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

V. Monbet

Master 1 - 2012-2013

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

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 $\bar{X}_n$ est exhaustive complète pour λ puis utiliser le théorème de Lehman-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 $g(\theta)$.

Exercice 3

On dispose d'un n-échantillon $(X_1, \cdots, X_n)$ d'une loi de densité

$$(1-\theta)\mathbb{I}_{]-0.5,0]}(x) + (1+\theta)\mathbb{I}_{]0,0.5]}(x)$$

où θ est un paramètre réel inconnu.

1. Quelles conditions doit vérifier θ?

2. Déterminer l'estimateur du maximum de vraisemblance de θ.

3. Donner la forme du test de rapport de vraisemblance pour $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$

Exercice 4

Soit $p > 1$ un paramètre fixé et soient $X_1, \cdots, X_n$ n variables aléatoires indépendantes ayant la même loi, de densité

$$f_p(x) = \frac{p-1}{x^p}\mathbb{I}_{[1,+\infty[}(x).$$

1. Montrer qu'il existe deux constantes a et $b > 0$ telles que la suite

$$\frac{1}{b\sqrt{n}} \sum_{j=1}^{n} (\ln X_j - a)$$

converge en loi vers une variable gaussienne standard.

2. On désire tester l'hypothèse $p = 2$ contre l'alternative $p = 3$. Quelle est la forme des tests de Neyman-Pearson?

3. Déduire de ce qui précède une région critique de niveau asymptotique 5% pour le test envisagé.

Exercice 5

Dans le contexte de l'évaluation du nombre X d'accidents cérébraux consécutifs au vaccin anticoqueluche (sur 500000 vaccins), on veut déterminer si ce nombre est inférieur ou égal à 5 ou supérieur à 10. On modélise la distribution de X par une loi de Poisson de paramètre λ de sorte qu'on peut poser les hypothèses de test suivantes :

$$H_0 : \lambda \leq 5 \text{ contre } H_1 : \lambda \geq 10$$

1. Calculer le risque α et le manque de puissance β associés à la règle de décision suivante :
   on choisit H_0 si $x \leq 7$ et H_1 si $x \geq 8$ avec x le nombre d'accidents observé.

2. Chaque année 500000 enfants de moins de un an sont vaccinés en France. Une année, on observe 6 accidents. Que concluez-vous? Donner le degré de signification réel du test.

3. Même question pour 12 accidents.