EXAMEN STATISTIQUE - 1TR

Mercredi 1 Décembre 2014 (8h-9h45)

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi continue de densité

$$p(x;\theta) = \begin{cases} \frac{1}{2} - \theta \text{ si } -1 \leq x < 0 \\ \theta + \frac{1}{2} \text{ si } 0 \leq x < 1 \\ 0 \text{ sinon} \end{cases}$$

avec $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$.

1. Montrer que la vraisemblance des observations $x_1, ..., x_n$ s'écrit

$$p(x_1, ..., x_n; \theta) = \left(\frac{1}{2} - \theta\right)^{n-y_n} \left(\theta + \frac{1}{2}\right)^{y_n}$$

   où y_n est le nombre d'observations x_i vérifiant $x_i \geq 0$. Montrer que cette vraisemblance admet un unique maximum global pour une valeur de θ que l'on déterminera. On définit la variable aléatoire associée Y_n comme suit

$$Y_n = \text{card}\{i \in \{1, ..., n\} | X_i \geq 0\}.$$

   où card$\{A\}$ est le nombre d'éléments de l'ensemble A. Exprimer l'estimateur du maximum de vraisemblance de θ noté $\hat{\theta}_n$ en fonction de n et Y_n.

2. Montrer que Y_n suit une loi binomiale $\mathcal{B}(n, p)$ avec $p = \frac{1}{2} + \theta$. En déduire que $\hat{\theta}_n$ est un estimateur sans biais et convergent du paramètre θ.

3. Déterminer la borne de Cramer-Rao d'un estimateur non biaisé de θ. L'estimateur $\hat{\theta}_n$ est-il l'estimateur efficace du paramètre θ ?

4. On désire maintenant estimer le paramètre $a = \theta + \frac{1}{2} \in]0, 1[$.

   - Quel est l'estimateur des moments du paramètre a noté $\widetilde{a}_n$ construit à partir de $E[X_i]$?
   - Quel est l'estimateur du maximum de vraisemblance du paramètre a noté $\hat{a}_n$?
   - On suppose qu'on dispose d'une information a priori sur le paramètre a résumée dans la loi beta de densité

$$p(a) = \frac{a^{\alpha-1}(1-a)^{\beta-1}}{B(\alpha, \beta)} \mathcal{I}_{]0,1[}(a)$$

     où $\mathcal{I}_{]0,1[}(a)$ est la fonction indicatrice sur l'intervalle $]0, 1[$ et où $B(\alpha, \beta)$ est la fonction beta dont l'expression n'est pas importante pour cet exercice. Montrer que la loi a posteriori de $a|x_1, ..., x_n$ est une loi beta dont on précisera les paramètres. Quel est l'estimateur du maximum a posteriori du paramètre a noté a_n^* ?

Exercice 2: Tests Statistiques

Comme dans l'exercice précédent, on considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi continue de densité

$$p(x;\theta) = \begin{cases} \frac{1}{2} - \theta \text{ si } -1 \leq x < 0 \\ \theta + \frac{1}{2} \text{ si } 0 \leq x < 1 \\ 0 \text{ sinon} \end{cases}$$

avec $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$. Dans un premier temps, on cherche à effectuer le test d'hypothèses simples

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0 \\ H_1 & : & \theta = \theta_1 \end{array}$$

avec $\theta_1 > \theta_0$.

1. Montrer que la statistique du test de Neyman Pearson est $Y_n = \text{card}\{i \in \{1, ..., n\} | X_i \geq 0\}$ et déterminer la région critique du test. Commenter la forme de ce test à l'aide de l'expression de la densité $p(x;\theta)$.

2. En remarquant que $Y_n = \sum_{i=1}^n Z_i$, où Z_i est une variable aléatoire binaire telle que $Z_i = 1$ si $X_i \geq 0$ et $Z_i = 0$ si $X_i < 0$, donner la loi approchée de Y_n pour n "grand" découlant de l'application du théorème de la limite centrale. On supposera que cette approximation est suffisamment précise pour être utilisée dans la suite de ce exercice.

3. En utilisant la loi approchée déterminée à la question précédente, exprimer les risques de première et seconde espèce α et β en fonction du seuil déterminant la région critique du test, des paramètres n, θ_0 et θ_1, et de la fonction $\phi(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$.

4. Déterminer les courbes COR et analyser leur comportement en fonction de n. Quelles sont les deux autres quantités notées $A(\theta_0, \theta_1)$ et $B(\theta_0, \theta_1)$ dont dépendent les courbes COR ?