{\theta} \right)^n = \frac{(\theta - x)^{n-1}n}{\theta^n}.$$

Теперь мы должны устремить функцию риска к минимуму: $\arg \min_c E_{\theta}(cX_{(1)} - \theta)^2 = ?$ Дифференцируя эту функцию по θ и приравнявая полученное к 0, находим, что $c = \frac{\theta E_{\theta}X_{(1)}}{E_{\theta}X_{(1)}^2} = \frac{n+2}{2}$. \square

Определение. Если $\theta \in \Theta \subset \mathbb{R}^k$ – многомерный параметр, то оценки $\theta^*(X)$ лучше $\hat{\theta}(X)$ в среднеквадратичном подходе, если $\forall \theta \in \Theta \quad \forall a \in \mathbb{R}^k$ выполнено $E_{\theta}(\langle \theta^* - \theta, a \rangle)^2 \leq E_{\theta}(\langle \hat{\theta} - \theta, a \rangle)^2$ и для каких-то a и θ неравенство строгое.

Эффективные оценки.

Пусть X – наблюдение с неизвестным распределением P_{θ} , $\theta \in \Theta$, причём для этого наблюдения имеется функция правдоподобия $f_{\theta}(X)$.

Определение. Величина $I_X(\theta) = E_{\theta} \left(\frac{\partial}{\partial \theta} \ln f_{\theta}(X) \right)^2$ называется количеством информации (по Фишеру), содержащемся в наблюдении X .

Задача. Пусть $X = (X_1, \dots, X_n)$ – выборка. Покажите, что $I_X(\theta) = ni(\theta)$, где $i(\theta)$ – количество информации в одном наблюдении X_i .

Решение.

Считаем по условию (это одно из условий регулярности, о котором речь пойдёт позднее и на лекциях), что мы можем менять дифференцирование и интеграл местами. Тогда

$$E_{\theta} \frac{\partial}{\partial \theta} \ln \underbrace{p_{\theta}(X_i)}_{\text{плотность } X_i} = \int_{\mathcal{X}} \left(\frac{\partial}{\partial \theta} p_{\theta}(x) \right) \frac{1}{p_{\theta}(x)} p_{\theta}(x) dx = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} p_{\theta}(x) dx = \frac{\partial}{\partial \theta} (1) = 0.$$

Отсюда

$$I_X(\theta) = D_{\theta} \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln p_{\theta}(X_i) \right) = |X_i \text{ независимы и одинаково распределены}| = \sum_{i=1}^n D_{\theta} \left(\frac{\partial}{\partial \theta} \ln p_{\theta}(X_i) \right) = ni(\theta). \quad \square$$

Теорема 1. (Неравенство Рао-Крамера)

Пусть выполнены условия регулярности (см. лекции), а $\theta^*(X)$ – несмещённая оценка для $\tau(\theta)$. Тогда для $\forall \theta \in \Theta$

$$D_{\theta}(\theta^*(X)) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}.$$

Определение. Если для оценки $\theta^*(X)$ выполнено равенство ($\forall \theta \in \Theta$) в неравенстве Рао-Крамера, то $\theta^*(X)$ называется эффективной оценкой $\tau(\theta)$.

Критерий эффективности. $\theta^*(X)$ – эффективная оценка для $\tau(\theta) \iff$ имеет место равенство $\theta^*(X) - \tau(\theta) = c(\theta) \frac{\partial}{\partial \theta} \ln f_{\theta}(X)$, где $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$.

Т.е. с помощью этого критерия можно сразу, без вычисления дисперсии, находить эффективные оценки, а также из формулы для $c(\theta)$ находить количество информации для данной оценки.

А смысл эффективности таков: эффективная оценка параметра $\tau(\theta)$ – наилучшая в среднеквадратичном подходе в классе несмещённых оценок $\tau(\theta)$.

Задача. Пусть выборка $X_1, \dots, X_n \sim \text{Bern}(\theta)$. Найти эффективную оценку θ и информацию одного наблюдения $i(\theta)$.

Решение.

Найдём функцию правдоподобия: $f_{\theta}(X_1, \dots, X_n) = \theta^{\sum X_i} (1 - \theta)^{n - \sum X_i}$. Тогда

$$\frac{\partial}{\partial \theta} \ln f_{\theta}(X_1, \dots, X_n) = \frac{n\bar{X}}{\theta} - \frac{n - n\bar{X}}{1 - \theta} = \frac{n}{\theta(1 - \theta)} (\bar{X} - \theta) = \frac{1}{c(\theta)} (\bar{X} - \theta),$$

т.е. $\frac{\partial}{\partial \theta} \ln f_{\theta}(X_1, \dots, X_n) = \frac{1}{c(\theta)} (\bar{X} - \theta)$, откуда по критерию эффективности получаем, что \bar{X} – эффективная оценка θ . С другой стороны, $\frac{\theta(1 - \theta)}{n} = c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$, где $I_X(\theta) = ni(\theta) \implies i(\theta) = \frac{1}{\theta(1 - \theta)}$. \square