Université d'Orléans - Licence Economie et Gestion
Statistique Mathématique

C. Hurlin. Examen Terminal Janvier 2010

Exercice 1 *Maximum de Vraisemblance et loi exponentielle. Barême : 5 points*

Une société de vente à distance cherche à étudier la durée, notée D, qui s'écoule entre deux appels sur une plateforme téléphonique. On suppose que D suit une loi exponentielle de paramètre θ admettant pour fonction de densité :

$$f_D\left(d;\theta\right)=\frac{1}{\theta}\exp\left(-\frac{d}{\theta}\right),\quad \forall d\geq 0, \tag{1}$$

Elle dispose pour cela d'un N-échantillon aléatoire $\{D_1,..,D_N\}$ *i.i.d.* de même loi que D. On admet que les deux premiers moments de la v.a.r. D sont définis par :

$$E\left(D\right)=\theta \qquad V\left(D\right)=\theta^2. \tag{2}$$

On suppose que le paramètre θ est inconnu et l'on se propose de l'estimer par la méthode du maximum de vraisemblance.

Question 1 (1 point) Etant donnée la définition de la variable D, quel contrainte doit satisfaire le coefficient θ ?

Question 2 (1.5 points) Montrez que la log-vraisemblance associée au N-échantillon $\{D_1,..,D_N\}$ s'écrit :

$$\ln L\left(D_1,..,D_N;\theta\right)=-N\ln\left(\theta\right)-\frac{1}{\theta}\sum_{i=1}^{N}D_i. \tag{3}$$

Question 3 (1.5 points) Montrez que l'estimateur du maximum de vraisemblance du paramètre θ, noté $\widehat{\theta}$, correspond à la moyenne empirique de l'échantillon, *i.e.* $\overline{D}_N=\left(1/N\right)\sum_{i=1}^{N}d_i$.

Question 4 (1 point) Montrez que l'estimateur du maximum de vraisemblance est sans biais et convergent.

Exercice 2 *Tests paramétriques: Barème : 10 points*

Une société cherche à établir le profil des ventes d'un produit A en fonction des ventes d'autres produits de sa gamme. Elle considère pour cela un modèle de régression linéaire liant le logarithme des ventes mensuelles du produit A, notées q_A, au logarithme des ventes mensuelles d'un produit B, notées q_B. Ce modèle de régression s'écrit sous la forme suivante :

$$q_A=a+b\ q_B+\varepsilon \tag{4}$$

où ε désigne un terme d'erreur aléatoire, et où a et b sont deux paramètres réels. A partir d'un échantillon de T observations $\{q_{A,1}, .., q_{A,T}\}$ et $\{q_{B,1}, .., q_{B,T}\}$, la direction commerciale souhaite effectuer le test suivant :

$$H_0 : b = 1 \tag{5}$$

$$H_1 : b > 1 \tag{6}$$

On admet que la région critique du test UPP de niveau α de l'hypothèse nulle $H_0 : b = 1$ contre $H_1 : b > 1$ est de la forme :

$$W = \{ q_{A,1}, .., q_{B,T} \mid t_b \geq A\} \tag{7}$$

où A est une constante déterminée par le risque de première espèce et où t_b désigne une statistique (dite statistique de Student) dont la loi sous H_0 est supposée connue et telle que :

$$t_b \underset{H_0}{\rightsquigarrow} t\,(T-1) \tag{8}$$

où $t\,(T-1)$ désigne la loi de Student à $T-1$ degrés de liberté.

Question préliminaire A (0.5 point) Quelle est la signification économique du paramètre a ?

Remarque : on rappelle que si

$$\log(Y) = c + d\log(X) \tag{9}$$

alors :

$$\frac{\partial Y}{\partial X} \times \frac{X}{Y} = d \tag{10}$$

Question préliminaire B (0.5 point) Commentez la forme de la région critique.

Question 1 (1 point) Déterminez le seuil critique A en fonction du risque de première espèce α, de la taille de l'échantillon T, et de la fonction de répartion de la loi de Student $t\,(T-1)$, notée $G_{T-1}(x)$, $\forall x \in \mathbb{R}$.

Question 2 (1 point) On suppose que la société observe une réalisation de la statistique de Student t_b égale à 2.34 dans un échantillon comportant $T = 30$ observations de ventes mensuelles. Quelle conclusion doit-elle tirer quant à la valeur de b pour un niveau de risque α de 5% ?

Question 3 (1 point) A partir des seules hypothèses fournies, est-il possible d'exprimez la fonction puissance de ce test ? *Justifiez précisèment votre réponse.*

Question 4 (1 point) A partir des données historiques de ventes du produit B et de trajectoires tirées au hasard pour les résidus ε, on simule 100 pseudos-échantillons de 30 valeurs de ventes q_A pour chacune des valeurs suivantes du paramètre b : $b = 1.5$, $b = 2$ et $b = 3$. Pour chaque pseudo-échantillon, on calcule la statistique de Student t_b. Dans le tableau ci-dessous, les 100 valeurs de la statistique t_b sont réparties par rapport à la valeur du seuil critique A pour un niveau de risque de 10%, à savoir $A = 1.311$. A partir de ces informations, tracez la fonction puissance du test pour un niveau de risque de première espèce de 10%.

Tableau 1 : Puissance empirique

	$b = 1.5$	$b = 2$	$b = 3$
Nombre d'échantillons avec $t_b \geq 1.311$	17	36	95
Nombre d'échantillons avec $t_b < 1.311$	83	64	5

Question 5 (2 points) On considère finalement le test bilatéral :

$$H_0 : b = 1 \tag{11}$$

$$H_1 : b \neq 1 \tag{12}$$

Déterminez la région d'acceptation de ce test en fonction du risque de première espèce α, de la taille de l'échantillon T, et de la fonction de répartion de la loi $t\,(T-1)$, notée $G_{T-1}\,(x)$, $\forall x \in \mathbb{R}$.

Question 6 (1 point) Si l'on suppose que la société dispose d'un échantillon de très grande taille ($T \to \infty$), sous quelles conditions sur la valeur absolue de la statistique de Student t_b, pourra-t-on dire que le paramètre b est significativement différent de 1 (autrement dit, que l'on rejette $H_0 : b = 1$) au seuil de risque de 5% ?

Question 7 (2 points) On suppose que la société dispose d'un échantillon de très grande taille ($T \to \infty$) et que la réalisation de la statistique de Student t_b est égale à -1.645. A partir de quel seuil de risque de première espèce α pourrait-on conclure au non rejet de l'hypothèse nulle $H_0 : b = 1$ contre une alternative $H_1 : b \neq 1$? *Remarque : montrez que ce seuil de risque vaut :*

$$\alpha = 2 \times [1 - G_\infty\,(|t_b|)] \tag{13}$$

où $G_\infty\,(x)$ désigne la fonction de répartition de la de Student avec un degré de liberté infini (loi normale).

Exercice 3 *Test d'Adéquation : Barème : 5 points.*

On souhaite tester si une variable X suit une loi binomiale négative de paramètre n et p, notée $\text{NegBin}(n,p)$. On rappelle que la loi binomiale négative est une distribution de probabilité discrète. Elle décrit la situation suivante: une expérience consiste en une série de tirages indépendants, donnant un succès avec probabilité p et un échec avec une probabilité complémentaire. Cette expérience se poursuit jusqu'à l'obtention d'un nombre donné n de succès. La variable aléatoire représentant le nombre d'échecs (avant l'obtention du nombre donné n de succès) suit alors une loi binomiale négative[1]. Ainsi, si X suit une loi $\text{NegBin}(n,p)$, avec $n > 0$ et $0 < p < 1$, alors la probabilité que X soit égale à k est égale à :

$$\Pr(X = k) = C_{n+k-1}^{k} \times p^n \times (1-p)^k\,, \quad \forall k \in \{0, 1, 2, ...\}\,, \tag{14}$$

où le coefficient binomial C_{n+k-1}^{k} est défini par :

$$C_{n+k-1}^{k} = \frac{(k+n-1)!}{k!\,(n-1)!}. \tag{15}$$

On considère un échantillon de 100 variables aléatoires indépendantes $\{X_1, ..X_N\}$ de même loi que X. Pour ces 100 variables, on observe les effectifs empiriques suivants :

X_i	0	1	2	3	4	5	6	7	Total
Nombres d'observations	16	18	19	16	11	9	6	5	100

Testez au seuil de 5%, l'hypothèse selon laquelle la variable X suit une loi binomiale négative de paramètres $p = 0.5$ et $n = 3$.

[1] Source : Wikipedia. Article "loi binomiale négative".