Statistiques
Master Statistique et econométrie
TD - Feuille n^o 3

M. Emily, V. Monbet

Master 1 - 2011

Exercice 1

On peut modéliser la durée d'une connection sur le site "www.Cpascher.com" par une loi gamma $\gamma(2, 1/\theta)$ de densité

$$f(x; \theta) = \theta^{-2} x \exp(-x/\theta) \mathbb{I}_{\mathbb{R}^+}(x).$$

Pour fixer vos tarifs publicitaires, vous voulez estimer le paramètre θ à partir d'un échantillon $X_1, ..., X_n$ de n durées de connexion.

1. Vérifiez que f est bien une densité de probabilité.

2. Vous décidez d'estimer θ par l'estimateur du maximum de vraisemblance $\hat{\theta}_n$. Déterminer $\hat{\theta}_n$.

3. Sachant que $E_\theta[X_1] = 2\theta$, l'estimateur $\hat{\theta}_n$ est-il biaisé?

4. Vous vous interrogez sur la performance de votre estimateur. On vous donne $Var_\theta(X_1) = 2\theta^2$. Comparez la variance de $\hat{\theta}_n$ à la borne de Cramer-Rao. Conclusion ?

Exercice 2

Information et exhaustivité.
Soit $X_1, \cdots, X_n$ des v.a.i.i.d de loi de Gauss de moyenne μ et de variance σ^2. Considérons

$$T_n = T(X_1, \cdots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

et définissons $Y_n = (n-1)\frac{T_n}{\sigma^2}$

1. Verifier que T_n est un estimateur sans biais de σ^2.

2. Donner la loi de Y_n ainsi que sa densité de probabilité.

3. En déduire la densité de probabilité $g(t; \sigma^2)$ de la loi de T_n en utilisant un changement de variables simple.

4. Calculer l'information de Fisher

$$I_{T_n}(\sigma^2) = E\left(\frac{\partial^2 g(t;\sigma^2)}{\partial(\sigma^2)^2}\right).$$

5. Donner l'infomation de Fisher sur σ^2 continue dans l'échantillon.

6. Utiliser le théorème ci dessous pour en déduire que T_n n'est pas exhaustive pour σ^2.

Théorème 1 *Pour toute statistique T, on a*

$$I_T(\theta) \leq I_n(\theta)$$

et $I_T(\theta) = I_n(\theta)$ si et seulement si T est exhaustive,
$I_T(\theta) = 0$ si et seulement T est libre (c'est à dire que sa loi ne dépend pas de θ.

Exercice 3

On considère le modèle de régression simple

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n$$

où $\epsilon_1, \cdots, \epsilon_n$ sont des variables indépendantes de loi de Gauss centrées et de variance σ^2.

1. Calculer $E(Y_i)$ et $Var(Y_i)$.

2. Donner les estimateurs $\hat{\alpha}$ et $\hat{\beta}$ du maximum de vraisemblance de α et β.

3. Vérifier que les estimateurs du maximum de vraisemblance de α et β sont identiques aux estimateurs des moindres carrés.

4. Montrer que $\hat{\alpha}$ et $\hat{\beta}$ sont des estimateurs sans biais de α et β.

5. Calculer la variance de $\hat{\alpha}$ et $\hat{\beta}$.

6. Sous quelle condition, ces estimateurs sont-ils efficaces?

Exercice 4

Soit X une variable aléatoire géométrique de paramètre π. Sa loi est donnée par

$$P(X=k)=(1-\pi)^{k-1}\pi,\ \ k=1,2,3,...$$

On peut reparamétriser cette loi en posant $\theta=1/\pi$, ce qui donne

$$P(X=k)=\frac{1}{\theta}\left(\frac{\theta-1}{\theta}\right)^{k-1},\ \ k=1,2,3,...$$

1. Montrer que
$$\frac{\partial}{\partial\theta}\left(\frac{\theta-1}{\theta}\right)^{k}=k\frac{1}{\theta^2}\left(\frac{\theta-1}{\theta}\right)^{k-1}$$
et que
$$\sum_{k=1}^{\infty}\left(\frac{\theta-1}{\theta}\right)^{k}=\theta-1$$

2. Supposons que l'on observe $x_1,...,x_n$ un échantillon issu de X. Donner l'estimateur du maximum de vraisemblance de θ.

3. Calculer l'information de Fisher $I_n(\theta)$ et en déduire la borne de Cramér-Rao.

4. L'estimateur du maximum de vraisemblance atteint-il la borne de Cramér-Rao? Commenter : qu'est ce que cela signifie ?

Exercice 5

On observe le nombre de cas graves traités chaque jour par un vétérinaire sur une période de 200 jours.

X : Nbre de cas graves (i)	0	1	2	3	4	5 et plus
Nbre de jours (ni)	50	74	50	21	4	1

Notons X la variable aléatoire observée. On modélise la loi de X par une loi de Poisson de paramètre λ.

Partie 1 - Considérons l'estimation de λ.

1. Donner l'estimateur du maximum de vraisemblance de λ. En déduire l'estimation correspondant aux observations ci-dessus. Puis calculer la probabilité d'avoir plus de 3 cas graves un jour donné.

2. L'estimateur du maximum de vraisemblance est-il efficace? Justifier votre réponse.

Partie 2 - Considérons maintenant l'estimation de $g(\lambda) = \lambda^2$ par l'estimateur

$$T_n = (\bar{X}_n)^2 - \frac{1}{n}\bar{X}_n$$

1. Montrer que T_n est un estimateur sans biais de $g(\theta)$.
2. Montrer que T_n est un estimateur de variance minimale pour $g(\theta)$.
   Indication : montrer que T_n est exhaustive complète puis utiliser le théorème de Lehmann-Scheffé.
3. Calculer la borne de Cramér-Rao pour T_n.
4. Prouver que $X_1 + \cdots + X_n$ suit une loi de Poisson de paramètre $n\theta$. Et montrer que
   $$Var(T_n) = \frac{4\theta^3}{n} + \frac{2\theta^2}{n^2}$$
5. En déduire que T_n n'est pas efficace pour $g(\theta)$.

Eléments de Corrigé

Exercice 4

L'espérance d'une variable aléatoire suivant une loi géométrique de paramètre p est $\frac{1}{p}$. La variance est $\frac{q}{p^2}$ avec $q = 1 - p$.

Démonstration - Calculs préliminaires, pour tout $x \in [0;1[$,

$$f(x) = \sum_{k=1}^{+\infty} x^{k-1} = \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x} \tag{1}$$

$$f'(x) = \sum_{k=1}^{+\infty} kx^{k-1} = \sum_{k=0}^{+\infty} (k+1)x^k = \frac{1}{(1-x)^2} \tag{2}$$

$$f''(x) = \sum_{k=1}^{+\infty} k(k+1)x^{k-1} = \sum_{k=1}^{+\infty} k^2x^{k-1} + \sum_{k=1}^{+\infty} kx^{k-1} = \frac{2}{(1-x)^3} \tag{3}$$

En particulier $\sum_{k=1}^{+\infty} k^2x^{k-1} = \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}$
On peut alors utiliser ces identités pour le calcul de l'espérance :

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} kq^{k-1}p \tag{4}$$

$$= \frac{p}{(1-q)^2} \tag{5}$$

$$= \frac{1}{p}, \tag{6}$$

car 1 ? q = p, et pour celui de la variance :

$$V(X) = \left(\sum_{k=1}^{+\infty} k^2q^{k-1}p \right) - \mathbb{E}[X]^2 \tag{7}$$

$$= \frac{2p}{(1-q)^3} - \frac{p}{(1-q)^2} - \frac{1}{p^2} \tag{8}$$

$$= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} \tag{9}$$

$$= \frac{q}{p^2}. \tag{10}$$