Université 7 Novembre à Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information à Tunis

Année Universitaire
2010-2011
Première Année

Méthodes d'Estimation : Examen Session Principale

Enseignants : Mme Mallek et Mr Malouche **Durée : 1h30**

Exercice 1.

Pour modéliser la durée du phénomène du chômage X, on utilise généralement la distribution de Weibull dont la densité de probabilité a pour expression :

$$f(x,\theta) = \lambda\alpha x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mathbb{1}_{\{x>0\}} \qquad (\alpha,\lambda)\in\mathbb{R}_+^*\times\mathbb{R}_+^*$$

où α représente le coefficient de forme et l'inverse de λ désigne l'échelle du temps.

1. Soit $\underline{X} = (X_1,\ldots,X_n)$ un $n-$échantillon de la variable aléatoire X.
   (a) Ecrire la vraisemblance du modèle.
   (b) En déduire une condition sur le coefficient de forme pour que le modèle appartienne à la famille exponentielle à un paramètre.

   On suppose maintenant $\alpha = 2$.

   (a) Après avoir donné une statistique exhaustive pour le modèle, calculer son espérance et sa variance.
   (b) Calculer l'information de Fisher et la borne de Cramer-Rao. conclure.
   (c) Pouvait-on s'attendre à ce résultat.
2. On s'intéresse à l'estimation de λ
   (a) Montrer que $\widehat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)^{-1}$ est l'unique estimateur du maximum de vraisemblance de λ.
   (b) En admettant que $\sum_{i=1}^{n}X_i^2 \rightsquigarrow \gamma(n,\lambda)$, démontrer que que $\widehat{\lambda}$ est un estimateur biaisé de λ.
   (c) Déduire de $\widehat{\lambda}$ un estimateur sans biais $\widetilde{\lambda}$ de λ.
   (d) Montrer que $\widetilde{\lambda}$ est l'unique estimateur uvmb de λ.
   (e) Est-il efficace de λ? Est-il asymptotiquement efficace ?
3. On considère la variable aléatoire $Q = 2\lambda X^2$.
   (a) Sachant que $Q \rightsquigarrow \chi_2^2 = \gamma\left(1,\frac{1}{2}\right)$, trouver une statistique pivotale.
   (b) Donner un intervalle de confiance de niveau 0.95 Pour λ.

Exercice 2.

On dispose d'un n-échantillon d'une variable aléatoire X désignant le séquençage de l'ADN. X peut présenter 4 séquences possibles avec des probabilités respectives : $p_1 = 1 - \theta$; $p_2 = \theta - \theta^2$; $p_3 = \theta^2 - \theta^3$ *et* $p_4 = \theta^3$ où $0 \leq \theta \leq 1$.

Soient N_1, N_2, N_3 et N_4 les variables aléatoires "nombre d'occurences pour chacune des séquences $\left(\sum_{i=1}^{4} N_i = n\right)$.

1. Donner l'expression de $P_\theta [X = x]$.
2. Ecrire la vraisemblance associée à l'échantillon.
3. Déterminer $\widehat{\theta}$ l'estimateur par substitution de fréquence de θ et donner sa loi asymptotique.

Corrigé de l'exercice 1 *:*

1. $f(x,\theta)=\lambda\alpha x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mathbb{1}_{\{x>0\}}$ $\qquad\qquad \theta=(\alpha,\lambda)\in\mathbb{R}_+^*\times\mathbb{R}_+^*$

   (a) $\mathcal{L}\left(x_1,\ldots,x_n,\theta\right)=\lambda^n\alpha^n\prod\limits_{i=1}^{n}x_i^{\alpha-1}\exp\left(-\lambda\sum\limits_{i=1}^{n}x_i^{\alpha}\right)\mathbb{1}_{\left\{\underline{x}\in\left(\mathbb{R}_+^*\right)^n\right\}}$

   (b) $\mathcal{L}\left(x_1,\ldots,x_n,\theta\right)=\exp\left(-\lambda\sum\limits_{i=1}^{n}x_i^{\alpha}+n\ln\lambda+n\ln\alpha+(\alpha-1)\sum\limits_{i=1}^{n}\ln x_i\right)\mathbb{1}_{\left\{\underline{x}\in\left(\mathbb{R}_+^*\right)^n\right\}}$

   Pour que le modèle appartienne à la famille exponentielle, il faudrait que α soit constante. Dans ce cas, on aura

   $T\left(\underline{x}\right)=-\sum\limits_{i=1}^{n}x_i^{\alpha}\qquad c\left(\lambda\right)=\lambda\qquad d\left(\lambda\right)=n\ln\lambda\qquad S\left(\underline{x}\right)=(\alpha-1)\sum\limits_{i=1}^{n}\ln x_i$

   $A=\left(\mathbb{R}_+^*\right)^n$ *indépendante de* λ.

   On suppose maintenant $\alpha=2$.

2. $\mathcal{L}\left(x_1,\ldots,x_n,\lambda\right)=\exp\left(-\lambda\sum\limits_{i=1}^{n}x_i^2+n\ln\lambda+n\ln 2+\sum\limits_{i=1}^{n}\ln x_i\right)\mathbb{1}_{\left\{\underline{x}\in\left(\mathbb{R}_+^*\right)^n\right\}}.$

   (a) $T\left(\underline{X}\right)=-\sum\limits_{i=1}^{n}X_i^2$ *est la statistique exhaustive complète pour le modèle*

   La famille exponentielle est sous forme canonique. On a alors,

   $E_\lambda\left(T\left(\underline{X}\right)\right)=-d'\left(\lambda\right)=-\dfrac{n}{\lambda}$

   $Var\left(T\left(\underline{X}\right)\right)=-d"\left(\lambda\right)=\dfrac{n}{\lambda^2}$

   (b) $I_n\left(\lambda\right)=-E_\lambda\left[\dfrac{\partial^2}{\partial\lambda^2}\left(-\lambda\sum\limits_{i=1}^{n}x_i^2+n\ln\lambda+n\ln 2+\sum\limits_{i=1}^{n}\ln x_i\right)\right]$

   $I_n\left(\lambda\right)=-E_\lambda\left[\left(-\dfrac{n}{\lambda^2}\right)\right]=\dfrac{n}{\lambda^2}.$

   $BCR\left(-\dfrac{n}{\lambda}\right)=\dfrac{\left(\dfrac{\partial}{\partial\lambda}\left(-\dfrac{n}{\lambda}\right)\right)^2}{I_n\left(\lambda\right)}=\dfrac{\left(\dfrac{n}{\lambda^2}\right)^2}{\dfrac{n}{\lambda^2}}=\dfrac{n}{\lambda^2}=Var\left(T\left(\underline{X}\right)\right).$

   $T\left(\underline{X}\right)=-\sum\limits_{i=1}^{n}X_i^2$ *est donc efficace pour son espérance.*

   (c) *Ce résultat était prévisible car il s'agit d'une famille exponentielle à un paramètre sous forme canonique où $d\longmapsto n\ln\lambda$ est deux fois dérivable. $T\left(\underline{X}\right)$ est donc efficace pour son espérance.*

3. .

   (a) *La famille est exponentielle à un paramètre .L'application c est l'identité, elle est donc injective et de classe $\mathcal{C}^2$; d est aussi de classe $\mathcal{C}^2$. Comme l'ensemble des valeurs de λ est un ouvert, alors si l'équation $E_\lambda\left(T\left(\underline{X}\right)\right)=T\left(\underline{x}\right)$ possède une solution ;, celle-ci correspondra à l'unique estimateur du maximum de vraisemblance.*

   $E_\lambda\left(T\left(\underline{X}\right)\right)=T\left(\underline{x}\right)\Longleftrightarrow E_\lambda\left(-\sum\limits_{i=1}^{n}X_i^2\right)=-\sum\limits_{i=1}^{n}x_i^2\Longleftrightarrow-\dfrac{n}{\lambda}=-\sum\limits_{i=1}^{n}x_i^2\ \ (\lambda>0)$

   $\Longleftrightarrow\dfrac{n}{\sum\limits_{i=1}^{n}x_i^2}=\lambda.\left(x_i>0,\ 1\leq i\leq n\right).$

L'estimateur du maximum de vraisemblance est donc $\widehat{\lambda} = \left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2\right)^{-1}$.

(b) $\sum\limits_{i=1}^{n} X_i^2 \rightsquigarrow \gamma(n, \lambda)$

$$E_\lambda\left(\widehat{\lambda}\right) = \int_0^{+\infty} \frac{n}{t}\frac{\lambda^n}{\Gamma(n)} t^{n-1} \exp -\lambda t \; dt = \frac{n\lambda\Gamma(n-1)}{\Gamma(n)} \int_0^{+\infty} \frac{\lambda^{n-1}}{\Gamma(n-1)} t^{n-2} \exp -\lambda t \; dt$$

$$E_\lambda\left(\widehat{\lambda}\right) = \frac{n\lambda\Gamma(n-1)}{\Gamma(n)} = \frac{n}{n-1}\lambda \neq \lambda.$$

(c) $\widetilde{\lambda} = \frac{n-1}{n}\widehat{\lambda}$ $\quad E_\lambda\left(\widetilde{\lambda}\right) = \lambda$. $\widetilde{\lambda}$ *est sans biais.*

(d) $\widetilde{\lambda} = \frac{n-1}{\sum\limits_{i=1}^{n} X_i^2} = -\frac{n-1}{T(\underline{X})}$.

$\widetilde{\lambda}$ *est un estimateur sans biais de* λ *et fonction d'une statistique exhaustive complète. Il est donc uvmb (*$T(\underline{X})$ *est la statistique exhaustive complète associée à la famille exponentielle). Comme* $Var\,(T(\underline{X})) = \frac{n}{\lambda^2} < \infty$, *alors* $Var\left(\widetilde{\lambda}\right) < \infty$ *et* $\widetilde{\lambda}$ *est l'unique estimateur uvmb.*

(e) $\widetilde{\lambda}$ *ne peut être efficace car l'estimateur du maximum de vraisemblance est biaisé et donc il n'existe pas d'estimateur efficace.*

L'estimateur du maximum de vraisemblance $\widehat{\lambda}$ *est asymptotiquement efficace et comme* $\widetilde{\lambda} = \frac{n-1}{n}\widehat{\lambda}$ *avec* $\lim\limits_{n\to\infty} \frac{n-1}{n} = 1$, *alors* $\widetilde{\lambda}$ *est asymptotiquement efficace.*

4. $Q = 2\lambda X^2$.

(a) $Q \rightsquigarrow \chi_2^2 = \gamma\left(1, \frac{1}{2}\right)$

$2\lambda \sum\limits_{i=1}^{n} X_i^2 \rightsquigarrow \chi_{2n}^2$ *(somme de* n *variables aléatoires indépendantes de même loi* χ_2^2*)*

(b) $P_\lambda\left[F_{\chi_{2n}^2}^{-1}(\beta) \leq 2\lambda \sum\limits_{i=1}^{n} X_i^2 \leq F_{\chi_{2n}^2}^{-1}(0.95+\beta)\right] = 0.95$

La distribution étant unimodale, on supposera $0.05 \leq 2\min\left(F(x^*), 1-F(x^*)\right)$, *où* x^* *désigne le mode. De ce fait, l'intervalle optimal est associé à* $\beta = \frac{\alpha}{2}$.

$P_\lambda\left[F_{\chi_{2n}^2}^{-1}(0.025) \leq 2\lambda \sum\limits_{i=1}^{n} X_i^2 \leq F_{\chi_{2n}^2}^{-1}(0.975)\right] = 0.95$

D'où l'intervalle de confiance pour λ :

$$IC_{0.95}(\lambda) = \left[\frac{F_{\chi_{2n}^2}^{-1}(0.025)}{2\sum\limits_{i=1}^{n} X_i^2} ; \frac{F_{\chi_{2n}^2}^{-1}(0.975)}{2\sum\limits_{i=1}^{n} X_i^2}\right]$$

Corrigé de l'exercice 2 $p_1 = 1-\theta$; $p_2 = \theta - \theta^2$; $p_3 = \theta^2 - \theta^3$ *et* $p_4 = \theta^3$ *où* $0 \leq \theta \leq 1$.

1. $P[X) = x] = (1-\theta)^{1\!\!1_{\{x=1\}}} \left(\theta-\theta^2\right)^{1\!\!1_{\{x=2\}}} \left(\theta^2-\theta^3\right)^{1\!\!1_{\{x=3\}}} \theta^{3 1\!\!1_{\{x=4\}}} 1\!\!1(x)_{\{1,2,3,4\}}$

2. $\mathcal{L}(\underline{x}, \theta) = (1-\theta)^{\sum\limits_{i=1}^{n} 1\!\!1_{\{x_i=1\}}} \left(\theta-\theta^2\right)^{\sum\limits_{i=1}^{n} 1\!\!1_{\{x_i=2\}}} \left(\theta^2-\theta^3\right)^{\sum\limits_{i=1}^{n} 1\!\!1_{\{x_i=3\}}} \theta^{3\sum\limits_{i=1}^{n} 1\!\!1_{\{x_i=4\}}} 1\!\!1(\underline{x})_{\{1,2,3,4\}^n}$

$\mathcal{L}(\underline{x}, \theta) = (1-\theta)^{N_1} \left(\theta-\theta^2\right)^{N_2} \left(\theta^2-\theta^3\right)^{N_3} \theta^{3N_4} 1\!\!1(\underline{x})_{\{1,2,3,4\}^n}$

3. On a

$$\begin{cases} p_1 = 1-\theta \\ p_2 = \theta - \theta^2 \\ p_3 = \theta^2 - \theta^3 \\ p_4 = \theta^3 \end{cases} \Longrightarrow \theta = p_2 + p_3 + p_4 = h\left(p_2, p_3, p_4\right)$$

$$\widehat{\theta} = \widehat{p}_2 + \widehat{p}_3 + \widehat{p}_4 = \tfrac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=2\}} + \tfrac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=3\}} + \tfrac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=4\}}$$

$\widehat{\theta} = N_2 + N_3 + N_4$

$\sqrt{n}\left(\widehat{\theta} - \theta\right)$ *converge en loi vers la loi normale* $\mathcal{N}\left(0, \sigma^2\right)$

avec $\sigma^2 = \sum\limits_{i=2}^{4} p_i \left(\dfrac{\partial h}{\partial p_i}\right)^2 - \left(\sum\limits_{i=2}^{4} p_i \dfrac{\partial h}{\partial p_i}\right)^2$

$\sigma^2 = p_2 + p_3 + p_4 - \left(p_2 + p_3 + p_4\right)^2$