Statistique de Base - M1 Pro - Rattrapage du 5 mai 2009

Aucun document autorisé
Calculatrice autorisée

Exercice 1 (Estimation du paramètre d'échelle d'une loi gamma et tests)
Soit X une variable aléatoire de loi gamma avec paramètres de forme $\alpha = 4$ et paramètre d'échelle inconnu $\theta > 0$. On rappelle que la densité d'une loi gamma de paramètre α et θ est :

$$f = (x; \alpha, \theta) = \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\theta^{\alpha}} \mathbb{1}_{]0,+\infty[}(x).$$

On dispose d'un échantillon i.i.d. de cette distribution, $X_1, \ldots, X_n$.

1. Montrer qu'il s'agit d'un modèle exponentiel. Donner une statistique exhaustive.

2. Donner l'information de Fisher $I(\theta)$.

3. On s'intéresse à l'estimation par maximum de vraisemblance de θ.

a) Donner l'estimateur du maximum de vraisemblance $\hat{\theta}$ associé à θ.

b) Est-ce que $\hat{\theta}$ est un estimateur efficace de θ ?

c) Donner la loi asymptotique de $\sqrt{n}(\hat{\theta} - \theta)$.

d) En déduire une région critique asymptotique de niveau de signification α pour le test suivant

$$\begin{cases} H_0 : & \theta = 1 \\ H_1 : & \theta > 1. \end{cases}$$

e) Expliciter la fonction puissance du test précédent en utilisant c).

f) Tracer le graphique associé à cette fonction puissance.

4. On se propose maintenant de résoudre le test précédent de façon exacte et optimale.

a) Dans un premier temps, obtenir *la forme* (et seulement la forme) de la région critique optimale pour le test intermédiaire suivant (ne pas oublier de justifier) :

$$\begin{cases} H_0 : & \theta = 1 \\ H_1 : & \theta = \theta_1 \quad (\theta_1 > 1). \end{cases}$$

b) En déduire la région critique uniformément plus puissante pour le test

$$\begin{cases} H_0 : & \theta = 1 \\ H_1 : & \theta > 1. \end{cases}$$

(à expliciter complètement pour un niveau de signification α).

c) Exprimer la fonction puissance comme une intégrale.

d) Sans aucun calcul, comment doit se situer graphiquement cette fonction puissance par rapport à la fonction puissance asymptotique obtenue à la question 3.f) ?

Exercice 2 (Tests)
Soient $X_1, \cdots X_n$ n observations i.i.d. sur $[0,1]$. On veut tester l'hypothèse H_0 que les X_i sont uniformes sur $[0,1]$. Pour tester cela, on partage $[0,1]$ en trois intervalles égaux : $A_1 = [0, 1/3[$, $A_2 = [1/3, 2/3[$ et $A_3 = [2/3, 1]$ et l'on compte le nombre d'observations $N_j = \sum_{i=1}^{n} \mathbb{1}_{A_j}(X_i)$ pour $j \in [\![1,3]\!]$.

1. Quelle est la loi du vecteur $N = (N_1, N_2, N_3)$ sous H_0 ?

2. Ecrire la statistique du test ξ_n du χ^2 fondée sur N pour tester H_0. Quelle est sa loi lorsque $n \to +\infty$? Donner la région de rejet du test au niveau 5%.

3. Supposons maintenant que l'hypothèse H_0 soit fausse et que la vraie loi des X_i ait la densité :

$$f(x) = (1 - e^{-1})^{-1} e^{-x} \mathbb{1}_{[0,1]}(x)$$

Quelle est alors la loi du vecteur N ? Obtenir la limite de ξ_n/n quand $n \to +\infty$. En déduire que sous cette alternative la puissance du test tend vers 1 lorsque $n \to +\infty$.

4. On veut maintenant tester l'hypothèse H_0 en utilisant le test de Kolmogorov-Smirnov.

4.1. Calculer la fonction de répartition F associée à la densité f et $M = \sup_{x \in [0,1]} |F(x) - x|$.

4.2. La région de rejet du test au seuil α est $W = \{\sqrt{n} \sup_{x \in [0,1]} |F_n(x) - x| > C_\alpha\}$ où F_n est la fonction de répartition empirique des X_i et où C_α est une constante tabulée. En utilisant l'inégalité suivante :

$$\mathbb{P}\big(\sup_{x \in [0,1]} |F(x) - F_n(x)| \geq t \big) \leq 2 \exp(-2nt^2),$$

comment faut-il choisir n pour que la puissance du test en F soit supérieure à 95% ? (on ne résoudra pas l'équation).