University of Lille 2023-2024
M1 Data Science - Stats 2 March 13th, 2024

TD2 - Efficient estimators, Exponential families

Exercise 1 In the Gaussian model

$$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma_0^2)$$

where σ_0^2 is known, we recall that the MLE is $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

1. Compute the Fisher information of the model, $I_n(\theta)$.
2. Show that the MLE is an efficient estimator.
3. Prove that the family of Gaussian distribution of variance σ_0^2

$$\mathcal{P}_{\sigma_0^2} = \left\{\mathcal{N}(\mu, \sigma_0^2), \mu \in \mathbb{R}\right\}$$

forms an exponential family.

Exercice 2 Same questions for the Poisson model $X_1, \dots, X_n \sim \mathcal{P}(\lambda)$. We recall that

$$\mathbb{P}(X_1 = k) = \frac{\lambda^k}{k!}e^{-\lambda} \text{ for all } k \in \mathbb{N}.$$

Exercise 3 We are given a i.i.d sample $X = (X_1, \dots, X_n) \sim f_\theta$ where $\theta > 0$ is an unknown parameter, and

$$f_\theta(x) = \frac{1}{2\theta}\exp\left(-\frac{|x|}{\theta}\right), \quad x \in \mathbb{R}.$$

1. Show that $\mathcal{P} = \{f_\theta, \theta \in \mathbb{R}^+\}$ forms an exponential family. What is the canonical statistic?
2. Find the MLE $\widehat{\theta}_n$ of θ. Is it unbiased?
3. Find Var($\widehat{\theta}_n$). Then using the CLT find the asymptotic distribution of $\widehat{\theta}_n$.
4. Find the Fisher information $I^X(\theta)$, and use it to obtain the Cramer-Rao lower bound. Is $\widehat{\theta}_n$ an efficient estimator of θ?

Exercise 4 Set $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(\theta)$, with $\theta \in (0,1)$, and $g(\theta) = \theta(1-\theta)$.

1. Letting $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, justify that a plug-in estimator of $g(\theta)$ is $\widehat{g}_n = \overline{X}_n(1 - \overline{X}_n)$.
2. Compute the biais of $\widehat{g}_n$ and prove that $\widetilde{g}_n = \frac{n}{n-1}\widehat{g}_n$ is unbiaised.
3. Prove that the minimal variance of an unbiased estimators of $g(\theta)$ is $\frac{\phi(\theta)}{n}$, where

$$\phi(\theta) = \theta(1-\theta)(1-2\theta)^2.$$

4. We admit that

$$\mathrm{Var}_\theta[\widetilde{g}_n] - \frac{\phi(\theta)}{n} = \frac{2\theta^2(1-\theta)^2}{n(n-1)}.$$

   Is $\widetilde{g}_n$ an efficient estimator?

5. Using the Δ-method, compute the asymptotic distribution of $\widetilde{g}_n$. Comment.

Exercise 5 Let $\{P_\theta \mid \theta \in \Theta\}$ be an exponential family where the density of P_θ with respect to the Lebesgue measure in $\mathbb{R}$ is

$$f_\theta(x) = e^{a(\theta)T(x)-b(\theta)},$$

where $a : \Theta \to \mathbb{R}$, $b : \Theta \to \mathbb{R}$ and $T : \mathbb{R} \to \mathbb{R}$ are three real-valued functions and $\Theta \subset \mathbb{R}$. We assume that these functions are regular enough for the model to be regular. In particular, a and b are differentiable and we write $a'(\theta)$ and $b'(\theta)$ their derivative.

We observe a n sample $X = (X_1, \dots, X_n) \overset{iid}{\sim} f_\theta$ and seek to estimate $g(\theta) = \mathbb{E}_\theta[T(X_1)]$. We denote by $I_n(\theta)$ the Fisher information of the observation X and by $I(\theta)$ the Fisher information of X_1.

1. Prove that the score function satisfies

$$s(X_1;\theta) = a'(\theta)T(X_1) - b'(\theta)\ .$$

   Deduce that

$$a'(\theta)g(\theta) = b'(\theta).$$

2. Prove that

$$I(\theta) = (a'(\theta))^2\mathrm{Var}_\theta[T(X_1)].$$

3. Show that

$$g'(\theta) = a'(\theta)\mathbb{E}_\theta\left[(T(X_1))^2\right] - b'(\theta)\cdot g(\theta),$$

   and deduce that

$$g'(\theta)^2 = (a'(\theta))^2 \cdot (\mathrm{Var}_\theta\left[T(X_1)\right])^2\ .$$

   Hint. Write $g(\theta)$ as in integral and swap derivation and integration.

4. Write down the minimal variance of an unbiased estimator of $g(\theta)$ based on the observation X.

5. Propose an efficient estimator of $g(\theta)$.