Statistiques
Master Statistique et econométrie
TD - Feuille n^o 2

V. Monbet

Master 1 - 2012

Exercice 1

On a observé les durées de vie (en millier d'heures) de 30 composants électroniques. les résultats sont les suivants :

0.1 ; 7.4 ; 1.0 ; 7.9 ; 2.1 ; 1.8 ; 17.9 ; 9.3 ; 6.5 ; 3.3 ; 5.6 ; 7.7 ; 0.1 ; 24.3 ; 8.1 ; 10.0; 11.9 ; 1.6 ; 2.7 ; 0.5 ; 5.6 ; 42.5 ; 5.2 ; 2.0 ; 0.2 ; 15.0 ; 3.5 ; 6.4 ; 0.6 ; 3.3

On admet que $\sum_{i=1}^n x_i = 223.5$ et $\sum_{i=1}^n x_i^2 = 3826.8$.

Première partie - On suppose dans cette partie que la durée de vie des composants électroniques suit une loi exponentielle de paramètre inconnu $\theta > 0$. On rappelle que la densité de cette loi est donnée par :

$$f_\theta(x) = \frac{1}{\theta}\exp(-\frac{x}{\theta}) \text{ pour } x \geq 0,\ 0 \text{ sinon}$$

1. Calculer l'estimateur du maximum de vraisemblance T de θ ainsi que l'estimation correspondante.
2. Cet estimateur du maximum de vraisemblance est-il une statistique exhaustive?
3. Calculer le biais et l'erreur en moyenne quadratique de T.
4. Etudier les propriétés asymptotiques de T (consistance, normalité asymptotique).
5. Calculer la fonction de répartition de la loi exponentielle, puis en déduire un estimateur de la probabilité que la durée de vie d'un composant soit supérieur à une durée quelconque $t \geq 0$. En déduire une estimation que la durée de vie d'un composant soit supérieur à 20000h, 30000h et 40000h et comparer ces résultats avec les fréquences empiriques calculées à partir des données.
6. Tracer sur un même graphique la densité de la loi exponentielle ajustée et un histogramme décrivant la répartition des durées de vie observées (on utilisera un découpage en classes de largeur 3). Discuter la qualité de l'ajustement.

Deuxère partie - On suppose maintenant que la durée de vie des composants électroniques suit une loi dont la densité de probabilité est donnée par :

$$f_\theta(x) = \frac{x}{\theta^2}\exp(-\frac{x}{\theta}) \text{ pour } x \geq 0,\ 0 \text{ sinon}$$

avec θ un paramètre inconnu.

1. Calculs préliminaires. On pose pour $n \geq 0$,

$$J_n(\theta) = \int_0^{+\infty} x^n \exp\left(-\frac{x}{\theta}\right) \mathrm{d}x$$

   (a) Montrer, à l'aide d'une intégration par partie, que pour $n \geq 0$, $J_{n+1} = (n+1)\theta J_n(\theta)$. En déduire que $J_n(\theta) = \theta^{n+1} n!$

   (b) En déduire que f_θ définit bien une densité de probabilité, puis que si X est une variable aléatoire dont la loi admet la densité de probabilité f_θ alors $E[X] = \theta$ et $Var(X) = 2\theta^2$.

2. Calculer l'estimateur du maximum de vraisemblance T de θ, ainsi que l'estimation correspondante.

3. Caluler le biais et l'erreur en moyenne quadratique de T.

4. Etudier les propriétés asymptotiques de T.

Exercice 2

Il y a en France 17800 passages à niveau et on a relevé le nombre d'accidents mortels (hors suicides) sur ces passages entre 1985 et 1997. Les nombres observés sont les suivants :

1985	3
1988	2
1991, 1993, 1995, 1997	1

On suppose que le nombre d'accicents X au cours d'une année suit une loi de Poisson de paramètre $\theta > 0$ inconnu. On a alors, pour $k \in \mathbb{N}$,

$$P_\theta(X = k) = \frac{\theta^k \exp(-\theta)}{k!}$$

On rappelle que $E[X] = \theta$ et $Var(X) = \theta$.

1. Calculer l'estimateur du maximum de vraisemblance de T_n de θ, puis l'estimation de θ.

2. Montrer que l'estimateur obtenu à la question précédente est une statistique exhaustive pour θ.

3. Calculer le biais et l'erreur en moyenne quadratique de T_n, puis étudier ses propriétés asymptotiques.

4. Etudier les propriétés asymptotiques de $\sqrt{T_n}$

Exercice 3 (CC 2 - 2011-2012)

On considère le modèle de régression simple

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \cdots, n$$

où $\epsilon_1, \cdots, \epsilon_n$ sont des variables indépendantes de loi de Gauss centrées et de variance σ^2 connue et $x_1, \cdots, x_n$ sont des quantités connues déterministes.

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

2. Montrer que les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ du maximum de vraisemblance de β_0 et β_1 sont :

$$\hat{\beta}_0 = \overline{Y_n} - \hat{\beta}_1.\overline{x}_n \quad \text{et} \quad \hat{\beta}_1 = \frac{\displaystyle\sum_{i=1}^n x_i(Y_i - \overline{Y_n})}{\displaystyle\sum_{i=1}^n x_i(x_i - \overline{x}_n)}.$$

avec $\overline{Y_n} = \frac{1}{n}\sum_{i=1}^n Y_i$ et $\overline{x_n} = \frac{1}{n}\sum_{i=1}^n x_i$

Indication - On remarque de la matrice $\begin{pmatrix} 1 & \overline{x}_n \\ \overline{x}_n & \frac{1}{n}\sum_{i=1}^n x_i^2 \end{pmatrix}$ est définie positive.

3. Montrer que $\hat{\beta}_1$ s'écrit sous la forme :

$$\hat{\beta}_1 = \frac{\displaystyle\sum_{i=1}^n (x_i - \overline{x}_n)(Y_i - \overline{Y_n})}{\displaystyle\sum_{i=1}^n (x_i - \overline{x}_n)^2} = \beta_1 + \frac{\displaystyle\sum_{i=1}^n (x_i - \overline{x}_n)\epsilon_i}{\displaystyle\sum_{i=1}^n (x_i - \overline{x}_n)^2}.$$

4. Montrer que $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs sans biais de β_0 et β_1.

5. Calculer la variance de $\hat{\beta}_0$ et $\hat{\beta}_1$.

Exercice 4 (CC 1 - 2011-2012)

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

$$\sum_{k=1}^{\infty}\left(\frac{\theta - 1}{\theta}\right)^k = \theta - 1,$$

   en déduire le calcul de $\mathbb{E}[X]$ et $V(X)$.

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. Étudier les propriétés asymptotiques de cet estimateur (consistance, normalité asymptotique).

4. Calculer l'information de Fisher $I_n(\theta)$ et en déduire la borne de Cramer-Rao de cet estimateur.

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