Echantillonnage

Exercice 1.

Soit $(X_1, ..., X_n)$ un $n-$échantillon de loi de Bernoulli de paramètre θ

$$P(X_i = x_i) = \theta^{x_i}(1-\theta)^{1-x_i} \qquad \forall x_i \in \{0,1\}\ ,\ \theta \in]0,1[$$

1. Ecrire la vraisemblance de l'échantillon
2. Montrer que $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$ est une statistique exhaustive pour θ

Correction Exercice 1.

1. $L(x_1, ...x_n, \theta) = \prod\limits_{i=1}^n P(X_i = x_i) = \prod\limits_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$
2. $L(x_1, ...x_n, \theta) = \left(\dfrac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i}(1-\theta)^n = g_\theta(T(x_1, ..., x_n))h(x_1, ..., x_n)$

   où $g_\theta(T(x_1, ..., x_n)) = \left(\dfrac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i}(1-\theta)^n$ et $h(x_1, ..., x_n) = 1$

   D'après le théorème de factorisation, $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$ est une statistique exhaustive pour θ

Exercice 2.

Pour chacune des lois suivantes, écrire la vraisemblance de l'échantillon $(X_1, ..., X_n)$ et donner une statistique exhaustive.

1. Loi de Poisson de paramètre θ
2. Loi de Uniforme sur $[0, \theta]$

Correction Exercice 2.

1. $L(x_1, ...x_n, \theta) = \prod\limits_{i=1}^n P(X_i = x_i) = \prod\limits_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^n x_i}}{\prod\limits_{i=1}^n x_i!}$

   $L(x_1, ...x_n, \theta) = \left(e^{-n\theta}\theta^{\sum_{i=1}^n x_i}\right)\left(\frac{1}{\prod\limits_{i=1}^n x_i!}\right) = g_\theta(T(x_1, ..., x_n))h(x_1, ..., x_n)$

   où $g_\theta(T(x_1, ..., x_n)) = e^{-n\theta}\theta^{\sum_{i=1}^n x_i}$ et $h(x_1, ..., x_n) = \frac{1}{\prod\limits_{i=1}^n x_i!}$

   D'après le théorème de factorisation, $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$ est une statistique exhaustive pour θ

2. $L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}\frac{1}{\theta}1_{0\le x_i\le\theta}=\frac{1}{\theta^n}1_{\min x_i\ge 0}1_{\max x_i\le\theta}$
   $L(x_1,...x_n,\theta)=\left(\frac{1}{\theta^n}1_{\max x_i\le\theta}\right)\left(1_{\min x_i\ge 0}\right)=g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$
   où $g_\theta(T(x_1,...,x_n))=\frac{1}{\theta^n}1_{\max x_i\le\theta}$ et $h(x_1,...,x_n)=1_{\min x_i\ge 0}$
   D'après le théorème de factorisation, $T(X_1,...,X_n)=\max X_i$ est une statistique exhaustive pour θ

Exercice 3.

Soit $(X_1,...,X_n)$ un $n-$échantillon de loi Uniforme sur $\left[\theta-\frac{1}{2},\theta+\frac{1}{2}\right]$
Montrer que la statistique $T(X_1,...,X_n)=(\max X_i,\min X_i)$ est exhaustive pour le paramètre θ.

Correction Exercice 3.

$L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}1\times 1_{\theta-\frac{1}{2}\le x_i\le\theta+\frac{1}{2}}=1_{\min x_i\ge\theta-\frac{1}{2}}1_{\max x_i\le\theta+\frac{1}{2}}$
$L(x_1,...x_n,\theta)=\left(1_{\min x_i\ge\theta-\frac{1}{2}}1_{\max x_i\le\theta+\frac{1}{2}}\right)(1)=g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$
où $g_\theta(T(x_1,...,x_n))=1_{\min x_i\ge\theta-\frac{1}{2}}1_{\max x_i\le\theta+\frac{1}{2}}$ et $h(x_1,...,x_n)=1$
D'après le théorème de factorisation, $T(X_1,...,X_n)=(\max X_i,\min X_i)$ est une statistique exhaustive pour θ

Exercice 4.

Soit $(X_1,...,X_n)$ un $n-$échantillon issu d'une population de densité :

$$f_\theta(x_i)=\begin{cases}\theta e^{-x_i+\theta} & \text{si} \quad x_i>\theta\\ 0 & \text{sinon}\end{cases}$$

Montrer que $T(X_1,...,X_n)=\min X_i$ est une statistique exhaustive pour θ.

Correction Exercice 4.

$L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}\left(\theta e^{-x_i+\theta}1_{x_i>\theta}\right)=\theta^n e^{n\theta-\sum_{i=1}^{n}x_i}1_{\min x_i>\theta}$

$L(x_1,...x_n,\theta)=\left(\theta^n e^{n\theta}1_{\min x_i>\theta}\right)\left(e^{-\sum_{i=1}^{n}x_i}\right)=g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$
où $g_\theta(T(x_1,...,x_n))=\theta^n e^{n\theta}1_{\min x_i>\theta}$ et $h(x_1,...,x_n)=e^{-\sum_{i=1}^{n}x_i}$
D'après le théorème de factorisation, $T(X_1,...,X_n)=\min X_i$ est une statistique exhaustive pour θ

Exercice 5.

Soit $(X_1,...,X_n)$ un $n-$échantillon issu d'une population de densité :

$$f(x_i)=\begin{cases}\frac{\theta^p}{\Gamma(p)}x_i^{-(p+1)}e^{-\frac{\theta}{x_i}} & \text{si} \quad x_i>0\\ 0 & \text{sinon}\end{cases}$$

Montrer que $T\left(X_1,...,X_n\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}$ est une statistique exhaustive pour le paramètre θ.

Correction Exercice 5.

$L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}\left(\frac{\theta^p}{\Gamma(p)}x_i^{-(p+1)}e^{-\frac{\theta}{x_i}}1_{x_i>0}\right)$

$=(\theta^{np})\left(\frac{1}{(\Gamma(p))^n}\left(\prod\limits_{i=1}^{n}x_i\right)^{-(p+1)}1_{\min x_i>0}\right)\times e^{(-n\theta)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)}=C(\theta)h(\mathbf{x})\exp\left(\alpha(\theta)T(\mathbf{X})\right)$

où $C(\theta)=\theta^{np}$, $\quad h(\mathbf{x})=\frac{1}{(\Gamma(p))^n}\left(\prod\limits_{i=1}^{n}x_i\right)^{-(p+1)}1_{\min x_i>0}$

et $\alpha(\theta)=-n\theta$ $\quad$, $\quad T(\mathbf{X})=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$

Donc $T\left(X_1,...,X_n\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}$ est une statistique exhaustive pour θ.

Exercice 6.

Soit $(X_1,...,X_n)$ un $n-$échantillon ou X_i a pour densité :

$$f_\theta(x_i)=\begin{cases}\frac{\theta}{e^{\theta^2-1}}e^{\theta x_i} & \text{si} \qquad 0\leq x_i\leq\theta \\ 0 & \text{sinon}\end{cases}$$

Montrer que $T\left(X_1,...,X_n\right)=\left(\max X_i,\sum_{i=1}^{n}X_i\right)$ est une statistique exhaustive pour θ est une statistique exhaustive pour le paramètre θ.

Correction Exercice 6.

$L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}\left(\frac{\theta}{e^{\theta^2-1}}e^{\theta x_i}\times 1_{0\leq x_i\leq\theta}\right)$

$=\left(\frac{\theta^n}{\left(e^{\theta^2-1}\right)^n}e^{\theta\sum_{i=1}^{n}x_i}\times 1_{\max_i x\leq\theta}\right)\left(1_{\min x_i\geq 0}\right)=g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$

où $g_\theta(T(x_1,...,x_n))=\frac{\theta^n}{\left(e^{\theta^2-1}\right)^n}e^{\theta\sum_{i=1}^{n}x_i}\times 1_{\max_i x\leq\theta}$ et $h(x_1,...,x_n)=1_{\min x_i\geq 0}$

D'après le théorème de factorisation, $T\left(X_1,...,X_n\right)=\left(\max X_i,\sum_{i=1}^{n}X_i\right)$ est une statistique exhaustive pour θ

Exercice 7.

Soit $(X_1,...,X_n)$ un $n-$échantillon issu d'une population de densité :

$$f_\theta(x_i)=\begin{cases}\theta x_i^{\theta-1} & \text{si} \qquad 0\leq x_i\leq 1 \\ 0 & \text{sinon}\end{cases}$$

Montrer que la statistique $T(X_1,...,X_n)=\prod_{i=1}^{n}X_i$ est exhaustive.

Correction Exercice 7.

$L(x_1,...x_n,\theta)=\prod_{i=1}^{n}f_\theta(x_i)=\prod_{i=1}^{n}\left(\theta x_i^{\theta-1}\times 1_{0\leq x_i\leq 1}\right)=\left(\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}\right)(1_{\min x_i\geq 0}1_{\max_i x\leq 1})=$
$g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$

où $g_\theta(T(x_1,...,x_n))=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}$ et $h(x_1,...,x_n)=1_{\min x_i\geq 0}1_{\max_i x\leq 1}$

D'après le théorème de factorisation, $T(X_1,...,X_n)=\prod_{i=1}^{n}X_i$ est une statistique exhaustive pour θ

Exercice 8.

Soit $(X_1,...,X_n)$ un $n-$échantillon de loi de Poisson de paramètre λ. Montrer à partir de la définition d'exhaustivité que $T(X_1,...,X_n)=\sum_{i=1}^{n}X_i$ est une statistique exhaustive pour λ.

Correction Exercice 8.

Loi de Poisson.

$P\left((X_1,...,X_n)=(x_1,...,x_n)\ /\ T(X_1,...,X_n)=t\right)$
$=\frac{P((X_1,...,X_n)=(x_1,...,x_n)\text{ et }T(X_1,...,X_n)=t)}{P(\ T(X_1,...,X_n)=t)}$
$=\frac{P\left(X_1=x_1,...,X_n=x_n\text{ et }\left(\sum_{i=1}^{n}X_i\right)=t\right)}{P\left(\ \left(\sum_{i=1}^{n}X_i\right)=t\right)}$
$=\frac{P(X_1=x_1,...,X_n=x_n\text{ et }X_1+...+X_n=t)}{P\left(\ \left(\sum_{i=1}^{n}X_i\right)=t\right)}$
$=\frac{P(X_1=t-x2-...-x_n,X_2=x_2,X_3=x_3,...,X_n=x_n)}{P\left(\ \left(\sum_{i=1}^{n}X_i\right)=t\right)}$
Or $X_i\rightarrow P(\lambda)\Longrightarrow\sum_{i=1}^{n}X_i\rightarrow P(n\lambda)$ et les X_i sont indépendantes on aura :

$P\left((X_1,...,X_n)=(x_1,...,x_n)\ /\ T(X_1,...,X_n)=t\right)$
$=\frac{P(X_1=t-x2-...-x_n)(X_2=x_2)(X_3=x_3)...)(X_n=x_n)}{P\left(\ \left(\sum_{i=1}^{n}X_i\right)=t\right)}$
$=\frac{\left(e^{-\lambda}\frac{\lambda^{t-x_2-...-x_n}}{(t-x_2-...-x_n)!}\right)\left(e^{-\lambda}\frac{\lambda^{x2}}{x_2!}\right)...\left(e^{-\lambda}\frac{\lambda^{xn}}{xn!}\right)}{e^{-n\lambda}\frac{(n\lambda)^t}{t!}}$
$=\frac{t!}{n^t(t-x_2-...-x_n)!x2!...x_n!}$
qui ne dépend pas de λ donc $T(X_1,...,X_n)=\sum_{i=1}^{n}X_i$ est une statistique exhaustive pour λ.

Loi de Bernoulli.

Soit $(X_1,...,X_n)$ un $n-$échantillon de loi de Bernoulli de paramètre p. Montrer à partir de la définition d'exhaustivité que $T(X_1,...,X_n)=\sum_{i=1}^{n}X_i$ est une statistique exhaustive pour p.

$P\left((X_1,...,X_n)=(x_1,...,x_n)\ \ /\ T(X_1,...,X_n)=t\right)$
$= \frac{P((X_1,...,X_n)=(x_1,...,x_n)\text{ et }T(X_1,...,X_n)=t)}{P(\ T(X_1,...,X_n)=t)}$
$= \frac{P\left(X_1=x_1,...,X_n=x_n\text{ et }\left(\sum_{i=1}^n X_i\right)=t\right)}{P\left(\ \left(\sum_{i=1}^n X_i\right)=t\right)}$
$= \frac{P(X_1=x_1,...,X_n=x_n\text{ et }X_1+...+X_n=t)}{P\left(\ \left(\sum_{i=1}^n X_i\right)=t\right)}$
$= \frac{P(X_1=t-x2-...-x_n,X_2=x_2,X_3=x_3,...,X_n=x_n)}{P\left(\ \left(\sum_{i=1}^n X_i\right)=t\right)}$
Or $X_i \to B(p)$ et $\sum_{i=1}^n X_i \to B(n,p)$ on aura :

$P\left((X_1,...,X_n)=(x_1,...,x_n)\ \ /\ T(X_1,...,X_n)=t\right)$
$= \frac{P(X_1=t-x2-...-x_n)(X_2=x_2)(X_3=x_3)...)(X_n=x_n)}{P\left(\ \left(\sum_{i=1}^n X_i\right)=t\right)}$
$= \frac{\left(p^{t-x2-...-x_n}(1-p)^{1-(t-x2-...-x_n)}\right)\left(p^{x_2}(1-p)^{1-x_2}\right)...\left(p^{x_n}(1-p)^{1-x_n}\right)}{C_n^t p^t(1-p)^{n-t}}$
$= \frac{\left(p^{t-x2-...-x_n}(1-p)^{1-(t-x2-...-x_n)}\right)\left(p^{x_2}(1-p)^{1-x_2}\right)...\left(p^{x_n}(1-p)^{1-x_n}\right)}{C_n^t p^t(1-p)^{n-t}}$
$= \frac{1}{C_n^t}$
qui ne dépend pas de p donc $T(X_1,...,X_n) = \sum_{i=1}^n X_i$ est une statistique exhaustive pour p.

Exercice 9.

Soit $(X_1,...,X_n)$ un $n-$échantillon issu d'une population de densité :

$$f_\theta(x_i) = \begin{cases} 2\theta x_i e^{-\theta x_i^2} & \text{si} \quad 0 \le x_i \le \theta \\ 0 & \text{sinon} \end{cases}$$

Montrer que la statistique $T(X_1,...,X_n) = \frac{n}{\sum X_i^2}$ est exhaustive pour θ.

Correction Exercice 9.

$L(x_1,...x_n,\theta) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \left(2\theta x_i e^{-\theta x_i^2} \times 1_{0\le x_i\le\theta}\right)$
$= (2\theta)^n \left(\prod_{i=1}^n x_i\right) e^{-\theta\sum x_i^2} \times 1_{\max_i x\le\theta} 1_{\min x_i\ge 0}$
$\sum X_i^2$ est une statistique exhaustive pour θ. L'application $x \mapsto \frac{n}{x}$est une bijection sur R_+^* d'où la statistique $\frac{n}{\sum X_i^2}$ est aussi exhaustive pour θ.

Exercice 10.

Soit $(X_1,...,X_n)$ un $n-$échantillon issu d'une population de densité :

$$f_\theta(x_i) = \begin{cases} \frac{x_i^{\frac{1}{\theta}-1}}{\theta a^{\frac{1}{\theta}}} & \text{si} \quad 0 \le x_i \le a \\ 0 & \text{sinon} \end{cases}$$

Montrer que la statistique $T\left(X_1,...,X_n\right)=\frac{1}{n}\sum\ln\left(\frac{X_i}{a}\right)$ est exhaustive pour θ.

Correction Exercice 10.

$L(x_1,...x_n,\theta)=\prod\limits_{i=1}^{n}f_\theta(x_i)=\prod\limits_{i=1}^{n}\left(\frac{x_i^{\frac{1}{\theta}-1}}{\theta a^{\frac{1}{\theta}}}1_{0\leq x_i\leq a}\right)$

$=\left(\frac{1}{\theta a^{\frac{1}{\theta}}}\right)^n\left(\prod\limits_{i=1}^{n}x_i\right)^{\frac{1}{\theta}-1}1_{\max_i x\leq a}1_{\min x_i\geq 0}=g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$

où $g_\theta(T(x_1,...,x_n))=\left(\frac{1}{\theta a^{\frac{1}{\theta}}}\right)^n\left(\prod\limits_{i=1}^{n}x_i\right)^{\frac{1}{\theta}-1}$ et $h(x_1,...,x_n)=1_{\max_i x\leq a}1_{\min x_i\geq 0}$

D'après le théorème de factorisation, $T\left(X_1,...,X_n\right)=\prod\limits_{i=1}^{n}X_i$ est une statistique exhaustive pour θ

ou :

$f_\theta(x_i)=\frac{x_i^{\frac{1}{\theta}-1}}{\theta a^{\frac{1}{\theta}}}1_{0\leq x_i\leq a}=\frac{1}{\theta a^{\frac{1}{\theta}}}1_{0\leq x_i\leq a}\times e^{\left(\frac{1}{\theta}-1\right)\ln x_i}=C(\theta)h(x)\exp\left(\alpha(\theta)T(X)\right)$

où $C(\theta)=\frac{1}{\theta a^{\frac{1}{\theta}}}$, $h(x)=1_{0\leq x_i\leq a}$, $\alpha(\theta)=\frac{1}{\theta}-1$ et $T(X)=\ln X$

Donc $T\left(X\right)=\ln X$ est une statistique privilégiée et dans le cas d'un modèle d'échantillonnage la statistique $T\left(X_1,...,X_n\right)=\sum_{i=1}^{n}\ln X_i$ est exhaustive pour le paramètre θ.

L'application $x\mapsto\frac{1}{n}\left(\ln\frac{x}{a}\right)$est une bijection sur R_+^* d'où la statistique $\frac{1}{n}\sum\ln\left(\frac{X_i}{a}\right)$ est aussi exhaustive pour θ.