University of Lille 2025-2026
M1 Data Science - Stats 2

TD2 - Efficient estimators, Exponential families

Exercise 1 In the Gaussian model

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

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 $X_1, \ldots, X_n \sim \mathcal{P}(\lambda)$. 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 $X = (X_1, \ldots, X_n) \sim f_\theta$ 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 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(\theta)$, 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 (exam 2025) A Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $\Gamma(\alpha, \beta)$, has a density with respect to the Lebesgue measure given by

$$\forall x \in \mathbb{R}, \quad f_{\alpha,\beta}(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha m(\alpha)} \mathbb{1}_{[0,+\infty[}(x)$$

where $m(\alpha)$ is some normalizing constant. We observe $X_1, \ldots, X_n$, n iid realizations from $\Gamma(\alpha, \beta)$.

1. Prove that $\mathbb{E}_{\alpha,\beta}[X_1] = \alpha\beta$.
2. We admit that $\mathrm{Var}_{\alpha,\beta}[X_1] = \alpha\beta^2$. Deduce a moment estimator of the parameter (α, β).

In the rest of the exercise, we assume that $\alpha = 2$, which reduces the model for the distribution of X_i to the set of densities

$$f_\beta(x) = \frac{xe^{-\frac{x}{\beta}}}{\beta^2} \mathbb{1}_{[0,+\infty[}(x) \quad \text{for} \quad \beta > 0.$$

We are interested in estimating the variance $\sigma^2 = 2\beta^2$.

3. Prove that the maximum likelihood estimator of β is $\widehat{\beta}_n = \frac{1}{2n}\sum_{i=1}^n X_i$.
4. Deduce a consistent estimator of σ^2.
5. Using the Delta method, derive the asymptotic distribution of σ^2.
6. Is this estimator asymptotically efficient?
7. Propose a test of $(\sigma^2 = 1)$ against $(\sigma^2 > 1)$ that is asymptotically of level 0.05.