ENSAE

ECOLE NATIONALE
DE LA STATISTIQUE
ET DE L'ADMINISTRATION
ECONOMIQUE

Cursus Intégré
2004-2005

Rappels de statistique mathématique
Enoncé des travaux dirigés n°6

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 un échantillon $X_1, \ldots, X_n$ i.i.d. tiré de la loi de Poisson de paramètre λ.

☞ Q1 Donner $\mathbb{E}(X_i)$ et $\mathbb{V}(X_i)$.
Calculer $\phi(t) = \mathbb{E}\left(e^{tX_i}\right)$.
En déduire que

$$\begin{aligned}
\mathbb{E}(X) &= \lambda \\
\mathbb{E}\left(X^2\right) &= \lambda(1+\lambda) \\
\mathbb{E}\left(X^3\right) &= \lambda\left(1+3\lambda+\lambda^2\right) \\
\mathbb{E}\left(X^4\right) &= \lambda\left(1+7\lambda+6\lambda^2+\lambda^3\right)
\end{aligned}$$

☞ Q2 (a) On pose $\widehat{\lambda_1}(x) = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$ et $\widehat{\lambda_2}(x) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$.
Par quelle méthode d'estimation ont été obtenus $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$?

(b) Les estimateurs $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$ estiment-ils λ sans biais ?
Proposer un autre estimateur sans biais de λ, que l'on notera $\widehat{\lambda_3}$.

(c) Donner la loi (à distance finie) de $\widehat{\lambda_1}$.

(d) Donner la loi asymptotique jointe de $\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_2} \end{pmatrix}$, puis celle de $\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_3} \end{pmatrix}$.

(e) Soit $\widehat{\lambda_\infty} = \alpha\widehat{\lambda_1} + \beta\widehat{\lambda_3}$, $(\alpha, \beta) \in \mathbb{R}^{+2}$.
Donner la valeur de (α, β) telle que $\widehat{\lambda_\infty}$ soit le meilleur estimateur asymptotiquement sans biais de λ.
Ce résultat est-il surprenant (donner l'e.m.v. de λ) ?

☞ Q3 Dans cette question, on se place dans le cas où $n = 2$.

(a) Calculer les erreurs quadratiques moyennes pour les trois estimateurs $\widehat{\lambda_1}$, $\widehat{\lambda_2}$ et $\widehat{\lambda_3}$.

(b) Déterminer en fonction de λ lequel de ces estimateurs est le meilleur selon la critère de l'erreur quadratique moyenne.

(c) Comparer explicitement l'EQM de $\hat{\lambda}_3$ avec la borne FDCR.

Enoncé de l'exercice 2

On considère une population de n individus infectés par un virus ; on étudie leurs durées d'incubation $(T_i)_{i\in[\![1,n]\!]}$, dont on suppose qu'elle est observable.
Pour modéliser l'hétérogénéité de la population, on suppose qu'on peut caractériser chaque individu i par un "facteur de risque" inobservable, réalisation de la variable aléatoire Λ_i, de telle sorte que :

– la loi de T_i, conditionnellement à Λ_i est la loi exponentielle de paramètre λ_i (de densité $\lambda_i e^{-\lambda_i t}\mathbb{1}_{t\geq 0}$) ;
– la famille $(\Lambda_i)_{i\in[\![1,n]\!]}$ est identiquement distribuée selon la loi $\mathcal{L}$ de densité

$$f(\lambda) = \frac{\alpha^r}{\Gamma(r)}\lambda^{r-1}e^{-\alpha\lambda}\mathbb{1}_{\mathbb{R}^+}$$

où $\alpha > 0$ et $r > 2$;
– les couples (T_i, Λ_i) sont indépendants entre eux.

☞ Q1 Donner la vraisemblance de $(T_1, \ldots, T_n)$.

☞ Q2 Calculer, lorsqu'il existe, le moment d'odre k $\mathbb{E}\left(T_i^k\right)$.

☞ Q3 Calculer l'information de Fisher du modèle.
Dans le cas où α est connu, calculer l'estimateur du maximm de vraisemblance de r.
Que se passe-t-il si α et r sont tous deux inconnus ?

☞ Q4 On suppose α connu.
Déterminer au moyen de la méthode des moments un estimateur convergent de r.
Cet estimateur est-il sans biais ? Est-il asymptotiquement efficace ?

☞ Q5 On suppose α et r inconnus.

(a) En utilisant les deux premiers moments de T_i, trouver des estimateurs convergents $\tilde{\alpha}$ et $\tilde{r}$ de α et r.

(b) Donner la loi limite du vecteur

$$\sqrt{n}\begin{pmatrix} \bar{T} - \mathbb{E}(T) \\ \bar{T^2} - \mathbb{E}(T)^2 \end{pmatrix}$$

En déduire la loi asymptotique du vecteur $(\tilde{\alpha}, \tilde{r})$.

Enoncé de l'exercice 3

On considère les réalisations de T variables aléatoires i.i.d. $Y_1, Y_2, \ldots, Y_T$, issues d'une loi de Poisson de paramètre λ inconnu.
On s'intéresse au test de l'hypothèse :

$$H_0 : \text{“ } \lambda = \lambda_0\text{” contre } H_a : \text{“}\lambda \neq \lambda_0\text{”}$$

où λ_0 est un réel donné.

☞ Q1 Pour tester ce type d'hypothèse, on dispose de trois tests asymptotiques usuels : test de **Wald**, du **score** et du **rapport de maximum de vraisemblance**. Rappeler rapidement leur principe et définir les statistiques de test sur lesquelles ils s'appuient.

☞ Q2 Pratiquer explicitement chacun de ces tests, en calculant l'expression de la statistique de test ξ_T^{test} et en définissant la région critique au seuil $\alpha \in [0,1]$: $W_\alpha = \{(y_1, \ldots, y_T) \,/\, \xi_T^{test}(y_1, \ldots, y_T) > q_{1-\alpha}^{\chi_1^2}\}$

☞ Q3 On veut tester $\lambda_0 = 1$ au seuil $\alpha = 5\%$.
Décider de chacun des tests dans le cas où l'échantillon observé $(y_1, \ldots, y_T)$ est tel que :
- $T = 100$ et $\bar{y} = 1,2$;
- $T = 200$ et $\bar{y} = 1,1$.

Discuter de l'acceptation de H_0 et expliquer pourquoi certains tests ne conduisent pas à la même décision.
Conclure.