

Université d'Orléans - Licence Economie et Gestion

Statistique Mathématique

C. Hurlin. Correction de l'Examen Terminal Janvier 2007

Exercice 1 *Tests Paramétriques.*

Question 1 (1.5 point) On sait que $(N - 1) S_N^2 / \sigma^2$ suit un $\chi^2(N - 1)$. Par ailleurs, la variable correspond à une somme de variables normales *i.i.d.* et donc suit elle aussi une loi normale de paramètres (0.5 point) :

$$E(\bar{X}_N) = m \quad V(\bar{X}_N) = \sigma^2 / N \quad (1)$$

Par ailleurs, si l'on suppose que les deux variables $(N - 1) S_N^2 / \sigma^2$ et $(\bar{X}_N - m) / \sigma \sqrt{N}$ sont indépendantes (0.5 point), alors la variable transformée z vérifie :

$$\frac{(\bar{X}_N - m) / \sigma \sqrt{N}}{\sqrt{\frac{(N-1)S_N^2 / \sigma^2}{N-1}}} \sim Student(N - 1) \quad 0.5 \text{ point}$$

ou de façon équivalente (0.5 point) :

$$\frac{\bar{X}_N - m}{S_N / \sqrt{N}} \sim Student(N - 1) \quad (2)$$

Question 2 (1 point) On peut traduire ce souhait par un test bilatéral de la forme :

$$H_0 : m = 0 \quad (3)$$

$$H_1 : m \neq 0 \quad (4)$$

Dans ce test, le risque de première espèce correspond alors à la probabilité de rejeter l'hypothèse de nullité des rendements moyens alors que les rendements moyens de l'actif sont effectivement nuls.

Question 3 (2 points) On sait que :

$$W = \{X_1, \dots, X_N \mid \bar{X}_N > K\} \quad (5)$$

Or sous H_0 , on a $m = 0$ et donc :

$$\frac{\bar{X}_N}{S_N / \sqrt{N}} \sim Student(N - 1) \quad 0.5 \text{ point} \quad (6)$$

Dès lors, par définition du risque I , on a (0.5 point) :

$$\alpha = \Pr \left[\bar{X}_N > K \mid \frac{\bar{X}_N}{S_N / \sqrt{N}} \sim Student(N - 1) \right] \quad (7)$$

ou encore :

$$\alpha = \Pr \left[\frac{\bar{X}_N}{S_N / \sqrt{N}} > \frac{K}{S_N / \sqrt{N}} \mid \frac{\bar{X}_N}{S_N / \sqrt{N}} \sim Student(N - 1) \right] \quad (8)$$

D'où l'on tire que (0.5 point) :

$$K = F^{-1}(1 - \alpha) \frac{S_N}{\sqrt{N}} \quad (9)$$

où $F(\cdot)$ désigne la fonction de répartition de la loi de Student à $N - 1$ degrés de liberté. Le seuil critique de ce test dépend de l'échantillon via la quantité S_N (0.5 point).

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

$$W = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} > A \right\} \quad (10)$$

où A est une constante déterminée par le niveau de risque de première espèce égal à $\alpha\%$. Par définition, on a :

$$\alpha = \Pr \left(\frac{\bar{X}_N}{S_N/\sqrt{N}} > A \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} \sim Student(N - 1) \right) \quad (0.5point) \quad (11)$$

Donc :

$$A = F^{-1}(1 - \alpha) \quad (12)$$

Ici, on a pour $N - 1 = 15$:

$$A = F^{-1}(0.95) = 1.7531 \quad (13)$$

et donc :

$$W = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} > 1.7531 \right\} \quad (0.5 point)$$

Or ici, on a :

$$\frac{\bar{X}_N}{S_N/\sqrt{N}} = \frac{1.5}{0.2/\sqrt{16}} = 3 \quad (0.5 point) \quad (14)$$

La réalisation de la statistique de Student appartient à la RC W , donc pour un risque de première espèce de 5%, on rejete l'hypothèse de nullité des rendements moyens contre une alternative de rendements strictement positifs. (0.5 point)

Question 5 (1 point) On considère à présent le test bilatéral $H_0 : m = 0$ contre $H_1 : m \neq 0$. (i). La région d'acceptation (RA) de ce test fondée correspond à l'intersection des deux RA des deux tests unilatéraux $H_0 : m = 0$ contre $H_1 : m > 0$ et $H_0 : m = 0$ contre $H_1 : m < 0$, c'est à dire (0.5 point) :

$$H_0 : m = 0 \text{ contre } H_1 : m > 0 : \bar{W} = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} < B \right\} \quad (15)$$

$$H_0 : m = 0 \text{ contre } H_1 : m < 0 : \bar{W} = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} > A \right\} \quad (16)$$

Toutefois les seuils A et B doivent être déterminés pour un niveau de risque de $\alpha/2\%$ afin de garantir un risque I de $\alpha\%$ pour le test bilatéral. Soient (0.5 point) :

$$B = F^{-1}(1 - \alpha/2) \quad (17)$$

$$A = F^{-1}(\alpha/2) \quad (18)$$

Dès lors, l'intersection des deux RA nous donne pour le test bilatéral une RA définie par :

$$\bar{W} = \left\{ X_1, \dots, X_N \mid A < \frac{\bar{X}_N}{S_N/\sqrt{N}} < B \right\} \quad (19)$$

ou encore

$$\bar{W} = \left\{ X_1, \dots, X_N \mid F^{-1}(\alpha/2) < \frac{\bar{X}_N}{S_N/\sqrt{N}} < F^{-1}(1 - \alpha/2) \right\} \quad (20)$$

Dans notre échantillon, cela donne ($N = 16$) une RC de la forme :

$$W = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} \notin \left[F^{-1}\left(\frac{\alpha}{2}\right) ; F^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \right\} \quad (21)$$

ou encore

$$W = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} \notin [-2.1314 ; 2.1314] \right\} \quad 0.5 \text{ point} \quad (22)$$

Puisque ici, on a :

$$\frac{\bar{X}_N}{S_N/\sqrt{N}} = \frac{1.5}{0.2/\sqrt{16}} = 3 \quad (23)$$

La réalisation de la statistique de Student appartient à la RC W , donc pour un risque de première espèce de 5%, on rejete l'hypothèse de nullité des rendements moyens contre une alternative de rendements strictement positifs. (0.5 point).

Question 6 (1.5 points) On sait que

$$\frac{\bar{X}_N - m}{S_N/\sqrt{N}} \sim Student(N - 1) \quad (24)$$

Dès lors on a par définition :

$$\Pr \left[t_{\alpha/2} < \frac{(\bar{X}_N - m)}{S_N/\sqrt{N}} < t_{1-\alpha/2} \right] = 1 - \alpha \quad (25)$$

où $t_{\alpha/2}$ et $t_{1-\alpha/2}$ désignent respectivement les fractiles d'ordre $\alpha/2$ et $1-\alpha/2$ de la loi $Student(N - 1)$. On en déduit finalement (0.5 point):

$$IC_{1-\alpha} = \left[\bar{X}_N - \frac{S_N}{\sqrt{N}} t_{1-\alpha/2}; \bar{X}_N - \frac{S_N}{\sqrt{N}} t_{\alpha/2} \right] \quad (26)$$

puisque la loi de Student est symétrique. Application Numérique : $S_N^2 = 4$ et $\bar{X}_N = 1.5$, $N = 16$, $\alpha = 5\%$

$$t_{1-\alpha/2} = 2.131 \quad (27)$$

D'om l'on tire que :

$$IC_{0.95} = \left[1.5 - \frac{2}{\sqrt{16}} 2.131 ; 1.5 - \frac{2}{\sqrt{16}} 2.131 \right] = [0.4343; 2.5657] \quad (28)$$

On remarque que cet intervalle ne comprend pas la valeur nulle.

Question 7 (2 points) On admet que

$$IC_{1-\alpha} = \left[\bar{X}_N - \frac{S_N}{\sqrt{N}} F^{-1} \left(1 - \frac{\alpha}{2} \right); \bar{X}_N - \frac{S_N}{\sqrt{N}} F^{-1} \left(\frac{\alpha}{2} \right) \right] \quad (29)$$

Supposons que cet intervalle de confiance ne couvre pas la valeur nulle ($0 \notin IC_{1-\alpha}$), alors cela implique que

$$\bar{X}_N - \frac{S_N}{\sqrt{N}} F \left(1 - \frac{\alpha}{2} \right) > 0 \quad (30)$$

Dès lors, :

$$\frac{\bar{X}_N}{S_N/\sqrt{N}} > F \left(1 - \frac{\alpha}{2} \right) \quad 1 \text{ point} \quad (31)$$

Or, rappelons que la RC du test bilatéral est :

$$W = \left\{ X_1, \dots, X_N \mid \frac{\bar{X}_N}{S_N/\sqrt{N}} \notin \left[F^{-1} \left(\frac{\alpha}{2} \right) ; F^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \right\} \quad (32)$$

Donc si $\bar{X}_N / (S_N/\sqrt{N}) > F(1 - \frac{\alpha}{2})$, alors $\bar{X}_N / (S_N/\sqrt{N})$ appartient à la RC, ce qui conduit à rejeter H_0 . On doit donc rejeter l'hypothèse nulle de nullité des rendements (1 point). A l'inverse, si l'intervalle de confiance à $1-\alpha\%$ couvre la valeur nulle on est conduit à ne pas rejeter $H_0 : m = 0$ pour un risque I de $\alpha\%$. Au vu de l'IC de la question 6, on conclut au rejet de $H_0 : m = 0$ (+ 1 point bonus)

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)

- 1 point pour le tableau de contingence
- 0.5 point pour la condition de regroupement
- 1 point pour la statistique
- 0.5 point pour le seuil critique
- 1 pour la conclusion

Figure 1:

- Tableau des effectifs théoriques

Si les 2 variables sont indépendantes, $p_{ij} = p_i \times p_j = \frac{n_i}{n} \times \frac{n_j}{n} \Leftrightarrow T_{ij} = np_{ij} = \frac{n_i \times n_j}{n}$. D'où :

X \ Y	- de 25 ans	25-50 ans	50 ans et +	Total
rare	31	45	24	100
occasionnelle	77,5	112,5	60	250
fréquente	46,5	67,5	36	150
Total	155	225	120	500

Tous les effectifs théoriques sont supérieurs à 4. On peut donc utiliser la loi du chi-deux pour effectuer un test d'indépendance entre les 2 variables.

- Hypothèse : H_0 : indépendance des variables X et Y
- Statistique et loi sous H_0 :

$$\Delta = \sum_i \sum_j \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \xrightarrow{H_0} \chi_{(i-1)(j-1)}^2 = \chi_4^2 \text{ car } T_{ij} \geq 4 \quad \forall i, j \text{ (voir ci-dessus)}$$

- Règle de décision :

W tel que $\Delta > c$

C : CPE $\rightarrow c = \chi_{4,\alpha}^2 = 9.49$ au seuil $\alpha = 5\%$.

- Conclusion :

$$\text{Ici, } \Delta_{obs} = \frac{(40 - 31)^2}{31} + \dots + \frac{(50 - 36)^2}{36} = 14.16 > c \quad \forall \alpha$$

On rejette donc H_0 aux deux seuils. Les deux variables ne sont pas indépendantes. On retrouve donc à ce niveau les résultats de la question 2.

Exercice 3 *Maximum de Vraisemblance . Barème : 5 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 :

$$f(x; \theta) = \begin{cases} (1 - \frac{1}{\theta})^{x-1} \frac{1}{\theta} & \text{si } x \in \{1, 2, 3, \dots\} \\ 0 & \text{sinon} \end{cases} \quad (33)$$

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

Question 1 (1.5 point) La vraisemblance s'écrit de la façon suivante :

$$L(X_1, \dots, X_N; \theta) = \prod_{i=1}^N f(x_i; \theta) = \frac{1}{\theta^N} \prod_{i=1}^N \left(1 - \frac{1}{\theta}\right)^{x_i-1} \quad (34)$$

Ce qui peut se réécrire sous la forme :

$$L(X_1, \dots, X_N; \theta) = \prod_{i=1}^N f(x_i; \theta) = \frac{1}{\theta^N} \left(1 - \frac{1}{\theta}\right)^{\sum_{i=1}^N x_i-1} \quad (35)$$

La log-vraisemblance devient :

$$\begin{aligned} \log L(X_1, \dots, X_N; \theta) &= -N \log(\theta) + \sum_{i=1}^N (x_i - 1) \log\left(\frac{\theta - 1}{\theta}\right) \\ &= -N \log(\theta) + \sum_{i=1}^N (x_i - 1) \log(\theta - 1) - \sum_{i=1}^N (x_i - 1) \log(\theta) \end{aligned} \quad (36)$$

En simplifiant, il vient :

$$\begin{aligned} \log L(X_1, \dots, X_N; \theta) &= -N \log(\theta) + \sum_{i=1}^N (x_i - 1) \log(\theta - 1) - \sum_{i=1}^N x_i \log(\theta) + N \log(\theta) \\ &= \left(\sum_{i=1}^N x_i - N\right) \log(\theta - 1) - \sum_{i=1}^N x_i \log(\theta) \end{aligned} \quad (37)$$

Question 2 (2 points) Par définition :

$$\hat{\theta} = \arg \max \log L(X_1, \dots, X_N; \theta)$$

Les conditions CN et CS de la maximisation de la vraisemblance sont les suivantes.

$$\frac{\partial \log L(X_1, \dots, X_N; \theta)}{\partial \theta} = \left(\frac{1}{\theta - 1}\right) \left(\sum_{i=1}^N x_i - N\right) - \frac{1}{\theta} \sum_{i=1}^N x_i = 0 \quad (38)$$

En réarrangeant les termes, il vient :

$$\theta \sum_{i=1}^N x_i - \theta N - (\theta - 1) \sum_{i=1}^N x_i = 0 \quad (39)$$

$$\iff \theta \sum_{i=1}^N x_i - \theta N - \theta \sum_{i=1}^N x_i + \sum_{i=1}^N x_i = 0 \quad (40)$$

D'où l'on tire que (1 point) :

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{X}_N \quad (41)$$

On vérifie que c'est un maximum puisque (1 point) :

$$\begin{aligned} \frac{\partial^2 \log L(X_1, \dots, X_N; \theta)}{\partial \theta^2} &= -\frac{1}{(\theta - 1)^2} \left(\sum_{i=1}^N x_i - N \right) + \frac{1}{\theta^2} \sum_{i=1}^N x_i \\ &= -\frac{1}{(\theta - 1)^2} (N \bar{X}_N - N) + \frac{1}{\theta^2} N \bar{X}_N \\ &= \frac{1}{\theta^2 (\theta - 1)^2} \left[-\theta^2 (N \bar{X}_N - N) + N \bar{X}_N (\theta - 1)^2 \right] \\ &= \frac{1}{\theta^2 (\theta - 1)^2} \left[-\theta^2 N \bar{X}_N + N \theta^2 + N \bar{X}_N \theta^2 - 2N \bar{X}_N \theta + N \bar{X}_N \right] \\ &= \frac{1}{\theta^2 (\theta - 1)^2} \left[N \theta^2 - 2N \bar{X}_N \theta + N \bar{X}_N \right] \end{aligned} \quad (42)$$

Or au point optimal $\hat{\theta} = \bar{X}_N$, donc on a :

$$\begin{aligned} \left. \frac{\partial^2 \log L(X_1, \dots, X_N; \theta)}{\partial \theta^2} \right|_{\hat{\theta} = \bar{X}_N} &= \frac{1}{\bar{X}_N^2 (\bar{X}_N - 1)^2} \left[N \bar{X}_N^2 - 2N \bar{X}_N^2 + N \bar{X}_N \right] \\ &= \frac{1}{\bar{X}_N^2 (\bar{X}_N - 1)^2} \left[-N \bar{X}_N^2 + N \bar{X}_N \right] \\ &= \frac{-N \bar{X}_N (\bar{X}_N - 1)}{\bar{X}_N^2 (\bar{X}_N - 1)^2} \end{aligned} \quad (43)$$

D'où l'on tire finalement :

$$\left. \frac{\partial^2 \log L(X_1, \dots, X_N; \theta)}{\partial \theta^2} \right|_{\hat{\theta} = \bar{X}_N} = \frac{-N}{\bar{X}_N (\bar{X}_N - 1)} < 0 \quad (44)$$

puisque $\bar{X}_N > 0$.

Question 3 (1.5 point) Montrez que cet estimateur $\hat{\theta}$ est sans biais et convergent. On sait que :

$$E(\hat{\theta}) = E(\bar{X}_N) = \frac{1}{N} \sum_{i=1}^N E(X_i) = \theta \quad (45)$$

L'estimateur est sans biais (0.5 point). Sa variance vaut (0.5 point) :

$$V(\hat{\theta}) = V(\bar{X}_N) = \frac{1}{N^2} \sum_{i=1}^N V(X_i) = \frac{\theta(\theta - 1)}{N} \quad (46)$$

On en déduit que :

$$\lim_{N \rightarrow \infty} V(\hat{\theta}) = 0 \quad (47)$$

L'estimateur est donc convergent en probabilité (0.5 point).

$$\hat{\theta} \xrightarrow[N \rightarrow \infty]{p} \theta \quad (48)$$

Question 4 (1 point) Calculons la borne FDCR dans ce cas :

$$FDCR = \frac{1}{I_N} \quad (49)$$

avec :

$$I_N = E \left[- \frac{\partial^2 \log L(X_1, \dots, X_N; \theta)}{\partial \theta^2} \Big|_{\hat{\theta}} \right] = E \left[\frac{N}{\bar{X}_N (\bar{X}_N - 1)} \right] \quad (50)$$

Dès lors, on a (attention l'espérance est un opérateur linéaire, on ne peut effectuer le calcul directement) :

$$\begin{aligned} FDCR &= \frac{1}{N} E [\bar{X}_N (\bar{X}_N - 1)] \\ &= \frac{1}{N} E [\bar{X}_N^2 - \bar{X}_N] \\ &= \frac{1}{N} E [\bar{X}_N^2 - E(\bar{X}_N)] \\ &= \frac{1}{N} V(X) \\ &= \frac{\theta(\theta - 1)}{N} \end{aligned} \quad (51)$$

On retrouve donc :

$$V(\hat{\theta}) = FDCR = \frac{\theta(\theta - 1)}{N} \quad (52)$$

L'estimateur du MV est efficace.