Master1 PSA

Série2

Exercise 1 *Soient les variables aléatoires indépendantes et de Bernoulli* $X_1, ..., X_n$. *Nous avons donc, pour tout* $i = 1, ..., n,\ \ P[X_i = 1] = p$ *et* $P[X_i = 0] = 1 - p$.

1. Montrez que $T = \sum_{i=1}^{n} X_i$ *est une statistique exhaustive pour le paramètre* p.

2. Pour $n = 3$, *la statistique* $\widetilde{T} = e^{X_1+X_2+X_3}$ *est-elle aussi exhaustive ? Même question pour* $S = X_1 + 2X_2 + X_3$.

Exercise 2 *Soient les variables aléatoires* $X_1, ..., X_n$ *i.i.d.* $\mathcal{P}(\lambda)$. *Montrez que* $T = \sum_{i=1}^{n} X_i$ *est une statistique exhaustive pour le paramètre* λ.

Exercise 3 *Soit* $X_1, ..., X_n$ *un échantillon aléatoire simple issu d'une population de densité* $f(x) = \theta x^{\theta-1}$ *si* $0 < x < 1$ *et 0 sinon où* $\theta > 0$. *Déterminez une statistique exhaustive pour le paramètre.*

Exercise 4 *Soient les variables aléatoires* $X_1, ..., X_n$ *iid* $\mathcal{U}[\theta - 1/2, \theta + 1/2]$. *Déterminez une statistique exhaustive pour le paramètre.*

Exercise 5 *Soit* $X_1, ..., X_n$ *un échantillon aléatoire simple issu d'une population de densité* $f(x) = \theta e^{\theta - x}$ *si* $x > \theta$ *et 0 sinon où* $\theta > 0$. *Déterminez une statistique exhaustive pour le paramètre* θ.

Exercise 6 *Les éléments d'une population possèdent un caractère* X *qui suit une loi de probabilité dont la densité est* $f(x) = 2\theta x e^{-\theta x^2}$ *si* $0 < x < \theta$ *et 0 sinon où* $\theta > 0$. *Une suite de* n *expériences indépendantes a donné les valeurs* $x_1, ..., x_n$.

1. Montrez que l'estimateur $\widetilde{T} = n / \sum_{i=1}^{n} X_i^2$ *est une statistique exhaustive pour le paramètre* θ.

2. Ecrire la fonction de répartition Y *en posant* $y = \theta x^2$ *ainsi que la fonction de densité de* Y. *Déduire la loi de* $Z = \sum_{i=1}^{n} Y_i$.

3. Etudier la convergence de $\widetilde{T}$.

Exercise 7 *Soit* X *une v.a de loi* $\mathcal{N}(\theta_1, \theta_2)$ *et* Y *une v.a de loi* $\mathcal{N}(\theta_3, \theta_4)$. *On suppose que* X *et* Y *sont indépendantes. On dispose d'un n-échantillon de* (X, Y).

1. Déterminer une statistique exhaustive et complète lorsque $\theta_1, \theta_2, \theta_3, \theta_4$ *sont inconnus.*

2. Déterminer une statistique exhaustive lorsque $\theta_1 = \theta_3$. *Montrer qu'elle n'est pas complète.*

3. On suppose $\theta_2 = \theta_4$. *Existe-t-il une statistique exhaustive et complète ?*

Exercise 8 *Cas gaussien,*

1. *Montrer que* $a(\underline{x}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ *est exhaustive pour* (m, σ^2).
2. *En déduire que* $a(\underline{x})$ *est une statistique complète.*

Exercise 9 *Ecrire la vraisemblance et déterminer une statistique exhaustive pour un échantillon de n observations i.i.d. de lois :*

1) loi de Poisson de paramètre λ *:* $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$*;* $k \in \mathbb{N}^*$

2) loi de Pareto de paramètres α *et* θ *avec* $\alpha > 1$*,* $\theta >$ *0 de densité :* $f(x) = \frac{\alpha-1}{\theta}\left(\frac{x}{\theta}\right)^{\alpha} 1_{[0,+\infty]}(x)$.

3) loi de Weibull de paramètre α *et* θ *avec* $\alpha > 0$*,* $\theta > 0$ *de densité :* $f(x) = \alpha\theta x^{\alpha-1}e^{-\theta x^{\alpha}} 1_{[0,+\infty]}(x)$.

Exercise 10 *Calculer l'information de Fisher dans les modèles statistiques suivants :*

1) une loi de Poisson de paramètre : $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$*;* $k \in \mathbb{N}^*$

2) une loi de Pareto de paramètres α *et* θ *avec* $\alpha > 0$*,* $\theta > 0$ *de densité :* $f(x) = \alpha\theta x^{\alpha-1}e^{-\theta x^{\alpha}} 1_{[0,+\infty]}(x)$.

3) une loi de Weibull de α *et* θ *avec* $\alpha > 0$*,* $\theta > 0$ *de densité :* $f(x) = \alpha\theta x^{\alpha-1}e^{-\theta x^{\alpha}} 1_{[0,+\infty]}(x)$.

4) loi uniforme sur $[0, \theta]$ *avec* $\theta > 0$ *inconnu.*

Exercise 11 *On étudie une variable aléatoire* X*, de densité* $f(., \theta)$.

1) Quelle est la fonction score du modèle, notée $S(X, \theta)$ *? Donner l'expression de l'information de Fisher* $I_X(\theta)$.

2) En fait, on ne parvient pas à observer X*, mais seulement* Y *définie par :*

$Y = \begin{cases} 1 \text{ si } X \geq s \\ 0 \text{ si } X < s \end{cases}$ *où s est un seuil connu.*

On suppose que l'on peut intervertir l'intégrale et la dérivée. Donner la fonction score du modµele, notée $S_Y(y, \theta)$. *En déduire que* $S_Y(y, \theta) = E(S(X, \theta)|Y = y)$

3) En déduire alors que $I_X(\theta) >> I_Y\theta)$*, oµu* $I_Y(\theta)$ *est l'information de Fisher associée à* Y *(l'inégalité s'entend au sens des matrices symétriques). Quelle interprétation pouvez-vous donner à l'inégalité ci-dessus ?*

Exercise 12 *1) Calculer l'estimateur du maximum de vraisemblance (e.m.v.)* $\widehat{p}$ *de* p *dans le modèle* X_i *iid et suivent la loi* $B(1, p)$ *et calculer la loi limite de* $\sqrt{n}(\widehat{p} - p)$.

2) Calculer l'e.m.v. $(\widehat{m}, \widehat{\sigma}^2)$ *de* (m, σ^2) *dans le modèle* X_i *iid et suivent la loi* $\mathcal{N}(m, \sigma^2)$ *et donner la loi limite du vecteur* $\sqrt{n}\begin{pmatrix} \widehat{m} - m \\ \widehat{\sigma}^2 - \sigma^2 \end{pmatrix}$

Exercise 13 *Soit* $x = (x_1, x_2, ..., x_n)$ *n-échantillon de* $X \sim U[0, \theta]$

1- Montrer que $T_1 = 2\overline{X}$ *et* $T_2 = \frac{n+1}{n} max_{1 \leq i \leq n}(X_i)$ *sont des estimateurs sans biais de* θ.

2- Montrer que T_2 *est un estimateur sans biais de variance minimale uniformément minimale (ESBVUM).*

Corrigé

EXO1

Par définition, une statistique $T(X)$ est exhaustive si

$$P(X = (x_1, ..., x_n)|T(X) = y)$$

est indépendant de la valeur du paramètre. Calculons donc

$P(X = (x_1, ..., x_n)|T(X) = y) = P(X_1 = x_1\ X_2 = x_2 ... X_n = x_n | \sum_{i=1}^{n} X_i = y)$

$= \frac{P(X_1=x_1 et X_2=x_2...X_n=x_n \text{ et } \sum_{i=1}^{n} X_i=y)}{P(\sum_{i=1}^{n} X_i=y)} = \frac{P(X_2=x_2...X_n=x_n \text{ et } X_1=y-x_2-...-x_n)}{P(\sum_{i=1}^{n} X_i=y)}$

Par indépendance, on obtient alors (en se rappelant que la somme de n v.a. de Bernouilli est une binomiale de paramètres (n, p), pour $x_1 = y - x_2 - \cdots - x_n$,

$P(X = (x_1, ..., x_n)|T(X) = y) = \frac{P(X_2=x_2)...P(X_n=x_n)P(X_1=y-x_2-...-x_n)}{P(\sum_{i=1}^{n} X_i=y)}$

$= \frac{p^{x_2}(1-p)^{1-x_2}...p^{x_n}(1-p)^{1-x_n}p^{y-x_2-...-x_n}(1-p)^{1-(y-x_2-...-x_n)}}{\binom{n}{y}p^y(1-p)^{n-y}} = \frac{1}{\binom{n}{y}}$

Cette quantité est indépendante de p. La statistique $T(X)$ est donc exhaustive. Remarquons que le calcul effectué ci-dessus revient exactement à calculer la vraisemblance conditionnelle et à vérifier que celle-ci ne dépend de p que via une fonction de y (apparaissant dans la condition).

Pour $n = 3$, la statistique $\widetilde{T}$ est en bijection avec T. Elle est donc exhaustive. Ce n'est pas le cas de S.

EXO2

On applique ici la meme construction qu'à l'exercice précédent...et on obtient (la somme de n v.a. de Poisson de paramètre λ est une v.a. de Poisson de paramètre $n\lambda$) :

$P(X = (x_1, ..., x_n)|T(X) = y) = \frac{P(X_2=x_2)...P(X_n=x_n)P(X_1=y-x_2-...-x_n)}{P(\sum_{i=1}^{n} X_i=y)}$

$= \frac{e^{-\lambda}\lambda^{x_2}(x_2!)^{-1}...e^{-\lambda}\lambda^{x_n}(x_n!)^{-1}e^{-\lambda}\lambda^{y-x_2-...-x_n}((y-x_2-...-x_n)!)^{-1}}{e^{-n\lambda}\frac{(n\lambda)^y}{y!}}$

$= \frac{(x_2!)^{-1}...(x_n!)^{-1}((y-x_2-...-x_n)!)^{-1}}{\frac{(n)^y}{y!}}$

qui est indépendant de la valeur de λ. La statistique est bien exhaustive.

EXO3

La méthode la plus simple pour calculer une statistique exhaustive passe par la condition nécessaire et suffisante. Il faut pour cela calculer la vraisemblance :

$$L_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} (\theta x_i^{\theta-1} 1_{(0<x_i<1)}) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \prod_{i=1}^{n} 1_{(0<x_i<1)}$$

$g_\theta(T(x_1,...,x_n)) = \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$

$h(x_1,...,x_n) = \prod_{i=1}^n 1_{(0<x_i<1)}$

Celle-ci peut s'´ecrire $g_\theta(T(x_1,...,x_n))h(x_1,...,x_n)$ pour la statistique (exhaustive donc) : $T = \prod_{i=1}^n X_i$.

EXO4

La fonction de densité de chacune des variables aléatoires est donnée par $f(x) = \frac{1}{\theta}1_{[\theta-1/2,\theta+1/2]}$.

$L_\theta(x_1,...,x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \left(\frac{1}{\theta}1_{[\theta-1/2\leq x_i\leq\theta+1/2]}\right) = \left(\frac{1}{\theta^n}1_{[\theta-1/2\leq x_{(1)}]}1_{x_{(n)}\leq\theta+1/2]}\right)$

$g_\theta(T(x_1,...,x_n)) = \frac{1}{\theta^n}1_{[\theta-1/2\leq x_{(1)}]}1_{x_{(n)}\leq\theta+1/2]}$

$h(x_1,...,x_n) = 1$

Une statistique exhaustive est donnée par

$$T(X) = \left(X_{(1)}, X_{(n)}\right).$$

EXO5

La vraisemblance est donnée par

$L_\theta(x_1,...,x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \left(\theta e^{\theta-x_i}1_{x_i>\theta}\right) = \theta^n \prod_{i=1}^n \left(e^{\theta-x_i}1_{x_i>\theta}\right) = \theta^n e^{-\sum_{i=1}^n x_i} e^{n\theta}1_{x_{(1)}>\theta}$

on pose $h(x) = e^{-\sum_{i=1}^n x_i}$ et $g_\theta(T(X)) = \theta^n e^{n\theta}1_{x_{(1)}>\theta}$

Une statistique exhaustive est donc donnée par $T(X) = X_{(1)}$.

EXO6

1. Le support de la distribution ne dépendant pas du paramètre inconnu θ, nous supposerons que tous les x_i sont strictement positifs. Ecrivons la vraisemblance du modèle :

$$L_\theta(X) = (2^n \prod_{i=1}^n x_i)\theta^n e^{-\theta\sum_{i=1}^n x_i^2}$$

Le critère de factorisation nous dit alors que $\sum_{i=1}^n x_i^2$ est une statistique exhaustive pour θ.

Par ailleurs, la fonction $x \to \frac{n}{x}$ est une fonction bijective sur $\mathbb{R}_0^+$, et $\sum_{i=1}^n x_i^2$ est toujours dans $\mathbb{R}_0^+$. Etant en bijection avec une statistique exhaustive, la statistique $\widetilde{T} = n / \sum_{i=1}^{n} X_i^2$ proposée est également exhaustive.

2. Le changement de variable $y = \theta x^2$, ce qui implique $dy = 2\theta x dx$, et donc en écrivant la fonction de répartition :

$P(Y \leq u) = 0$ si $u \leq 0$ et,

$$P(Y \leq u) = P(X \leq \sqrt{\tfrac{u}{\theta}}) = \int_0^{\sqrt{\frac{u}{\theta}}} 2\theta x e^{-\theta x^2} dx = \int_0^u e^{-y} dx \text{ si } u \geq 0.$$

On constate ainsi que la densité de Y est

$$f(y) = e^{-y} \text{ si } y > 0 \text{ et } 0 \text{ sinon,}$$

soit la densité d'une loi $\Gamma(1,1)$. Puisque les X_i sont indépendants, les Y_i le sont aussi et grace aux propriétés de la loi gamma nous savons donc que $Z = \sum_{i=1}^{n} Y_i \sim \Gamma(1, n)$.

3. Pour étudier la convergence de $\widetilde{T}$ il faut calculer la variance de cet estimateur :

$$E(\widetilde{T}) = E\left(\tfrac{n\theta}{Z}\right) = \int_{\mathbb{R}} \tfrac{n\theta}{z} f(z) dz$$

$$E(\widetilde{T}) = \int_0^{\infty} \tfrac{n\theta}{\Gamma(n)} z^{n-2} e^{-z} dz.$$

La présence de z^{n-2} dans l'expression de l'intégrande, par ailleurs fort proche de la densité d'une loi gamma, nous suggère de faire apparaıtre la densité d'une $\Gamma(1, n-1)$. En utilisant les propriétés de la fonction Γ :

$$E(\widetilde{T}) = \tfrac{n\theta}{n-1} \int_0^{\infty} \tfrac{1}{\Gamma(n-1)} z^{n-2} e^{-z} dz$$

l'intégrale ci dessus étant simplement l'intégrale d'une densité $\Gamma(1, n-1)$ sur son domaine, elle vaut exactement 1 et par conséquent $E(\widetilde{T}) = \frac{n}{n-1}\theta$. On a

$$E(\widetilde{T}^2) = \int_0^{\infty} \left(\tfrac{n\theta}{z}\right)^2 f(z) dz$$

Le raisonnement est le même que pour l'espérance, mais cette fois la puissance de z dans l'intégrale est abaissée de deux unités, et l'on fait donc apparaıtre la densité d'une loi $\Gamma(1, n-2)$

$$E(\widetilde{T}^2)=\frac{n^2\theta^2}{(n-1)(n-2)}\int_0^\infty \frac{1}{\Gamma(n-2)}z^{n-3}e^{-z}dz=\frac{n^2\theta^2}{(n-1)(n-2)}$$

ce qui implique

$$Var(\widetilde{T})=\frac{n^2\theta^2}{(n-1)(n-2)}-\left(\frac{n}{n-1}\right)^2\theta^2=\frac{n^2\theta^2}{(n-1)^2(n-2)}$$

cette quantité tend bien vers 0, ainsi $\widetilde{T}$ est convergent.

EXO7

X et Y sont indépendantes$\Longrightarrow P_{X,Y}=P_XP_Y\Longrightarrow P_{X,Y}=\mathcal{N}(\theta_1,\theta_2)\times\mathcal{N}(\theta_3,\theta_4)$

$f_{X,Y}(x,y)=\frac{1}{\sqrt{2\pi\theta_2}}e^{-\frac{1}{2\theta_2}(x-\theta_1)^2}\times\frac{1}{\sqrt{2\pi\theta_4}}e^{-\frac{1}{2\theta_4}(y-\theta_3)^2}=\frac{1}{2\pi\sqrt{\theta_2\theta_4}}e^{-\frac{1}{2}\left[\frac{1}{\theta_2}(x-\theta_1)^2+\frac{1}{\theta_4}(y-\theta_3)^2\right]}=$
$C\exp(\sum_{j=1}^{4}\alpha_j(\theta_1,\theta_2,\theta_3,\theta_4)a_j(x,y))h(x,y)$

or

$-\frac{1}{2}\left(\frac{1}{\theta_2}(x-\theta_1)^2+\frac{1}{\theta_4}(y-\theta_3)^2\right)=x\frac{\theta_1}{\theta_2}-\frac{1}{2}\frac{y^2}{\theta_4}-\frac{1}{2}\frac{\theta_1^2}{\theta_2}-\frac{1}{2}\frac{\theta_3^2}{\theta_4}-\frac{1}{2}\frac{x^2}{\theta_2}+y\frac{\theta_3}{\theta_4}=$
$-\frac{1}{2}\frac{\theta_1^2}{\theta_2}-\frac{1}{2}\frac{\theta_3^2}{\theta_4}+(-\frac{1}{2}\frac{x^2}{\theta_2}-\frac{1}{2}\frac{y^2}{\theta_4}+x\frac{\theta_1}{\theta_2}+y\frac{\theta_3}{\theta_4})$

$\Longrightarrow$

$f_{X,Y}(x,y)=\frac{1}{2\pi\sqrt{\theta_2\theta_4}}e^{-\frac{1}{2}\frac{\theta_1^2}{\theta_2}}e^{-\frac{1}{2}\frac{\theta_3^2}{\theta_4}}\exp(-\frac{1}{2}\frac{x^2}{\theta_2}-\frac{1}{2}\frac{y^2}{\theta_4}+x\frac{\theta_1}{\theta_2}+y\frac{\theta_3}{\theta_4})$

on pose

$\alpha_1(\theta_1,\theta_2,\theta_3,\theta_4)=-\frac{1}{2\theta_2}$; $\alpha_2(\theta_1,\theta_2,\theta_3,\theta_4)=\frac{\theta_1}{\theta_2}$; $\alpha_1(\theta_1,\theta_2,\theta_3,\theta_4)=\frac{1}{2\theta_4}$; $\alpha_2(\theta_1,\theta_2,\theta_3,\theta_4)=\frac{\theta_3}{\theta_4}$

et

$a_1(x,y)=x^2$; $a_2(x,y)=x$; $a_3(x,y)=y^2$; $a_4(x,y)=y$

et

$h(x,y)=1$

$\Longrightarrow T=(\sum_{i=1}^{n}X_i^2,\sum_{i=1}^{n}X_i,\sum_{i=1}^{n}Y_i^2,\sum_{i=1}^{n}Y_i)$ est exhaustive.

$\#\ dimT=dim\theta=4\Longrightarrow T$ est complète.

Exo8$\mapsto$ mess

les autres$\leftarrow$polytechnique

EXO (dernier)

$X\sim U[0,\theta]$ de densité $f(x)=\frac{1}{\theta}1_{[0,\theta]}(x)$

On a $E(X)=\frac{\theta}{2}$ et donc $T_1=2X$ est un estimateur sans biais de θ.

On peut aussi vérifier que $T_2=\frac{n+1}{n}\max_{1\leq i\leq n}(X_i)$ est aussi un estimateur sans biais de θ. En effet

$$f_{\max_{1\leq i\leq n}(X_i)}(x)=\frac{nx^{n-1}}{\theta^n}1_{[0,\theta]}(x)\Rightarrow E(T)=\frac{n}{n+1}\theta$$

On a :

$$var(T_1) = var(2\overline{X}) = \frac{4var(X)}{n} = \frac{\theta^2}{3n}$$

et

$$var(T_2) = \left(\frac{n+1}{n}\right)^2 E(\boxed{S^2}) - \theta^2 = \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

On conclut que $\boxed{var(T_2) < var(T_1)}$ dès que $n > 1$, et donc que T_2 est un meilleur estimateur sans biais de que T_1 au sens du risque quadratique.

$3n - n(n+2) = -n\,(n-1) < 0 \Rightarrow 3n < n(n+2) \Rightarrow \frac{1}{3n} > \frac{1}{n(n+2)}$