Maîtrise de Mathématiques 2003-2004
Statistiques
3. Statistiques d'ordre. Information de Kullback-Leibler

3.1. Soient $X_1,\ldots,X_n$ n variables aléatoires réelles indépendantes de même loi admettant une densité f. Pour chaque $\omega \in \Omega$, on définit $\sigma(\omega)$ la permutation de $\{1,\ldots,n\}$ qui classe par ordre croissant $X_1(\omega),\ldots,X_n(\omega)$. Ainsi

$$X_{\sigma(\omega)(1)}(\omega) \leqslant X_{\sigma(\omega)(2)}(\omega) \leqslant \ldots \leqslant X_{\sigma(\omega)(n)}(\omega).$$

On note pour simplifier $X_{\sigma(\omega)(i)}(\omega) := X_{(i)}(\omega)$. Soit R_i le rang de X_i dans le classement précédent.

a) Montrer que la variable aléatoire $\omega \mapsto \sigma(\omega)$ est p.s. définie et donner sa loi de probabilité. Calculer, pour $1 \leqslant i \leqslant n$ et $1 \leqslant k \leqslant n$, $\mathrm{P}(\sigma(i) = k)$.

b) Trouver la densité du n-uplet $(X_{(1)},\ldots,X_{(n)})$.

c) Trouver la loi de $(R_1,\ldots,R_n)$ et ensuite montrer que les deux n-uplets $(R_1,\ldots,R_n)$ et $(X_{(1)},\ldots,X_{(n)})$ sont indépendants.

d) Calculer $\mathrm{E}(R_i)$, $\mathrm{Var}(R_i)$ et $\mathrm{Cov}(R_i,R_j)$.

e) Trouver la densité de $X_{(i)}$ $(1 \leqslant i \leqslant n)$. Écrire les cas particuliers $i = 1$ et $i = n$.

f) Trouver la densité du couple $X_{(i)},X_{(j)}$, $i < j$. Écrire le cas particulier $(i,j) = (1,n)$.

g) Trouver la densité de $X_{(n)} - X_{(1)}$.

On suppose maintenant que la loi commune des $X_1,\ldots,X_n$ est la loi uniforme sur $[0,1]$.

h) Montrer que, pour i fixé et lorsque $n \to \infty$, $nX_{(i)}$ converge en loi vers une loi gamma dont on précisera les paramètres.

i) Montrer que $\ln X_{(i)}$ a la même loi que $-\sum_{k=i}^n Y_k/k$ où les variables aléatoires Y_k sont indépendantes de même loi exponentielle de paramètre 1.

j) Montrer que les variables $Z_1,\ldots,Z_n$ définies par $Z_i = (X_{(i)}/X_{(i+1)})^i$ $(1 \leqslant i \leqslant n-1)$, $Z_n = (X_{(n)})^n$ sont indépendantes et de loi uniforme sur $[0,1]$.

On suppose maintenant que la loi commune des $X_1,\ldots,X_n$ est la loi exponentielle de paramètre λ.

k) Montrer que les variables $Y_1 = nX_{(1)}$, $Y_i = (n+1-i)(X_{(i)} - X_{(i-1)})$ $(1 < i \leqslant n)$ sont indépendantes et que ce n-uplet $(Y_1,\ldots,Y_n)$ a la même loi que $(X_1,\ldots,X_n)$.

l) Si $X_1,\ldots,X_n$ sont les instants de panne de n machines identiques, mises en route simultanément, alors $X_{(i)}$ est donc l'instant de la i-ème panne observée. Montrer que les intervalles entre deux pannes observées sont indépendants.

On revient à la situation générale et on note $F(x) = \int_{-\infty}^x f(y)dy$.

m) Montrer que les variables aléatoires $U_i = [F(X_{(i)})/F(X_{(i+1)})]^i$ $(1 \leqslant i \leqslant n-1)$ et $U_n = [F(X_n)]^n$ sont indépendantes et de loi uniforme sur $[0,1]$. On pourra remarquer que les variables aléatoires sont de loi uniforme sur $[0,1]$ et que les variables aléatoires $-\ln F(X_i)$ sont exponentielles de paramètre 1.

3.2. Trouver l'information de Kullback-Leibler $H(\mathrm{P}_2 \mid \mathrm{P}_1)$ dans les cas suivants:

a) $\mathrm{P}_i = \mathcal{E}(\lambda_i)$, $i = 1,2$;

b) $\mathrm{P}_i = \mathcal{P}(\lambda_i)$, $i = 1,2$;
c) $\mathrm{P}_i = \mathcal{B}(1,p_i)$, $i = 1,2$;
d) $\mathrm{P}_i = \mathcal{N}(m_i,\sigma_i^2)$, $i = 1,2$;