University of Lille 2024-2025
M1 Data Science - Stats 2

TD3 - Asymptotic properties of estimators

Exercise 1 We are given two asymptotically normal estimators of the same parameter $\theta \in \Theta \subseteq \mathbb{R}$, denoted by $\widehat{\theta}_n^1$ and $\widehat{\theta}_n^2$, which satisfy, for all $\theta \in \Theta$,

$$\begin{aligned} \sqrt{n}\left(\widehat{\theta}_n^1 - \theta\right) &\rightsquigarrow \mathcal{N}(0, \sigma_1^2) \\ \sqrt{n}\left(\widehat{\theta}_n^2 - \theta\right) &\rightsquigarrow \mathcal{N}(0, \sigma_2^2) \end{aligned}$$

where $\sigma_1^2 < \sigma_2^2$.

1. Using each estimator, derive an asymptotic confidence interval on θ of level 0.95.
2. For any $\varepsilon > 0$, how many samples n are needed to get a precision ε on θ with the estimator $\widehat{\theta}_n^2$, with probability $\simeq 0.95$?
3. What fraction of this sample size is needed if we choose to use instead the estimator $\widehat{\theta}_n^1$?

Exercise 2 We are given an i.i.d sample $X_1, \ldots, X_n \sim f_\theta$ where $\theta > 0$ is an unknown parameter and

$$f_\theta(x) = \frac{1}{2}\theta^3 x^2 \exp(-\theta x); \quad x > 0.$$

Denote μ, σ^2 to be the mean and variance of f_θ respectively.

1. Show that $\mu = 3/\theta$ and $\sigma^2 = 3/\theta^2$.
2. Find the Cramer Rao bound for μ and show that the estimator $\widehat{\mu}_n = \frac{1}{n}\sum_i X_i$ achieves this bound.
3. Now we consider the estimator $\widehat{\sigma}_n^2 = \frac{\widehat{\mu}_n^2}{3}$ of σ^2.
   (a) Compute the bias of $\widehat{\sigma}_n^2$ and conclude that it is asymptotically unbiased.
   (b) Show that $\widehat{\sigma}_n^2$ is a consistent estimator of σ^2.
   (c) Show that $\widehat{\sigma}_n^2$ is asymptotically efficient.

Exercise 3 (exam 2024) We perform a survey independently on n individuals in a population. The outcome of the survey can be summarized by a score X_i (that can be either positive or negative) observed for each individual. We define by p the probability that an individual has a positive experience. Letting

$$Y_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \le 0 \end{cases}$$

we have $p = \mathbb{P}(Y_i = 1)$. We assume that the individual's scores $X_1, \ldots, X_n$ are iid from a $\mathcal{N}(\theta, 1)$ distribution, whose density is given by

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right), \quad \text{for all } x \in \mathbb{R}.$$

We recall that the maximum likelihood estimator of θ is given by $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

1. Compute the biais, the variance and the mean square error of $\widehat{\theta}_n$.
2. Prove that $p = \Phi(\theta)$ where Φ denotes the cdf of a standard normal distribution.

   Deduce a consistent estimator of p, denoted by $\widehat{p}_n$.
3. Using the Delta method, derive the asymptotic distribution of $\widehat{p}_n$.
4. We define $\widetilde{p}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Show that $\widetilde{p}_n$ is an unbiased estimator of p. What is its variance?
5. What is the asymptotic distribution of $\widetilde{p}_n$? Compare its asymptotic variance to that of $\widehat{p}_n$.
6. Now assume that the (X_i) are not really Gaussian but come from another distribution F with mean μ. Is $\widehat{p}_n$ still consistent? Same question for $\widetilde{p}_n$.
7. How would you check in practise that that (X_i) being Gaussian is a reasonnable assumption?

Exercise 4 (exam 2024) $X_1, \ldots, X_n$ are iid from a distribution whose density is

$$f_a(x) = \frac{a}{x^{a+1}}\mathbb{1}_{[1,+\infty[}(x)$$

for some parameter $a > 0$.

1. Justify that the expectation of X_1 is finite if and only if $a > 1$.
2. Prove that the family of distributions $\{f_a, a > 0\}$ forms an exponential family.

   What is its canonical statistic?
3. Prove that the maximum likelihood estimator of the parameter a is $\widehat{a}_n = \frac{n}{\sum_{i=1}^n \log(X_i)}$
4. Let $Y = \log(X_1)$. Prove that its cdf is given by $F_Y(t) = (1 - e^{-at})$ for all $t > 0$.

We recognize an exponential distribution of parameter a, for which $\mathbb{E}[Y] = \frac{1}{a}$ and $\mathrm{Var}[Y] = \frac{1}{a^2}$.

5. Using the Delta method, find the asymptotic distribution of $\widehat{a}_n$.
6. Prove that this estimator is asymptotically efficient.
7. Construct an asymptotic test of level α for $\mathcal{H}_0 : (a \leq 1)$ versus $\mathcal{H}_1 : (a > 1)$.