Université d'Orléans - Licence Economie et Gestion Statistique Mathématique

C. Hurlin. Examen Terminal Janvier 2007

Exercice 1 *Tests Paramétriques et Intervalle de Confiance. Barème : 12 points.*

On considère le rendement d'un actif financier à la date t, noté X. On suppose que la v.a.r. X suit une loi normale de moyenne m inconnue et de variance σ^2 elle aussi inconnue. On considère un échantillon de $N = 16$ observations historiques notées $\{X_1, .; , X_N\}$ supposées *i.i.d.* de même loi que X. On note S_N^2 la variance empirique corrigée associée à cet échantillon telle que :

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \overline{X}_N\right)^2 \tag{1}$$

où $\overline{X}_N = (1/N) \sum_{i=1}^{N} X_i$ désigne la moyenne empirique. On admet que :

$$(N-1) \frac{S_N^2}{\sigma^2} \sim \chi^2 (N-1) \tag{2}$$

On suppose que sur cet échantillon, on observe $S_N^2 = 4$ et $\overline{X}_N = 1.5$.

Question 1 (1.5 point) Montrez que sous les hypothèses énoncées ci-dessus, la variable centrée réduite z, définie par :

$$z = \frac{\overline{X}_N - m}{S_N / \sqrt{N}} \tag{3}$$

suit une loi de Student à $N - 1$ degrés de liberté.

Question 2 (1 point) On souhaite tester la nullité des rendements moyens associés à cet actif. (i) Proposez un test permettant de traduire au mieux ce souhait et (ii) donnez une interprétation économique au risque de première espèce.

Question 3 (2 points) On débute par un test unilatéral de la forme :

$$H_0 : m = 0 \tag{4}$$

$$H_1 : m > 0 \tag{5}$$

Afin de réaliser ce test, on considère une région critique de la forme :

$$W = \left\{ X_1, .., X_N | \, \overline{X}_N > K \right\} \tag{6}$$

où K est une constante déterminée par le niveau de risque de première espèce égal à $\alpha\%$.
(i) Montrez que :

$$K = F^{-1} (1 - \alpha) \frac{S_N}{\sqrt{N}} \tag{7}$$

où $F(.)$ désigne la fonction de répartition de la loi de Student à $N - 1$ degrés de liberté.
(ii) Ce seuil critique est-il le même pour deux échantillons différents de rendements ? Cela pose-t-il un problème particulier ?

Question 4 (2 points) On considère le même test unilatéral $H_0 : m = 0$ contre $H_1 : m > 0$, mais en utilisant cette fois une statistique, dite statistique de Student, et une région critique de la forme :

$$W = \left\{ X_1, .., X_N | \ \frac{\overline{X}_N}{S_N / \sqrt{N}} > A \right\} \tag{8}$$

où A est une constante déterminée par le niveau de risque de première espèce égal à $\alpha\%$. (i) Déterminez la valeur du seuil critique A et (ii) concluez quant à la nullité des rendements moyens pour un risque de première espèce de 5% ?

Question 5 (2 points) On considère à présent le test bilatéral $H_0 : m = 0$ contre $H_1 : m \neq 0$. (i) Montrez que la région critique de ce test fondée sur la statistique de Student $z = \overline{X}_N / \left(S_N / \sqrt{N} \right)$ pour un risque de première espèce de $\alpha\%$ est égale à :

$$W = \left\{ X_1, .., X_N | \ \frac{\overline{X}_N}{S_N / \sqrt{N}} \notin \left[F^{-1}\left(\frac{\alpha}{2}\right) \ ; \ F^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \right\} \tag{9}$$

où $F(.)$ désigne la fonction de répartition de la loi de Student à $N - 1$ degrés de liberté et (ii) concluez quant à l'hypothèse de nullité des rendements moyens pour un risque de première espèce $\alpha = 5\%$.

Question 6 (1.5 points) Construisez un intervalle de confiance sur la vraie valeur de m pour un niveau de risque de $\alpha\%$ et vérifiez que pour un risque de $\alpha = 5\%$, cet intervalle est défini par les bornes :

$$\Pr[0.4343 < m < 2.5657] = 0.95 \tag{10}$$

Vous détaillerez précisément votre démarche.

Question 7 (2 points) On admet que

$$IC_{1-\alpha} = \left[\overline{X}_N - \frac{S_N}{\sqrt{N}} F^{-1}\left(1 - \frac{\alpha}{2}\right) ; \overline{X}_N - \frac{S_N}{\sqrt{N}} F^{-1}\left(\frac{\alpha}{2}\right) \right] \tag{11}$$

(i) Montrez que si cet intervalle de confiance ne couvre pas la valeur nulle $(0 \notin IC_{1-\alpha})$, alors cela implique que pour un risque de première espèce de $\alpha\%$ vous êtes conduit à rejeter l'hypothèse nulle $H_0 : m = 0$ contre $H_1 : m \neq 0$. (ii) Que conclure quant à la nullité de m au vu de l'intervalle obtenu dans la question 6 ?

Exercice 2 *Tests d'Indépendance : le cas de la Presse. Barème : 4 points. D'après un exercice fourni par Mme Bessec (Université Paris IX dauphine)*

Un institut d'études est chargé d'identifier les causes de la crise de la presse écrite française. Souhaitant analyser la liaison entre la lecture de la presse (variable X) et l'âge des individus (variable Y), l'institut a contacté 500 personnes et a constitué le tableau de contingence suivant. Testez au seuil de 10% l'hypothèse nulle l'indépendance des deux variables. *Vous détaillerez précisément votre démarche.*

X \ Y	- de 25 ans	25-50 ans	50 ans et +	Total
rare	40	40	20	100
occasionnelle	75	125	50	250
fréquente	40	60	50	150
Total	155	225	120	500

Exercice 3 *Maximum de Vraisemblance . Barème : 6 points*

On considère une variable aléatoire X à valeur dans $\mathbb{N}$ supposée suivre une loi géométrique de paramètre $\theta > 1$. On rappelle que la fonction de densité de X est définie par :

$$f(x;\theta) = \begin{cases} \left(1 - \frac{1}{\theta}\right)^{x-1} \frac{1}{\theta} & \text{si } x \in \{1,2,3,..\} \\ 0 & \text{sinon} \end{cases} \tag{12}$$

On considère un N-échantillon $\{X_1, .., X_N\}$ où les variables X_i sont supposées *i.i.d.* de même loi que X. On admet que :

$$E(X) = \theta \qquad V(X) = \theta(\theta - 1) \tag{13}$$

Question 1 (1.5 point) Montrez que la log-vraisemblance de l'échantillon $\{X_1, .., X_N\}$ peut s'écrire sous la forme :

$$\log L(X_1, .., X_N; \theta) = \left(\sum_{i=1}^{N} x_i - N\right) \log(\theta - 1) - \sum_{i=1}^{N} x_i \log(\theta) \tag{14}$$

Question 2 (2 points) Montrez que l'estimateur du maximum de vraisemblance du paramètre θ, noté $\widehat{\theta}$, correspond à la moyenne empirique $\overline{X}_N = (1/N) \sum_{i=1}^{N} X_i$:

$$\widehat{\theta} = \overline{X}_N \tag{15}$$

Conseil : pour la condition suffisante calculez $\partial^2 \log L(X_1, .., X_N; \theta) / \partial\theta^2$, *puis évaluez cette quantité au point optimal* $\widehat{\theta} = \overline{X}_N$.

Question 3 (1.5 point) Montrez que cet estimateur $\widehat{\theta}$ est sans biais et convergent (en probabilité).

Question 4 (1 point) On admet que :

$$\left.\frac{\partial^2 \log L\left(X_1, .., X_N; \theta\right)}{\partial \theta^2}\right|_{\widehat{\theta}=\overline{X}_N} = \frac{-N}{\overline{X}_N\left(\overline{X}_N - 1\right)} \tag{16}$$

Montrez que l'estimateur du maximum de vraisemblance est efficace. *Rappel : la quantité d'information de Fisher peut être définie par :*

$$I_N = E\left[-\left.\frac{\partial^2 \log L\left(X_1, .., X_N; \theta\right)}{\partial \theta^2}\right|_{\widehat{\theta}}\right] \tag{17}$$

Bon courage