На границе дозволенного

Информация Фишера. Эффективные оценки

Чертоги разума

Цель. Найти оценку с как можно меньшей *среднеквадратичной ошибкой (MSE):*

$$\mathsf{E}_\theta\left(\widehat{\theta}-\theta\right)^2 = \mathsf{D}_\theta\widehat{\theta} + \left(\mathsf{E}_\theta\widehat{\theta}-\theta\right)^2.$$

Будем впредь считать, что в нашей модели выполнены **условия регулярности.**

- Множество $\Theta \subset \mathbb{R}^k$ открыто и связно;
- Множество $\{\mathsf{P}_{\boldsymbol{\theta}}\}$ доминируемо по мере μ с плотностью $\rho_{\boldsymbol{\theta}}$, причём её носитель $\{x\colon \rho_{\boldsymbol{\theta}}(x) > 0\}$ не зависит от $\boldsymbol{\theta}$;
- Для всякой статистики $T(\mathbf{X})$ с ограниченной $f(\boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}}T^2(\mathbf{X})$ её матожидание можно дифференцировать по параметру:

$$\nabla_{\boldsymbol{\theta}}\int_{\mathbb{R}^n} T(\mathbf{x})\rho_{\boldsymbol{\theta}}(\mathbf{x})\,d\mathbf{x} = \int_{\mathbb{R}^n} T(\mathbf{x})\cdot\nabla_{\boldsymbol{\theta}}\rho_{\boldsymbol{\theta}}(\mathbf{x})\,d\mathbf{x}.$$

- Матрица

$$I_{\mathbf{X}}(\boldsymbol{\theta})_{ij} = \mathsf{E}_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial\theta_i}\ln\rho_{\boldsymbol{\theta}}(\mathbf{X})\cdot\frac{\partial}{\partial\theta_j}\ln\rho_{\boldsymbol{\theta}}(\mathbf{X})\right)$$

  существует, конечна и положительно определена. Её называют *информационной матрицей Фишера* (или просто информация Фишера).

Неравенство Рао-Крамера на дисперсию несмещённой оценки для $\tau(\theta)$.

$$\mathsf{E}_\theta\theta^*(\mathbf{X}) = \tau(\theta) \Longrightarrow \forall\theta\in\Theta\colon \mathsf{D}_\theta\theta^* \geqslant \frac{(\tau'(\theta))^2}{I_{\mathbf{X}}(\theta)},\ \text{где } I_{\mathbf{X}}(\theta) = \mathsf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln\rho_\theta(X)\right)^2$$

$$\left(\begin{matrix}\text{в многомерном}\\ \text{случае}\end{matrix}\colon \mathsf{D}_{\boldsymbol{\theta}}\boldsymbol{\theta}^* \succcurlyeq \nabla_{\boldsymbol{\theta}}\tau(\boldsymbol{\theta})\cdot I_{\mathbf{X}}(\boldsymbol{\theta})^{-1}\cdot\nabla_{\boldsymbol{\theta}}\tau(\boldsymbol{\theta})^\top\right)$$

Оценки, для которых неравенство выше обращается в равенство, называют *эффективными.*
Критерий эффективности.

$$\text{Эффективность} \iff \boldsymbol{\theta}^*(\mathbf{X}) - \tau(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\tau(\boldsymbol{\theta})\cdot I^{-1}_{\mathbf{X}}(\boldsymbol{\theta})\cdot\nabla_{\boldsymbol{\theta}}(\ln\rho_{\boldsymbol{\theta}}(\mathbf{X})).$$

Теорема. Пусть $\theta \subset \mathbb{R}$. Тогда эффективная оценка существует только для так называемых *экспоненциальных семейств*, то есть когда плотность имеет вид

$$\rho_\theta(x) = h(x)\cdot\exp\left(a(\theta) + b(\theta)T(x)\right),$$

Более того, в этом случае $\theta^*(\mathbf{X}) = \overline{T(\mathbf{X})}$ является эффективной оценкой для $-\frac{a'(\theta)}{b'(\theta)}$, а любая другая эффективная оценка будет линейной функцией от $\theta^*(\mathbf{X})$.

Problemas

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. (а) Покажите, что информация Фишера может быть найдена как ковариационная матрица градиента логправдоподобия:

$$I_{\mathbf{X}}(\boldsymbol{\theta}) = \mathsf{D}_\theta \nabla_{\boldsymbol{\theta}} \ln \rho_{\boldsymbol{\theta}}(\mathbf{X});$$

(б) Покажите, что $I_{\mathbf{X}}(\boldsymbol{\theta}) = nI_{X_1}(\boldsymbol{\theta})$.

3. Найдите $I_{\mathbf{X}}(p)$ для выборки $(X_1, \ldots, X_n)$ из $\mathrm{Geom}(p)$. Для каких функций $\tau(p)$ существует эффективная оценка?

4. В модели категориального распределения $\mathrm{Cat}(p_1, p_2, 1 - p_1 - p_2)$ посчитайте информационную матрицу Фишера для вектора параметров (p_1, p_2). Проверьте, эффективна ли ОМП, найденная в предыдущем листике.

5. 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_θ — 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?

6. 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

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