Université Paris Dauphine
Département MIDO
Licence 3 – Statistical modelling 2022–2023

Tutorial n°4 – Information, Sufficiency and Ancillarity

stoehr@ceremade.dauphine.fr

Exercise 1. Let X be a random variable admitting the following probability density:

$$f_X(x \mid \theta) = \frac{\theta}{x^{\theta+1}} \mathbb{1}_{\{x \geq 1\}}, \quad \text{where} \quad \theta > 0.$$

1. Show that the model is a regular exponential family.

2. Compute the Fisher information $\mathscr{I}_X(\theta)$ contained in X for the parameter θ. Deduce the information contained in a n-sample.

Instead of observing directly X, we observe a random variable Y defined by:

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

3. Compute the Fisher information $\mathscr{I}_Y(\theta)$ brought by Y for the parameter θ and show that $\mathscr{I}_X(\theta) > \mathscr{I}_Y(\theta)$.

Exercise 2. Consider the statistical model

$$\mathscr{P} = \left\{ f(x \mid b) = \exp\left[-\frac{|x-\mu|}{b} - \log(2b) \right], b \in \mathbb{R}_+^* \right\},$$

where $\mu \in \mathbb{R}$ is a known location parameter. Assuming X is distributed according to $f(\cdot \mid b)$, let set $T(X) = |X - \mu|$. Compute the Fisher information contained in $T(X)$ for the parameter b.

Exercise 3. We consider the statistical model $\mathscr{P} = \left\{ \mathscr{B}(p) \mid p \in]0,1[\right\}$.

1. Give the likelihood function associated with an observed sample $x_1, \ldots, x_n$.

2. **(a)** Compute the Fisher information contained in a random variable X following the Bernoulli distribution $\mathscr{B}(p)$. Deduce the information contained in the n-sample.

(b) What is the Fisher information for the natural parameter corresponding to the canonical form of this exponential family?

3. **(a)** Using the definition, show that $T = \sum_{i=1}^n x_i$ is a sufficient statistic for p.

(b) Derive the same result, using the factorisation theorem.

(c) Let set $n = 3$. Are the following statistics sufficient?

$$R = \exp(x_1 + x_2 + x_3) \quad \text{and} \quad S = 2x_1 + x_2 + 3x_3.$$

Exercise 4. Let consider $X_1,\dots,X_n$ a sample from the Cauchy distribution with location parameter $\theta \in \mathbb{R}$, *i.e.* whose density is given by

$$f_X(x \mid \theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}.$$

1. Write the likelihood function associated with a n-sample $x_1,\dots,x_n$.

2. Show that the order statistics is minimal sufficient.

Exercise 5. Let consider the discrete distribution $\mathbb{P}_\theta$ defined by

$$\mathbb{P}_\theta[X=1] = \frac{1-\theta}{6}, \quad \mathbb{P}_\theta[X=2] = \frac{1+\theta}{6}, \quad \mathbb{P}_\theta[X=3] = \frac{2-\theta}{6} \quad \text{and} \quad \mathbb{P}_\theta[X=4] = \frac{2+\theta}{6}$$

with $\theta \in [-1,1]$. Let $X_1,\dots,X_n$ be a sample from $\mathbb{P}_\theta$ and denote N_k, for $k \in \{1,2,3,4\}$, the number of times k appears in the sequence.

1. Show that $T = (N_1, N_2, N_3, N_4)$ is a sufficient statistic.

2. Show that $S = (N_1+N_2, N_3+N_4)$ is an ancillary statistic. Is this ancillary statistic unique?

◇ To do at Home ◇

Exercise 6. Let consider $X_1,\dots,X_4$ distributed according to a Poisson distribution of parameter $\lambda \in \mathbb{R}_+^*$. One set

$$T(X_1,\dots,X_4) = \log\left(1+\sum_{i=1}^{4} X_i\right) \quad \text{and} \quad R(X_1,\dots,X_4) = \log\left[\exp\left(-\frac{X_1+X_2}{2}\right) + \frac{1}{2}\exp\left(X_4-X_3\right)\right].$$

Show that $T(X_1,\dots,X_4)$ is a sufficient statistic for λ whereas $R(X_1,\dots,X_4)$ is not.

Exercise 7. Consider S and T that are minimal sufficient statistics for θ. Show the two following properties:

1. There exists a bijective transform between S and T.

2. Any sufficient statistic R for θ such that $R = g(S)$ is also minimal sufficient for θ.

Exercise 8. Given $\theta \in \mathbb{R}_+^*$, consider the random variables $U_1,\dots,U_n$ *i.i.d.* according to the uniform distribution on $[\theta, \theta+1]$ and $V_1,\dots,V_n$ *i.i.d.* according to the uniform distribution on $[\theta, 2\theta]$. Show that the following statistics are ancillary

$$R = \max(U_1,\dots,U_n) - \min(U_1,\dots,U_n) \quad \text{and} \quad S = \frac{V_2}{\max(V_1,\dots,V_n) - \min(V_1,\dots,V_n)}.$$