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

Final Exam

Exercise 1 We collect iid data $X_1, \dots, X_n$ from a uniform distribution $\mathcal{U}([\theta, 1])$ for some $\theta \in [0,1]$. We recall that the density of a uniform distribution over an interval $[a,b]$ is given by

$$\forall x \in \mathbb{R}, \ \ f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x) \ .$$

1. Compute $\mathbb{E}_\theta[X_1]$.
   Show that the moment estimator is given by $\widehat{\theta}_n = \frac{2}{n}\sum_{i=1}^n X_i - 1$.
2. Compute its biais, its variance and its mean square error.
   Hint. The variance of a uniform distribution over $[a,b]$ *is* $\frac{(b-a)^2}{12}$.
3. Prove that the maximum likelihood estimator of θ is $\widetilde{\theta}_n = \min_{i \in \{1,\dots,n\}} X_i$.
4. Compute $\mathbb{P}(\widetilde{\theta}_n > t)$ for all $t \in \mathbb{R}$ and deduce the density of $\widetilde{\theta}_n$.
5. Prove that $\widetilde{\theta}_n$ is a consistent estimator of θ.
6. Compute the mean square error of $\widetilde{\theta}_n$ and compare it to that of $\widehat{\theta}_n$. Which estimator is better?

Exercise 2 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 $X_1, \dots, X_n$, 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

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

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.

Exercise 3 We collect an iid sample $X_1, \dots, X_n$ distributed under a Binomial distribution with parameters $m \in \mathbb{N}\backslash\{0\}$ and $p \in [0,1]$ for which

$$\mathbb{P}(X_i = k) = \binom{m}{k} p^k (1-p)^{m-k}, \quad k \in \{0, \dots, m\} .$$

We assume that the parameter $m > 0$ is known.

1. Prove that the family of distributions $\{B(m,p), p \in [0,1]\}$ forms an exponential family.
2. Let p_0, p_1 be two values $(0,1)$ such that $p_1 < p_0$. Give the form of the Likelihood Ratio Test of

$$\mathcal{H}_0 : (p = p_0) \text{ against } \mathcal{H}_1 : (p = p_1)$$

3. Explain to which extend this test can be considered optimal.
4. Illustrate graphically the distribution of the test statistic $T(X)$ used in 2. under $\mathcal{H}_0$ and under $\mathcal{H}_1$. Represent graphically the type I error and the power for a given threshold.

   Plausible distributions are enough.

Exercise 4 We consider a regression model in which we observe iid pairs $(X_i, Y_i)_{1 \leq i \leq n}$ with $X_i = (X_{i,1}, X_{i,2}) \in \mathbb{R}^2$ and $Y_i \in \mathbb{R}$. The distribution of (X_i, Y_i) is the following:

$$\begin{aligned} X_i &\sim P_X \\ Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

where P_X is some distribution on $\mathbb{R}^2$. We assume that the variance σ^2 is known and we are interested in the parameter $\beta = (\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3$.

1. We recall the the density of a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ is given by

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) .$$

   Give the expression of $\widehat{\beta}_n$, the maximum likelihood estimator of β.

   A closed-form solution is not needed, as long as you can write it as the solution to some optimization problem, made as simple as possible.

2. We are interested in testing the significativity of the variable $X_{i,2}$ by testing

$$\mathcal{H}_0 : (\beta_2 = 0) \quad \text{against} \quad \mathcal{H}_1 : (\beta_2 \neq 0) .$$

   Propose a test that is asymptotically of level α.