На границе дозволенного

Информация Фишера. Эффективные оценки

1. Для выборки $\mathbf{X} = (X_1, \ldots, X_n)$ из распределения $\mathrm{Exp}(1/\theta)$ сравните оценки $\widehat{\theta}(\mathbf{X}) = \overline{\mathbf{X}}$ и $\widetilde{\theta}(\mathbf{X}) = nX_{(1)}$ в среднеквадратичном подходе.

2. Найдите $I_{\mathbf{X}}(p)$ для выборки $X_i \sim \mathrm{Geom}(p)$ и покажите, что $\overline{\mathbf{X}}$ является эффективной оценкой для $1/p$.

3. Из первой домашки мы знаем, что $\overline{I(\mathbf{X} = 0)}$ — несмещённая оценка функции $e^{-\theta}$ для выборки из $\mathrm{Pois}(\theta)$. Будет ли она эффективной?

4. Информация Фишера статистики $S(\mathbf{X})$ определяется как

$$I_{S(\mathbf{X})}(\theta) = \mathsf{D}_\theta \frac{\partial}{\partial \theta} \ln \rho_\theta(S(\mathbf{X})),$$

где ρ — плотность распределения $S(\mathbf{X})$ (в частности, для $S = \mathrm{Id}$ выражение выше равно просто информации Фишера). Для выборки $(X_1, \ldots, X_n)$ из $\mathcal{N}(0, \sigma^2)$ посчитайте $I_{S(\mathbf{X})}(\sigma^2)$ для статистик $S_1(\mathbf{X}) = \mathbf{X}$, $S_2(\mathbf{X}) = \overline{\mathbf{X}}$, $S_3(\mathbf{X}) = \overline{\mathbf{X}^2}$. Что можно заметить?

5. Докажите, что если в дополнение к условиям регулярности добавить равенство

$$\int \rho_\theta''(\mathbf{x})\, d\mathbf{x} = 0$$

(в частности, это верно, если интеграл плотности дважды дифференцируем по параметру), то информацию Фишера можно представить как

$$I_{\mathbf{X}}(\theta) = -\mathsf{E}_\theta \frac{\partial^2 \ln \rho_\theta(\mathbf{X})}{\partial \theta^2}.$$

6.* Имеется выборка, состоящая из одного вектора

$$\mathbf{X} = (X_1, \ldots, X_k) \sim \mathcal{N}(\boldsymbol{\theta}, E_k),$$

где $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ — неизвестный вектор параметров, а $k \geq 3$. Докажите, что смещённая *оценка Штейна*

$$\widehat{\boldsymbol{\theta}}(\mathbf{X}) = \mathbf{X} - \frac{k-2}{\|\mathbf{X}\|^2} \cdot \mathbf{X}$$

строго лучше несмещённой оценки $\boldsymbol{\theta}^*(\mathbf{X}) = \mathbf{X}$ в среднеквадратичном подходе, то есть $\forall \boldsymbol{\theta} \in \mathbb{R}^k\colon \mathsf{E}_{\boldsymbol{\theta}} \|\widehat{\boldsymbol{\theta}}(\mathbf{X}) - \boldsymbol{\theta}\|^2 < \mathsf{E}_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^*(\mathbf{X}) - \boldsymbol{\theta}\|^2$.