ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

Deuxième année
2002-2003

Théorie des tests
Enoncé des travaux dirigés n° 2

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ http://ensae.no-ip.com//SE207/

Exercice 1

☞ Q1 Dans l'exercice 1 du TD 9 du cours d'Estimation et tests, on a testé l'adéquation à un loi de Poisson du nombre d'appels par seconde dans un central téléphonique. Les donnée étaient les suivantes :

Nombre d'appels par seconde	Effectifs observés
0	6
1	15
2	40
3	42
4	37
5	30
6	10
7	9
8	5
9	3
10	2
11	1

La mise en oeuvre du test du χ^2 nécessitant un nombre fini de classes d'effectifs suffisants on a regroupé les modalités de la manière suivante :

$$\begin{cases} J_k = \{k\} \text{ pour } k = 0, 1, \ldots, K-1 \\ J_K = \{K, K+1, \ldots\}, \end{cases}$$

avec $K = 8$. On note n le nombre total d'observations et n_k les effectifs des $K+1$ classes

☞ Q2 Ecrire la statistique $\mathcal{T}_n^2$ du test de l'adéquation à la famille des lois de Poisson (le paramètr étant inconnu).

☞ Q3 On note μ_0 la vraie loi des observations et on pose comme dans le cours

$$\gamma_0 = 1 - \frac{1}{I_0} \sum_{k=0}^{K} \left(\int_{J_k} s_0 d\mu_0 \right)^2 \mu_0^{-1}(J_k)$$

(a) Expliciter, dans le cas où μ_0 est la loi de Poisson de paramètre λ, les quantités

$$q_k = \mu_0(J_k),\ \ s_0(k),\ \ \int_{J_k} s_0\, d\mu_0,\ \ I_0$$

en fonction de k, λ et $p_k(\lambda) = e^{-\lambda} \dfrac{\lambda^k}{k!}$.

(b) Montrer que γ_0 s'écrit aussi

$$\gamma_0 = \lambda \left(\sum_{k \geq K} p_k \right) \left[\frac{\sum_{k \geq K} p_k s_0(k)^2}{\sum_{k \geq K} p_k} - \left(\frac{\sum_{k \geq K} p_k s_0(k)}{\sum_{k \geq K} p_k} \right)^2 \right].$$

(c) Interpréter le terme entre crochets comme une variance. En déduire que $\gamma_0 > 0$.

☞ Q4 (a) Rappeler l'expression de la distribution limite sous l'hypothèse nulle de $\mathcal{T}_n^2$ en fonction de γ_0.
Quelle approximation le corrigé du TD 9 faisait-il implicitement ? Quelle conséquence l'erreur commise a-t-elle sur le niveau du test ?

(b) Estimer γ_0 à l'aide d'une calculatrice programmable ou d'un logiciel de calcul numérique (on utilisera les valeurs $K = 8$ et $\lambda = \hat{\lambda}_{\text{emv}} = 3,7$).
L'approximation faite était-elle justifiée ?

Exercice 2

On appelle loi de Pareto de paramètres $a > 0$ et $\gamma > 0$, et on note Pareto(a, γ), la mesure de probabilité sur l'ensemble $\mathbf{R}$ des réels dont la fonction de répartition $F_{a,\gamma}$ est définie par $F_{a,\gamma}(x) = 0$ pour $x < a$ et

$$F_{a,\gamma}(x) = 1 - \left(\frac{x}{a}\right)^{-\gamma} \quad \text{pour} \quad x \geq a$$

☞ Q1 On suppose que a est connu. On peut alors se ramener au cas $a = 1$, ce que l'on fera ici. Soit $X_1, \ldots, X_n$ un échantillon de variables aléatoires indépendantes et de même loi (iid) Pareto$(1, \gamma)$.

(a) Soient $\gamma_1 > 0$ et $\gamma_0 > 0$ deux valeurs distinctes du paramètre γ. On souhaite tester l'hypothèse simple $H_0 = [\gamma = \gamma_0]$ contre l'alternative simple $H_1 = [\gamma = \gamma_1]$.
Montrer que le test le plus puissant de niveau $0 < \alpha < 1$ de H_0 contre H_1 a une région de rejet de la forme $n^{-1} \sum_{i=1}^n \ln X_i > c_{\alpha,n}$ ou de la forme $n^{-1} \sum_{i=1}^n \ln X_i < c_{\alpha,n}$.
Discuter.

(b) Déterminer, sous H_0, l'espérance et la variance de la variable aléatoire $\ln X_1$. Déduire du théorème central limite une approximation de $c_{\alpha,n}$ utilisable lorsque n est assez grand.

(c) Quelle est, sous H_0, la loi de $\ln X_1$?
De quelle forme est la loi de $\sum_{i=1}^n \ln X_i$?

(d) Toujours dans le cas $a = 1$ connu, déterminer le test uniformément le plus puissant (UPP) de niveau α de H_0 : $[\gamma \leq \gamma_0]$ contre $H_1 = [\gamma > \gamma_0]$.

☞ Q2 Soit $X_1, \ldots, X_n$ un échantillon de variables aléatoires iid.

(a) On suppose encore $a = 1$ connu. On note γ_0 la vraie valeur de γ sous H_0, et on not $F_0 = F_{1,\gamma_0}$.

i. Calculer la fonction inverse $F_0^{-1}(\cdot)$.

ii. Déterminer l'estimateur du maximum de vraisemblance $\hat{\gamma}_n$ basé sur l'échantillon

iii. Déterminer la fonction de score $s_0(x) = s(x; \gamma_0)$ et l'information de Fisher $I_0 =$ $I(\gamma_0)$.

iv. En déduire une expression de la forme

$$n^{1/2}\,(\hat{\gamma}_n - \gamma_0) = n^{-1/2} \sum_{i=1}^n \phi_0(X_i) + o_P(1)$$

sous H_0, en précisant la fonction ϕ_0

v. Quelle est la limite en distribution de $n^{1/2}(\hat{\gamma}_n - \gamma_0)$ sous H_0, quand $n \to \infty$?

(b) Déduire des questions précédentes l'expression de la fonction de covariance du proces sus gaussien centré $\hat{\mathbf{B}}(\cdot)$, limite en distribution sous H_0, quand $n \to \infty$, de la suite d processus

$$\hat{\mathbf{B}}_n(t) = n^{\frac{1}{2}} \left[(\mathbf{F}_n - F_{1,\hat{\gamma}_n}) \circ F_0^{-1} \right](t), \ t \in [0, 1]$$

où $\mathbf{F}_n(\cdot)$ désigne la fonction de répartition empirique.

(c) Comparer le résultat obtenu à celui que l'on obtient lorsque l'on teste l'adéquatio du modèle exponentiel, et en déduire les valeurs critiques, pour $\alpha = 0.10$, 0.05 0.025 et 0.01, des tests de Kolmogorov-Smirnov et de Cramér-von Mises pour tes ter l'adéquation du modèle Pareto $(1, \gamma)$.

(d) Mettre en application du test d'adéquation du chi-deux pour tester l'adéquation d Pareto$(1, \gamma)$, avec γ inconnu.

☞ Q3 On se propose de tester l'adéquation de Pareto(a, γ), lorsque a et γ sont tous deux inconnus

(a) Soit $U_1, \ldots, U_n$ un échantillon de va iid uniformes sur $[0, 1]$, et soit $U_{(1,n)}$ la plus petit observation de cet échantillon. Montrer que $nU_{(1,n)}$ converge en distribution vers la lo exponentielle de moyenne 1 quand $n \to \infty$.

(b) Montrer que si $(A_j)_{j \geq 1}$ est une suite quelconque d'événements et si $\sum_{j=1}^{+\infty} P(A_j$ converge, alors presque sûrement seuls un nombre fini de A_j ont lieu.

(c) Proposer une test de l'adéquation de Pareto(a, γ).

Exercice 3

On rappelle la définition de l'inverse généralisée d'une fonction de répartition F :

$$F^{\text{inv}}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$$

☞ Q1 Soit X une variable aléatoire suivant la loi de Poisson de paramètre $\lambda > 0$. Déterminer et représenter graphiquement sa fonction de répartition F_λ et l'inverse généralisée F_λ^{inv} de F_λ.

☞ Q2 (a) Pour toute f.d.r. F, vérifier l'équivalence :

$$u \leq F(x) \Longleftrightarrow F^{\text{inv}}(u) \leq x$$

(b) Soit U une v.a. uniforme sur $[0, 1]$. On pose :

$$Y = F_\lambda^{\text{inv}}(U)$$

Montrer que Y suit la même loi que X.

(c) Expliquer par quel algorithme on construit Y à partir de U.

☞ Q3 Déterminer la f.d.r. G_λ de la v.a. $Z = F_\lambda(Y)$.
Cette v.a. est-elle uniforme sur $[0, 1]$? La f.d.r. G_λ dépend-elle de λ ?

☞ Q4 Peut-on utiliser les tests de Kolmogorov-Smirnov ou de Cramer-von Mises pour tester l'adéquation au modèle de Poisson ?