Université d'Orléans - Licence Economie et Gestion
Statistique Mathématique

C. Hurlin. Correction de l'Examen Terminal Janvier 2008

Exercice 1 *Scoring Bancaire et Tests UPP. Barème : 13 points.*

Question 1 (1.5 point) Cette idée se traduirait sous la forme :

$$H_0 : p = p_0 \quad \text{(0.5 point)} \tag{1}$$

$$H_1 : p = p_1 > p_0 \tag{2}$$

ou sous la forme :

$$H_0 : p = p_0 \quad \text{(0.5 point)} \tag{3}$$

$$H_1 : p > p_0 \tag{4}$$

L'hypothèse nulle correspond à l'hypothèse selon laquelle les clients de cette catégorie ont une faible probabilité de défaut, c'est à dire qu'ils correspondent à de "bons" clients auxquels on peut accorder un prêt. L'hypothèse alternative correspond donc à l'hypothèse de "mauvais" client. L'hypothèse nulle est donc l'hypothèse pour lequel on a le plus fort coût à la renonciation (0.5 point).

Question 2 (1 point) Le risque de première espèce, noté α, correspond la probabilité de rejeter l'hypothèse de "bon" client alors que le client est bon :

$$\alpha = \Pr\left(\text{Rejet } p = p_0 | \ p = p_0\right) \tag{5}$$

La puissance correspond la probabilité de rejeter l'hypothèse de "bon" client alors que le client est effectivement un mauvais client :

$$Puissance = \Pr\left(\text{Rejet } p = p_0 | \ p = p_1\right) \tag{6}$$

Question 3 (2.5 points) Soit $Z_N = (1/N) \sum_{i=1}^{N} Y_i$ la fréquence empirique de défauts associé au N-échantillon $\{Y_1, .., Y_N\}$. On sait que les variables Y_i sont *i.i.d.* (0.5 point) avec

$$E\left(Y_i\right) = p \; et \; V\left(Y_i\right) = p\left(1-p\right) \quad \text{(0.5 point)}$$

D'après le théorème central limite, on a :

$$\frac{Z_N - E\left(Z_N\right)}{\sqrt{V\left(Z_N\right)}} \xrightarrow[N \to \infty]{d} N\left(0,1\right) \tag{7}$$

Or, sait ici que :

$$E\left(Z_N\right) = \left(1/N\right) \Sigma_{i=1}^{N} E\left(Y_i\right) = p \tag{8}$$

$$V\left(Z_N\right) = \left(1/N^2\right) \Sigma_{i=1}^{N} V\left(Y_i\right) = \frac{p\left(1-p\right)}{N} \tag{9}$$

Donc, on a :

$$\frac{Z_N - p}{\sqrt{\frac{p(1-p)}{N}}} \xrightarrow[N \to \infty]{d} N\left(0,1\right) \quad \text{0.5 point} \tag{10}$$

Dès lors, le seuil critique vérifie :

$$\alpha = \Pr\left[Z_N > C \mid \frac{Z_N - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \quad 0.5 \text{ point} \tag{11}$$

ou encore

$$1-\alpha = \Pr\left[\frac{Z_N - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} < \frac{C - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \;\middle|\; \frac{Z_N - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \tag{12}$$

Ainsi, si N tend vers l'infini :

$$\Phi^{-1}(1-\alpha) = \frac{C - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \tag{13}$$

et donc finalement :

$$C = p_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{N}} \quad 0.5 \text{ point} \tag{14}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite.

Question 4 (1 point) A partir des éléments fournis, on déduit que le seuil critique du test est égal à :

$$\begin{aligned} C &= 0.2 + \Phi^{-1}(0.9)\sqrt{\frac{0.2(1-0.2)}{100}} \\ &= 0.2 + 1.2816 * \sqrt{\frac{0.2(1-0.2)}{100}} \\ &= 0.251 \end{aligned}$$

Si l'on admet que la pertinence de la distribution asymptotique pour cet échantillon, la région critique du test est donc la suivante :

$$W = \{Y_1, .., Y_N \mid Z_N > 0.251\} \quad 0.5 \text{ point} \tag{15}$$

Or dans le cas de l'échantillon considéré, la réalisation de la fréquence de défaut est :

$$Z_N = \frac{1}{100}\sum_{i=1}^{100} Y_i = \frac{23}{100} = 0.23 \tag{16}$$

Donc pour un risque de première espèce de 10%, on ne peut pas rejetter H_0. Donc on ne peut pas rejeter l'hypothèse que les individus de cette catégorie sont des bons clients : on accepte donc de leur accorder le prêt. (0.5 point).

Question 5 (1 point) La puissance du test est :

$$\begin{aligned} Puissance &= \Pr(W \mid H_1 \text{ vraie}) \\ &= \Pr\left(Z_N > C \mid \frac{Z_N - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right) \end{aligned} \tag{17}$$

Par conséquent, on a :

$$Puissance = 1 - \Phi\left(\frac{C - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}}\right) \quad \text{0.5 point} \tag{18}$$

AN : $C = 0.251$, $p_1 = 0.3$, $N = 100$:

$$\begin{aligned} Puissance &= 1 - \Phi(-1.0635) \\ &= 1 - 0.1438 \\ &= 0.856 \end{aligned} \tag{19}$$

Cette procédure permetrait ainsi de detecter 85,6% de "mauvais" clients. Plus précisement si les individus de la catégorie sont effectivement de "mauvais" clients, pour 100 échantillons, la procédure conduirait dans 85,6% des cas où rejet de l'hypothèse nulle de "bons" clients (0.5 point).

Question 6 (2 points) Le responsable du département des risques souhaite que la règle de décision permette d'atteindre un niveau donné de puissance, noté γ, pour un niveau donné de risque de première espèce α. Dès lors, par définition on a :

$$\gamma = 1 - \Phi\left(\frac{C - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}}\right) \tag{20}$$

ou encore :

$$1 - \gamma = \Phi\left(\frac{p_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{N}} - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}}\right) \quad \text{0.5 point} \tag{21}$$

$$\Longleftrightarrow \Phi^{-1}(1-\gamma)\sqrt{\frac{p_1(1-p_1)}{N}} = p_0 - p_1 + \Phi^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{N}} \tag{22}$$

$$\Longleftrightarrow \frac{1}{\sqrt{N}}\left[\Phi^{-1}(1-\gamma)\sqrt{p_1(1-p_1)} - \Phi^{-1}(1-\alpha)\sqrt{p_0(1-p_0)}\right] = p_0 - p_1 \tag{23}$$

D'où finalement

$$N = \left[\frac{\Phi^{-1}(1-\gamma)\sqrt{p_1(1-p_1)} - \Phi^{-1}(1-\alpha)\sqrt{p_0(1-p_0)}}{p_0 - p_1}\right]^2 \quad \text{1 point} \tag{24}$$

Pour garantir 90% de puissance pour un niveau de risque de première espèce de 10%, il faut :

$$N = \left[\frac{\Phi^{-1}(1-0.9)\sqrt{0.3(1-0.3)} - \Phi^{-1}(1-0.1)\sqrt{0.2(1-0.2)}}{0.2 - 0.3}\right]^2 \simeq 121 \quad \text{0.5 point} \tag{25}$$

Question 7 (1 point) On considère à présent un test d'hypothèse simple contre hypothèse multiple unilatéral :

$$H_0 : p = 0.2 \tag{26}$$

$$H_1 : p > 0.2 \tag{27}$$

La région critique du test UPP de niveau α d'hypothèse simple $H_0 : p = p_0$ contre hypothèse simple $H_0 : p = p_1$ avec $p_1 > p_0$ ne dépend pas de la valeur explicite de p_1. Dès lors la région critique du test UPP d'hypothèse simple $H_0 : p = p_0 = 0.2$ contre hypothèse multiple $H_0 : p > p_0$ est identique à la région critique du test simple contre simple. (0.5 point).Donc la région critique est :

$$W = \left\{ Y_1, .., Y_N \mid Z_N > p_0 + \Phi^{-1}(1-\alpha) \sqrt{\frac{p_0(1-p_0)}{N}} \right\} \tag{28}$$

AN : $N = 100$, $p_0 = 0.2$ et $\alpha = 10\%$:

$$W = \{ Y_1, .., Y_N \mid Z_N > 0.251 \} \quad \text{0.5 point} \tag{29}$$

Question 8 (1 point) La fonction puissance, notée $G(p)$, est définie par :

$$G(p) = 1 - \Phi\left(\frac{C - p}{\sqrt{\frac{p(1-p)}{N}}} \right) \tag{30}$$

ou encore :

$$G(p) = 1 - \Phi\left(\frac{p_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{N}} - p}{\sqrt{\frac{p(1-p)}{N}}} \right) \quad \text{0.5 point} \tag{31}$$

Dès lors pour $p_0 = 0.2$, $N = 100$ et $\alpha = 10\%$, si la vraie valeur de la probabilité de défaut de paiement est égale à 28%, la puissance du test vaut :

$$G(0.28) = 1 - \Phi\left(\frac{0.2 + \Phi^{-1}(0.9)\sqrt{\frac{0.2(1-0.2)}{100}} - 0.28}{\sqrt{\frac{0.28(1-0.28)}{100}}} \right) = 0.738 \quad \text{0.5 point} \tag{32}$$

On vérifie au passage que naturellement la valeur de la puissance pour $p_1 = 0.28$ (73,8%) est inférieure à celle obtenue pour l'hypothèse alternative $p_1 = 0.3$ (85%) à la question 5.

Question 9 (2 points) On souhaite tester la valeur de la probabilité de défaut comme suit :

$$H_0 : p = 0.2 \tag{33}$$

$$H_1 : p \neq 0.2 \tag{34}$$

La région d'acceptation du test bilatéral correspond à l'intersection des régions d'acceptation des deux tests unilatéraux associés $H_0 : p = 0.2$ contre $H_1 : p > 0.2$ et $H_0 : p = 0.2$ contre $H_1 : p < 0.2$ définies pour des risque de première espèce de $\alpha/2 = 2.5\%$. Soient K_1 et K_2 le deux seuils critiques correspondant (0.5 point) :

$$K_1 = p_0 + \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\frac{p_0(1-p_0)}{N}} = 0.1216 \tag{35}$$

$$K_2 = p_0 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{p_0(1-p_0)}{N}} = 0.2784 \tag{36}$$

Donc la RC du test bilatéral s'écrit (1 point) :

$$W = \{ Y_1, .., Y_N \mid Z_N \notin [0.134; 0.265] \} \quad \text{1 point} \tag{37}$$

Dès lors, $\sum_{i=1}^{100} Y_i = 23$, on a $Z_N = 0.23$ et donc la réalisation de Z_N appartient à la région d'acceptation. Pour un risque de première espcèce de 5%, on ne peut pas rejeter l'hypothèse nulle selon laquelle $p = 0.2$ (0.5 point).

Exercice 2 *Test d'Independance : Relation Salaire / Diplôme (à partir d'un examen de Mme Bessec, Université Paris IX Dauphine). Barème : 4 points.* (1 point pour tableau effectif théorique + 1 statistique + 1 seuil et conclusion)

- 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	Bac	BEP-CAP	Sixième	Total
600-750	137	266	26	429
750-900	59	115	11	185
900-1650	29	56	5	91
Total	226	437	42	705

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 :

$$\Delta = \sum_i \sum_j \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \underset{H_0}{\rightarrow} \chi^2_{(l-1)(c-1)} = \chi^2_4 \quad \text{car } T_{ij} \geq 4 \quad \forall i, j \text{ (voir ci-dessus)}$$

-Règle de décision :

W tel que $\Delta > c$

C : CPE $\rightarrow$ $c = \chi^2_{4;\alpha} = 9.49$ au seuil $\alpha = 5\%$ et 7.78 au seuil $\alpha = 10\%$.

- Conclusion :

Ici, $\Delta_{obs} = \frac{(77 - 115.01)^2}{115.01} + \ldots + \frac{(2 - 5.22)^2}{5.22} = 20.02 > c \quad \forall \alpha$

On rejette donc H_0 aux deux seuils. Les deux variables ne sont pas indépendantes. On

Exercice 3 Question 1 :

$$\begin{aligned} L &= \prod_{i=1}^{N} f(x_i) = \left(\frac{1}{\theta . c^{\frac{1}{\theta}}}\right)^N . \prod_{i=1}^{N} x_i^{\frac{1}{\theta}-1} \quad si\ \textbf{tous} \text{ les } \mathrm{x}_i\ v\acute{e}rfient\ \ 0 \leq x_i \leq c \\ &= 0 \, \text{sinon (1 point)} \end{aligned} \tag{38}$$

Question 2 : supposons que tous les x_i sont tels que $0 \leq x_i \leq c$, alors

$$\begin{aligned} \frac{\partial \log L}{\partial \theta} &= \frac{\partial}{\partial \theta}\left[-N \log(\theta) - \frac{N}{\theta}\log(c) + \left(\frac{1}{\theta} - 1\right)\sum \log(x_i)\right] \quad 0.5\ point \\ &= -\frac{N}{\theta} + \frac{N}{\theta^2}\log(c) - \frac{1}{\theta^2}\sum \log(x_i) = 0 \quad 0.5\ point \end{aligned}$$

D'où l'on tire

$$\hat{\Theta}_n = \ln(c) - \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \tag{39}$$

Question 3. On pose $Y = \ln(c) - \ln(X)$ et l'on admet que $E(Y) = \theta$ et $V(Y) = \theta^2$. On alors :

$$\begin{aligned} E\left(\hat{\Theta}_n\right) &= \ln(c) - \frac{1}{n}\sum_{i=1}^{n} E[\ln(X_i)] \\ &= \ln(c) - \frac{1}{n}\sum_{i=1}^{n} E[\ln(c) - \ln(Y_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[\ln(Y_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta = \theta \quad \text{1 point sans biais} \end{aligned}$$

$$\begin{aligned} V\left(\hat{\Theta}_n\right) &= \ln(c) - \frac{1}{n^2}\sum_{i=1}^{n} V[\ln(X_i)] \\ &= \ln(c) - \frac{1}{n^2}\sum_{i=1}^{n} V[\ln(c) - \ln(Y_i)] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} V[\ln(Y_i)] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \theta^2 \\ &= \frac{\theta^2}{n,} \quad convergent \text{ 1point} \end{aligned}$$