Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2006 – 2007

Statistiques I

Examen terminal, janvier 2007

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. On considère une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre p. On définit:

$$Y = \min\{k \in \mathbb{N}^*,\ X_k = 0\}.$$

   (a) Comment peut-on interpréter la variable Y? Montrer que la loi de Y est:

$$\mathbb{P}(Y = k) = p^{k-1}(1-p) \quad \text{pour} \quad k \in \mathbb{N}^*.$$

   (b) On suppose que $(Y_1, \ldots, Y_n)$ est un échantillon de variables aléatoires indépendantes et identiquement distribuées suivant la loi de Y avec $p \in]0,1[$ est inconnu. Déterminer le modèle statistique et sa mesure dominante. Montrer que ce modèle est exponentiel. En déduire, un estimateur $\widehat{p}_n$ de p sans biais et efficace. Déterminer la borne de Cramer-Rao et vérifier que cette borne est bien atteinte par $\widehat{p}_n$.

   (c) La variable Y est tronquée lorsqu'elle est trop "grande", par un paramètre $T \in \mathbb{N}^*$, c'est-à-dire que l'on définit une variable Y_T telle que $Y_T = \min(Y, T)$.

2. Soit X une variable aléatoire dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de densité :

$$f(x) = \frac{\lambda}{2}\exp\big(-\lambda|x-m|\big) \quad \text{pour } x \in \mathbb{R},$$

   avec $m \in \mathbb{R}$ et $\lambda > 0$, des paramètres inconnus.

   (a) Calculer l'espérance et la variance de X.

   (b) Calculer $P(X = m)$ et $P(X < m)$. En déduire la médiane (théorique) de la loi de X.

   (c) Soit une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X, dont on extrait un échantillon observé $(X_1, \ldots, X_{2n+1})$. Par ailleurs, on note $X_{(1)} \le X_{(2)} \le \ldots \le X_{(2n+1)}$ la statistique d'ordre associée. Soit :

$$\widehat{H}_n(a) = \frac{1}{2n+1}\sum_{i=1}^{2n+1}|X_i - a| \quad \text{pour } a \in \mathbb{R}.$$

   Calculer $\widehat{H}_n(X_{(n+1)})$ en fonction des $X_{(i)}$. Montrer que la fonction $a \mapsto \widehat{H}_n(a)$ est minimale en $X_{(n+1)}$ (on pourra développer $\widehat{H}_n(X_{(n+k)})$ en fonction des $X_{(i)}$ pour $k > 1$).

(d) On suppose ici que $m = 1$, donc que m est connu ($\lambda > 0$ restant inconnu). Quel est alors le modèle statistique ? Montrer que ce modèle appartient à la famille exponentielle, et en déduire une statistique exhaustive dont vous montrerez qu'elle est complète. Déterminer la matrice d'information de Fisher du modèle. Quelle est la fonction de λ (à une transformation affine près) que l'on peut estimer efficacement ? Déterminer l'estimateur de maximum de vraisemblance de λ et montrer qu'il vérifie un théorème de la limite centrale.

(e) On suppose désormais que $m \in \mathbb{R}$ est inconnu, tout comme $\lambda > 0$. Quel est alors le modèle statistique ? Montrer que ce modèle n'appartient pas à la famille exponentielle. A l'aide de la question 2.(c), déterminer un estimateur $(\widehat{m}_n, \widehat{\lambda}_n)$ du maximum de vraisemblance du couple (m, λ).

(f) Pour $a \in \mathbb{R}$, démontrer que $\widehat{H}_n(a)$ converge presque sûrement quand $n \to \infty$ vers $\mathbb{E}(|X - a|)$. Montrer que la fonction $a \in \mathbb{R} \mapsto \mathbb{E}(|X - a|)$ est minimale en $a = m$. En déduire que $\widehat{m}_n \xrightarrow[n\to+\infty]{p.s.} m$, puis que $\widehat{\lambda}_n \xrightarrow[n\to+\infty]{p.s.} \lambda$.