MATH-UA.0234 Mathematical Statistics Spring 2022
Dr. M. Nitzschner

Problem Set 6

Submission:
Thursday, 04/07/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **UMVU estimators** [4 Points]

   (a) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Ber(\theta)$, with unknown $\theta \in (0,1)$. We are looking for the UMVU for $f(\theta) = \theta^3$.

   (i) Verify that
   $$\widehat{\gamma}(X_1, ..., X_n) = \mathbb{1}_{\{X_1 = X_2 = X_3 = 1\}}$$
   is an unbiased estimator for $f(\theta)$.

   (ii) Explain why $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$ is a sufficient and complete statistic for θ.
   Hint: You can show completeness either directly or use the statement about exponential families from the notes.

   (iii) Calculate the UMVU estimator for $f(\theta)$, using the Lehmann-Scheffé theorem.

   (b) Now let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pois(\theta)$, with unknown $\theta > 0$. An unbiased estimator for $f(\theta) = e^{-\theta}$ is given by $\widehat{\gamma}(X_1, ..., X_n) = \mathbb{1}_{\{X_1 = 0\}}$, and a sufficient and complete statistic for θ is given by $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$. Use the Lehmann-Scheffé theorem to determine the UMVU estimator for $f(\theta)$.

2. **The Cramér-Rao lower bound I** [4 Points]

   Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Bin(N, p)$ with *known* $N \in \mathbb{N}$ and unknown $p \in (0,1)$.

   (a) Calculate the Maximum-likelihood estimator (MLE) $\widehat{p}_n$ for p.

   (b) Determine the Fisher information $I(p)$ in p.

   (c) Check whether the variance of the MLE in (a) saturates the Cramér-Rao lower bound.

3. **The Cramér-Rao lower bound II** [4 Points]

   Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with unknown $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

   (a) Determine the Fisher information matrix $I(\mu, \sigma^2)$ in (μ, σ^2).

   (b) Check whether the covariance matrix of the MLE $(\widehat{\mu}_n, S_n^2)$ for (μ, σ^2) saturates the Cramér-Rao lower bound.
   Hint: The MLE was calculated in Problem 1 (a) of Problem set 5.

4. **(R exercise) Maximum-Likelihood estimation for an autoregressive process** [4 Points]

   Note: Please provide your source code and images obtained with your solution.

   A *Gaussian autoregressive process* (more precisely a Gaussian AR(1) process) can be defined by considering random variables $X_0, X_1, X_2, ...$ with
   $$X_i = \theta X_{i-1} + \varepsilon_i, \text{ for } i = 1, 2, ... \qquad X_0 = 0.$$
   Here $\varepsilon_1, \varepsilon_2, ...$ are i.i.d. $\mathcal{N}(0, \sigma^2)$-distributed random variables. We will assume that $\sigma^2 > 0$ is *known* and $\theta \in \mathbb{R}$ is unknown.

(a) The joint probability density function of $X_1, ..., X_n$ is given by

$$f_{X_1,...,X_n;\theta}(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i - \theta x_{i-1}),$$

where $x_0 := 0$ and f is the density of a $\mathcal{N}(0, \sigma^2)$-distribution. Determine the Maximum-likelihood estimator $\widehat{\theta}_n$ (MLE) for θ.

(b) Simulate an autoregressive process with $\sigma^2 = 1$ and $\theta = 0.5$ with $n = 100$ time steps using R, and plot the outcome.

(c) Plot a histogram of $\widehat{\theta}_k$ for $k = 10$, $k = 100$ and $k = 1000$, $k = 5000$, each with $N = 500$ repetitions. What do you observe?