University of Lille | 2023-2024
M1 Data Science - Stats 2 | March 5th, 2024

TD1 - Estimation

Exercise 1 Let $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma_0^2)$ be i.i.d Gaussian random variables for an unknown $\mu \in \mathbb{R}$ but known σ_0.

1. Derive the maximum likelihood estimator $\widetilde{\mu}_n$ of μ.
2. Using $\widetilde{\mu}_n$, derive a confidence interval for μ at the confidence level $1 - \alpha$.

Exercise 2 A Poisson distribution with parameter $\lambda > 0$, denoted by $\mathcal{P}(\lambda)$, is a discrete distribution supported on $\mathbb{N}$ defined as

$$\mathbb{P}_{Z \sim \mathcal{P}(\lambda)}(Z = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

1. Compute the maximum likelihood estimator of λ given iid observations $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{P}(\lambda)$.
2. What other method(s) could you use to obtain the same estimator?
3. Compute its biais and its mean square error.

Exercise 3 Given iid samples $X_1 \dots, X_n$ from some distribution P in $\mathbb{R}$ whose cdf is F, we want to estimate the function F.

1. What is the underlying statistical model?
2. Estimating the function means estimating for each $x \in \mathbb{R}$, the value of $F(x)$. Justify that

   $$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

   is an unbiased estimator of this quantity. Compute its variance and mean-square error.
3. Using the central limit theorem, what is the asymptotic distribution of $\widehat{F}_n(x)$?

Exercise 4 Let $(X_1, \dots, X_n)$ be a n-sample drawn from the uniform distribution $\mathcal{U}(0, \theta)$, for an unknown parameter value $\theta > 0$.

1. Calculate $E_\theta[X_1]$ and deduce the moment estimator $\widehat{\theta}_n$ of θ.
2. Calculate the maximum likelihood estimator $\widetilde{\theta}_n$ of θ.
3. Are these estimators biased?
4. Compare the quadratic risks of these estimators.

Exercise 5 Let $(X_1, \dots, X_n)$ be a n-sample drawn from the uniform distribution over $[\theta - 1/2; \theta + 1/2]$, where $\theta \in \mathbb{R}$ is unknown. What is the MLE of θ?

Exercise 6 We are given an i.i.d sample $X_1, \dots, X_n \sim f_\theta$ where $\theta > 0$ is an unknown parameter, and

$$f_\theta(x) = \frac{2\theta^2}{x^3} \mathbb{1}_{[\theta, +\infty[}(x).$$

1. Show that f_θ is a density and find $\mathbb{E}_\theta[X]$.
2. Using the moment method find an unbiased estimator $\widetilde{\theta}_n$ of θ.
3. Show that the MLE $\widehat{\theta}_n$ is given by $\widehat{\theta}_n = \min_i X_i$. Evaluate the bias of $\widehat{\theta}_n$.
4. Obtain a confidence interval for θ using $\widehat{\theta}_n$ with confidence level $1 - \alpha$.
5. Using the above confidence interval, build a statistical test at level α for

   $$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta \neq \theta_0$$

   for $\theta_0 > 0$.

Exercise 7 (variance estimators) We let $\mathcal{P}(\mathbb{R})$ denote the set of probability distributions over $\mathbb{R}$. We consider the class of models $X_1, \dots, X_n \overset{iid}{\sim} P$ where P belongs to the set

$$\mathcal{M} = \{P \in \mathcal{P}(\mathbb{R}) : \mathbb{E}_{Z \sim P}[Z^2] < \infty\}$$

In particular, all distributions in $\mathcal{M}$ have a finite variance and we recall the expression of the variance of a random variable Z

$$\mathrm{Var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \mathbb{E}[Z] - (\mathbb{E}[Z])^2.$$

1. What is the parameter of this statistical model? Is it a parametric model?
2. Justify that the (unadjusted) empirical variance estimate

   $$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{\mu}_n)^2 \quad \text{where } \widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

   can also be recovered with the plug-in method.
3. Prove that this estimator satisfies $\mathbb{E}[\widehat{\sigma}_n^2] = \frac{n-1}{n} \sigma^2$.
4. Deduce the expression of the adjusted variance estimator $\widetilde{\sigma}_n^2$, which is proportional to $\widehat{\sigma}_n$ and is unbiased.
5. Computing the variances of these estimator can be complex if we make no further assumptions on the distributions. If we consider instead the parameteric model

   $$\mathcal{M} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

   we can do it and we get $\mathrm{Var}[\widetilde{\sigma}_n^2] = \frac{2\sigma^4}{n-1}$. Deduce the variance of $\widehat{\sigma}_n^2$.
6. Which estimator has the smallest mean square error?