7 Likelihood and Maximum Likelihood Estimation

Exercice 7.1. Let X be a random variable admitting the following probability density :

$$f_X(x;\theta) = \frac{\theta}{x^{\theta+1}} \mathbb{I}_{x\geq 1}$$

where $\theta > 0$. It is in fact a particular Pareto law. Consider a sample of size n : $X_1, \ldots, X_n$ iid from X.

1 Show that the model belongs to the exponential family and deduce that is *regular*.

2 Compute the Fisher information contained in X for the parameter θ. Deduce the information contained in the n-sample.

3 Actually we do not observe X, but a random variable Y defined by :

$$Y = \begin{cases} 1 & \text{if} \quad X \geq \exp(1) \\ 0 & \text{if} \quad X < \exp(1) \end{cases} .$$

Compute the Fisher information brought by Y for the parameter θ.

4 Show that $I_X(\theta) > I_Y(\theta)$ (we can use the fact that $\exp(x) - 2x > 0$ for every $x \geq 0$).

Exercice 7.2. We consider a random variable X following an exponential law with parameter $\theta > 0$:

$$f_X(x;\theta) = \theta \exp(-\theta x) \mathbb{I}_{x>0} .$$

Let $X_1, \ldots, X_n$ be an n-sample from X.

1 Show that $\sum_{i=1}^n X_i$ is a sufficient and minimal statistics for θ.

2 Admit that $g(\underline{X}) = \frac{X_n}{\sum_{i=1}^n X_i}$ is an ancillary statistics for θ.
Compute $\mathbb{E}(g(\underline{X}))$.

Exercice 7.3. Consider a random variable X with density

$$f_X(x;\theta) = kx^\theta \mathbb{I}_{]0,1]}(x) .$$

Let $X_1, \ldots, X_n$ be an n-sample from X.

1 Determine k as a function of θ, specifying the conditions on θ. Compute $\mathbb{E}_\theta(X)$ and $\mathbb{V}_\theta(X)$.

2 Show that $-\log(X)$ follows a known law and specify its parameter. Deduce that $\tilde{\theta}_n = -1-(n-1)\left(\sum_{i=1}^{n}\log(X_i)\right)^{-1}$ is an unbiased estimator for θ.

Exercice 7.4. Let X be a random variable admitting the following probability density :

$$f_X(x;\theta) = \frac{\theta\exp(\theta x)}{\exp(\theta^2)-1}\mathbb{I}_{[0,\theta]}(x)$$

where $\theta > 0$. Consider an n-sample $X_1, \ldots, X_n$ from X. Find, if it exists, a sufficient statistics for θ.

Exercice 7.5. Let X be a random variable with geometric law with parameter $p \in]0,1[$ (number of trials before the first success, p success probability) and $X_1, \ldots, X_n$ be an n-sample from X. Assume $p = 1-q$.

1 Compute the Fisher information brought by X on p and the one contained in the n-sample.

2 Show that $\overline{X_n}$ is sufficient and that $\overline{X_n}$ is an effective estimator for the parameter q/p.

Exercice 7.6. Let X be a Poisson random variable with parameter $\theta > 0$ and $X_1, \ldots, X_n$ an n-sample from X.

1 Show that $\overline{X_n}$ and S_{n-1}^2 are unbiased estimators of θ.

2 Show that $\overline{X_n}$ is the uniformly minimum-variance unbiased estimator of θ. Deduce that $\mathbb{V}_\theta(\overline{X_n}) \leq \mathbb{V}_\theta(S_{n-1}^2)$.

Exercice 7.7. Let X be a random variable in the set of the reals, admitting the density

$$f_X(x;\theta) = k\exp(-\theta|x|)$$

where $\theta > 0$ and $X_1, \ldots, X_n$ an n-sample from X.

1 Determine the constant k.

2 Compute the integrals $\int_{-\infty}^{0} t\exp(\theta t)dt$ and $\int_{0}^{+\infty} t\exp(-\theta t)dt$ then deduce the expressions of $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

3 Compute the estimator W_n of θ via the maximum likelihood method.

4 Show that W_n is convergent estimator for θ in quadratic mean.

Exercice 7.8. Consider an n-sample $(X_1, \ldots, X_n)$ iid from X, with density :

$$f_X(x;\theta) = \frac{3}{(x-\theta)^4}\mathbb{I}_{[\theta+1,+\infty[}\,.$$

where $\theta > 0$ is an unknown parameter.

1 Compute $\mathbb{E}_\theta(X)$ and $\mathbb{V}_\theta(X)$ (we can compute $\mathbb{E}_\theta((X-\theta))$ and $\mathbb{V}_\theta((X-\theta))$).

2 Give the maximum likelihood estimator $\hat{\theta}_n$ for θ.

3 Compute $\mathbb{E}_\theta\left(\hat{\theta}_n\right)$. Deduce an unbiased estimator $\theta_n^{\#}$ for θ as a function of $\hat{\theta}_n$.

Exercice 7.9. Let X be a random variable, distributed according to a Pareto law with parameters $\alpha > 0$ and $\beta > 0$:

$$f_X(x;\alpha,\beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}\mathbb{I}_{x\geq\beta}\,.$$

Let $X_1, \ldots, X_n$ be an n-sample from X. In the following $\theta = (\alpha, \beta)$.

1 Give the density of X, then compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$ giving the conditions on the existence on those moments.

2 Suppose β known.

a) Write the likelihood on the sample and give a sufficient statistics for α.

b) Find an estimator T_n of α with the maximum likelihood method.

c) Find the law of the random variable $Y = \log(X/\beta)$.

d) Show that T_n is a strongly consistent estimator of α.

e) Find the law of $Z_n = \sum_{i=1}^n \log(X_i/\beta)$. Deduce the expression of $\mathbb{E}(T_n)$ and $\mathbb{V}(T_n)$, then show that T_n is a quadratic mean convergent estimator for α.

f) Deduce from T_n an unbiased estimator T_n^* for α. Show that T_n^* is asymptotically efficient.

3 Suppose α known.

a) Find an estimator W_n of β via maximum likelihood.

b) Find the law of W_n and deduce that W_n is a quadratic mean convergent estimator of β.

Exercice 7.10. Consider a random variable X with density

$$f_X(x;\theta) = k|x|\exp\left(-\frac{x^2}{2\theta}\right)$$

where $\theta > 0$. Let $X_1, \ldots, X_n$ be as usual an n-sample from X.

1 Compute the normalising constant k.

2 Compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

3 Give the maximum likelihood estimator $\hat{\theta}_n$ of θ. Is it unbiased ?
Is it strongly consistent ?

5 Explain why the model for X is regular.

6 Compute the Fisher information given by X on θ, then the one given by the whole sample.

7 Compute $\mathbb{E}_\theta(X^4)$. $\hat{\theta}_n$ is efficient ? (Cramer-Rao bound)

8 Is $\hat{\theta}_n$ the unique unbiased estimator with uniformly minimum variance ?

Exercice 7.11. Let X be a real random variable with density
$f_X(x;\theta) = \frac{2\sqrt{\theta}}{\sqrt{\pi}}\exp(-\theta x^2)\mathbb{I}_{]0,+\infty[}(x)$ where $\theta > 0$.
Let $X_1, \ldots, X_n$ then be an n-sample of X.

1 Compute $\mathbb{E}_\theta(X)$, $\mathbb{E}_\theta(X^2)$ and $\mathbb{V}_\theta(X)$.

2 Find W_n, the maximum likelihood estimator for θ.

Exercice 7.12. Let X be a discrete random variable with values in $\{-1, 0, 1\}$ such that $\mathbb{P}(X = 0) = 1 - 2\theta$ et $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \theta$. Suppose $\theta \in [0, 1/2]$. Consider $X_1, \ldots, X_n$ an n-sample from X.

Name R the random variable euql to the number of X_i with a non-null value. Find the maximum likelihood estimator W_n of θ. Give the law of R and deduce $\mathbb{E}(W_n)$ and $\mathbb{V}(W_n)$.

Exercice 7.13. Let X be a random variable with values in $[-1, 1]$ with density $f_X(x; a, b) = a\mathbb{I}_{[-1,0]}(x) + b\mathbb{I}_{]0,1]}(x)$ where $a \leq 0$ and $b \leq 0$ and $X_1, \ldots, X_n$ is an n-sample from X.

1 Point out the relation between a and b.
in the following, we will write b as a function of a.

2 Compute $\mathbb{E}_a(X)$ and $\mathbb{V}_a(X)$.

3 Find the maximum likelihood estimator W_n of θ.

Exercice 7.14. Infection Markers

N infectious agents aggress simultaneously an organism that has Q defending agents . The Immune response is modeliez in the following way : every defending agent choose randomly an infectious agent (only one) in the N aggressors, independently from the other defendants. With probability $\vartheta \in (0,1)$ the infectious agent is nullified.

Only one surviving infectious agent is required for the organism to be infected.

1. Show that the probability that a given aggressive agent infect the the organism is

$$p_{Q,N}(\vartheta) = \Big(1 - \frac{\vartheta}{N}\Big)^Q.$$

In the lab, we repeat n independent scenarios of aggression. In every experiment, an infectious agent is **marked**. For experiment i, we note $X_i = 1$ if the given agent did infect the organism, 0 otherwise.

2. Consider having a sample $(X_1, \ldots, X_n)$, where ϑ is the unknown parameter and Q et N are known. Show that the likelihood can be written as

$$\vartheta \rightsquigarrow p_{Q,N}(\vartheta)^{\sum_{i=1}^n X_i}\big(1 - p_{Q,N}(\vartheta)\big)^{n-\sum_{i=1}^n X_i}.$$

3. Show that the model is regular and that its Fisher information is given by

$$\mathbb{I}(\vartheta) = \frac{\big(\partial_\vartheta p_{Q,N}(\vartheta)\big)^2}{p_{Q,N}(\vartheta)\big(1 - p_{Q,N}(\vartheta)\big)}.$$

4. Show that the maximum likelihood estimator for ϑ is well defined, asymptotically normal and compute its limiting variance.

Suppose now that N et Q are unknown parameters of interest, and we take the limit $N \approx +\infty$ supposing that $Q = Q_N \sim \kappa N$ for a $\kappa > 0$ (unknown).

6. Going at the limit for Nin the previous model, show that the observation of $(X_1, \ldots, X_n)$ allow (identifiability) the estimate of $\widetilde{\vartheta} = \kappa\vartheta$ and hence compute its maximum likelihood estimator.