ENSAE

Deuxième année - Cursus intégré

Introduction à la Statistique Mathématique

Ecrit. Deux heures. Tous documents autorisés

Nicolas CHOPIN

Mardi 25 mai 2004

EXERCICE 1

On rappelle la densité d'une loi Gamma(α, β), pour α, $\beta > 0$,

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x),$$

par rapport à la mesure de Lebesgue, avec $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t}\, dt$. Son espérance vaut α/β et sa variance α/β^2. En particulier, pour $\alpha = 1$ et $\beta = \theta$, on retrouve la loi exponentielle Exp(θ). Noter que $\Gamma(1) = 1$.

1) Calculer la fonction de répartition d'une loi exponentielle Exp(θ).

2) Soient X_1^*, ..., X_n^* variables indépendantes et identiquement distribuées selon Exp(θ), et soient X_1, ..., X_n leurs versions *tronquées*, c'est à dire

$$X_i = \begin{cases} X_i^* & \text{si } X_i^* \leq l \\ l & \text{sinon} \end{cases}$$

où l est une constante supposée connue. On suppose que seuls les X_i sont observés, et que la valeur de θ est inconnue.

Montrer que la vraisemblance de ce modèle s'écrit sous la forme

$$L(x_1, \ldots, x_n; \theta) = \theta^{n-k} \exp(-\theta \sum_{i=1}^n x_i)$$

où $k = k(x_1, \ldots, x_n)$ est le cardinal de l'ensemble $\{i : x_i \geq l\}$.

3) Ce modèle est-il exponentiel? Si oui, donner sa statistique canonique, et en déduire les fonctions de θ pour lesquelles on peut construire un estimateur sans biais efficace.

4) Calculer l'Information de Fisher du modèle, donner l'expression de l'estimateur du maximum de vraisemblance et son comportement asymptotique.

5) Donner l'expression de la médiane $M(\theta)$ de la loi $\text{Exp}(\theta)$. En déduire l'estimateur du maximum de vraisemblance de $M(\theta)$, et retrouver sa loi asymptotique.

Rappel : La médiane d'une variable aléatoire réelle continue Y est le réel M vérifiant $P(Y \leq M) = 1/2$.

6) Montrer que la famille des lois Gamma est conjuguée pour le modèle considéré, et donner la loi a posteriori correspondante (c'est à dire en supposant une loi a priori $\text{Gamma}(\alpha, \beta)$ pour le paramètre θ).

7) On souhaite tester $H_0 : M(\theta) < l$ contre $H_1 : M(\theta) \geq l$. Peut-on utiliser le théorème de Lehmann pour construire un test uniformément plus puissant dans ce cas précis? Justifier. En considérant à nouveau une loi a priori $\text{Gamma}(\alpha, \beta)$, exprimer le facteur de Bayes correspondant au test $H_0 : M(\theta) < l$ contre $H_1 : M(\theta) \geq l$ sous forme d'un rapport d'intégrales (que l'on ne cherchera pas à calculer).

EXERCICE 2

Soit $(X_1, ..., X_n)$ un échantillon de n variables aléatoires indépendantes et identiquement distribuées, suivant une loi de Pareto $\mathcal{P}(\alpha, \theta)$, de densité

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

par rapport à la mesure de Lebesgue (sur $\mathbb{R}$). La quantité $\alpha > 1$ est une constante supposée connue, et $\theta > 0$ est le paramètre à estimer.

1) Montrer que la fonction de répartition d'une loi de Pareto $\mathcal{P}(\alpha, \theta)$ vaut:

$$F(t) = P(X_1 < t) = 1 - (t/\theta)^{-(\alpha-1)}, \text{ pour } t \geq \theta$$

2) Donner une statistique exhaustive pour ce modèle (de dimension constante, quelque soit n).

3) Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ.

4) Déterminer la loi exacte de la variable aléatoire $\widehat{\theta}_n$, puis donner son comportement asymptotique. Commenter.

5) Calculer $E_\theta[\widehat{\theta}_n]$ (en précisant pour quelles valeurs de α cette espérance existe), et en déduire un estimateur sans biais de θ. La question 'Est-ce que cet estimateur sans biais est efficace?' a-t-elle un sens ici?

6) On souhaiter tester $H_0 : \theta = \theta_0$ contre $H_0 : \theta \neq \theta_0$, pour $\theta_0 > 0$. Peut-on utiliser les tests asymptotiques vus en cours dans ce cas? Justifier. A partir du comportement asymptotique de $\hat{\theta}_n$ obtenu en 4), construire un test asymptotique de niveau (asymptotique) α et de région critique $W_n = \{(x_1, \ldots, x_n) : \widehat{\theta}_n \notin [\theta_0; \theta_0 + c_n]\}$, où c_n est une constante à préciser. Préciser les propriétés asymptotiques de ce test et justifier la forme de région critique choisie ici.

7) On considère la loi a priori π de densité $\pi(\theta) \propto 1/\theta$. Quelle est la nature de cette loi? Calculer la densité de la loi a posteriori $\pi(\theta|x_1, ..., x_n)$, où $x_1, ..., x_n$ sont les valeurs observées de l'échantillon $X_1, ..., X_n$ (en précisant la constante de normalisation) et son espérance.