ENSAE

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

Cursus Intégré
2004-2005

Rappels de statistique mathématique
Enoncé des travaux dirigés n°2

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ `http://ensae.no-ip.com/SE222/`

Version du 20050421-15h38, révisée le 21 avril 2005

Enoncé de l'exercice 1

On considère n systèmes dont les durées de vie $X_1, \ldots, X_n$ suivent indépendamment une mê
loi de densité f. On observe uniquement les durées de vie $Y_1, ..., Y_r$ des r premiers systèr
tombant en panne.

☞ Q1 Ecrire la loi du r-uplet $(Y_1, ..., Y_r)$, puis celle de la variable Y_r.

☞ Q2 On suppose ici que la loi des X_i est exponentielle de densité

$$f(x) = \frac{1}{\theta} \exp\left(\frac{-(x-\alpha)}{\theta}\right) \mathbb{1}_{[\alpha,+\infty[}(x)$$

où $\theta > 0$ et $\alpha \geq 0$ sont des paramètres inconnus.
Ecrire la loi du r-uplet $(Y_1, ..., Y_r)$ dans ce cas.

☞ Q3 Trouver une statistique exhaustive pour les paramètres α et θ.

Enoncé de l'exercice 2

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

☞ Q1 une loi de Poisson de paramètre λ :

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{pour } k \in \mathbb{N}$$

☞ Q2 une loi de Pareto de paramètres α et θ avec $\alpha > 1$ et $\theta > 0$, de densité :

$$f(x) = \frac{\alpha - 1}{\theta} \left(\frac{\theta}{x}\right)^{\alpha} \mathbb{1}_{x \geq \theta}$$

☞ Q3 une loi de Weibull de paramètres α et θ avec $\alpha > 0$ et $\theta > 0$ de densité :

$$f(x) = \alpha\theta x^{\alpha-1} e^{-\theta x^{\alpha}}$$

☞ Q4 loi uniforme sur $[0, \theta]$ avec $\theta > 0$ inconnu.

Enoncé de l'exercice 3

On dispose de n observations $y_1, \ldots, y_n$ sur les durées de vie de certains composants industriels. On suppose que les variables aléatoires $Y_1, \ldots, Y_n$ associées sont i.i.d, de densité $f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \mathbb{1}_{t \geq 0}$.

☞ Q1 Soit F la fonction de répartition des Y_i. On cherche à estimer la “fonction de survie” de chaque composant $\bar{F}(t) = 1 - F(t)$.
Calculer $\bar{F}(t)$ en fonction de t et θ.

☞ Q2 Calculer l'estimateur du maximum de vraisemblance de θ et en déduire un estimateur convergent $\hat{F}(t)$ de $\bar{F}(t)$. Que peut-on dire du biais de $\hat{F}(t)$?

☞ Q3 Calculer la loi limite de $\sqrt{n}\left(\hat{F}(t) - \bar{F}(t)\right)$.

Soit T la variable aléatoire définie par :

$$T = \begin{cases} 1, \text{ si } Y_1 > t \\ 0, \text{ sinon.} \end{cases}$$

On note par ailleurs $S(Y) = Y_1 + \cdots + Y_n$.

☞ Q4 (a) Déterminer la loi de Y_1 conditionnellement à S.

(b) Calculer

$$T^*(Y) = \mathbb{E}\left(T | S(Y)\right)$$

Comment s'appelle cet estimateur ?

(c) Montrer que T^* est l'estimateur sans biais de $\bar{F}(t)$ optimal (parmi les estimateurs sans biais).

(d) Peux-on dire si T^* est efficace (à distance finie) ?

Enoncé de l'exercice 4

On étudie une variable aléatoire X, de densité $f(\cdot, \theta)$ (f est C^1).

☞ Q1 Quelle est la fonction score du modèle, notée $S(X; \theta)$?
Donner l'expression de l'information de Fisher $I_X(\theta)$.

☞ Q2 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 $\int_{\mathcal{X}}$ et $\frac{\partial}{\partial \theta}$; donner la fonction score du modèle, notée $S_Y(y; \theta)$.
En déduire que

$$S_Y(y; \theta) = \mathbb{E}\left(S_X(X; \theta) | Y = y\right)$$

☞ Q3 En déduire alors que $I_X(\theta) >> I_Y(\theta)$, où $I_Y(\theta)$ est l'information de Fisher associée à Y (l
égalité s'entend au sens des matrices symétriques).
Quelle interprétation pouvez-vous donner à l'inégalité ci-dessus ?
