Université d'Orléans - Licence Economie et Gestion
Statistique Mathématique

C. Hurlin. Correction Examen Décembre 2006

Exercice 1 *Tests UPP et Théorème de Neyman Pearson. Barème : 12 points.*

Soit X, une variable désignant le nombre d'incidents de paiement sur un crédit à la consommation observés sur la durée du prêt. On suppose que X suit une loi de Poisson de paramètre θ $(\theta > 0)$. On dispose d'un échantillon de N clients d'une banque noté $\{X_1, .., X_N\}$, où X_i désigne le nombre d'incidents observés pour l'individu i. On suppose que les variables X_i sont *i.i.d.* de même loi que X et l'on admet que :

$$P(X_i = k) = e^{-\theta} \frac{\theta^k}{k!} \tag{1}$$

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

$$H_0 : \theta = \theta_0 \tag{2}$$

$$H_0 : \theta = \theta_1 > \theta_0 \tag{3}$$

ou sous la forme :

$$H_0 : \theta = \theta_0 \tag{4}$$

$$H_0 : \theta > \theta_0 \tag{5}$$

Un bon client a en moyenne peu d'incidents de paiement.

Question 2 (2 points) On considère le test suivant :

$$H_0 : \theta = \theta_0 \tag{6}$$

$$H_0 : \theta = \theta_1 \tag{7}$$

avec $\theta_0 < \theta_1$. On sait que d'après le théorème de Neyman Pearson (1 point) :

$$\frac{L(X_1, .., X_N; \theta_0)}{L(X_1, .., X_N; \theta_1)} = e^{N(\theta_0 - \theta_1)} \left(\frac{\theta_0}{\theta_1}\right)^{\Sigma_{i=1}^N X_i} < A \tag{8}$$

où A est une constante. On en déduit que

$$N(\theta_0 - \theta_1) + \Sigma_{i=1}^N X_i \log\left(\frac{\theta_0}{\theta_1}\right) < \log(A) \tag{9}$$

$$\Longleftrightarrow \Sigma_{i=1}^N X_i > [\log(A) - N(\theta_0 - \theta_1)] \log\left(\frac{\theta_0}{\theta_1}\right) \tag{10}$$

car $\theta_1 > \theta_0$. Ce qui peut se réécrire sous la forme :

$$\overline{X}_N = (1/N) \Sigma_{i=1}^N X_i > K \tag{11}$$

où K est une constante déterminée par le niveau du risque de première espèce α. (1 point). La région critique du test UPP de niveau α est donc de la forme :

$$W = \left\{ X_1, .., X_N | \overline{X}_N > K \right\} \tag{12}$$

Question 3 (2 points) On sait que d'après le théorème central limite, on a :

$$\frac{\overline{X}_N - E\left(\overline{X}_N\right)}{\sqrt{V\left(\overline{X}_N\right)}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1) \tag{13}$$

Or, sait ici que :

$$E\left(\overline{X}_N\right) = (1/N)\,\Sigma_{i=1}^{N} E(X_i) = \theta \tag{14}$$

$$V\left(\overline{X}_N\right) = \left(1/N^2\right)\Sigma_{i=1}^{N} V(X_i) = \frac{\theta}{N} \tag{15}$$

Donc, on a (1 point) :

$$\frac{\overline{X}_N - \theta}{\sqrt{\theta/N}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1) \tag{16}$$

Dès lors, le seuil critique vérifie :

$$\alpha = \Pr\left[\overline{X}_N > K \Big| \frac{\overline{X}_N - \theta_0}{\sqrt{\theta_0/N}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \tag{17}$$

ou encore

$$1-\alpha = \Pr\left[\overline{X}_N < K \Big| \frac{\overline{X}_N - \theta_0}{\sqrt{\theta_0/N}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \tag{18}$$

D'où l'on tire que si N tend vers l'infini :

$$\Phi^{-1}(1-\alpha) \approx \frac{\overline{X}_N - \theta_0}{\sqrt{\theta_0/N}} \tag{19}$$

et donc finalement :

$$K \approx \theta_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{\theta_0}{N}} \tag{20}$$

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

Question 4 (1.5 points) On souhaite tester l'hypothèse nulle selon laquelle les clients de la banque A sont faiblement risqués sous la forme suivante : $H_0 : \theta = 1$ contre $H_1 : \theta = 2$. On sait que sur cet échantillon (0.5 point) :

$$\overline{X}_N = (0*510 + 500 + 2*226 + 3*70 + 4*19 + 5*6 + 6*1)/1332 = 0.9565 \tag{21}$$

Or la RC associée au test UPP de niveau α est définie par :

$$W = \left\{X_1, .., X_N \middle| \overline{X}_N > K\right\} \tag{22}$$

avec $K = \theta_0 + \Phi^{-1}(1-\alpha)\sqrt{\theta_0/N}$. On en déduit que pour $\alpha = 5\%$, on a (0.5 point) :

$$K = 1 + \Phi^{-1}(0.95)\sqrt{1/1332} = 1 + 1.64\sqrt{1/1332} = 1.0451 \tag{23}$$

Pour un risque de 5%, on ne peut pas rejetter H_0, c'est à dire l'hypothèse selon laquelle les clients de la banque A sont relativement peu risqués (0.5 point).

Question 5 (1 point) Il s'agit ici de déterminer le risque de deuxième espèce de notre test :

$$\beta = \Pr\left[\overline{X}_N < K \middle| \frac{\overline{X}_N - \theta_1}{\sqrt{\theta_1/N}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \approx \Phi\left(\frac{K-\theta_1}{\sqrt{\theta_1/N}}\right) \tag{24}$$

Ce qui conduit à :

$$\beta \approx \Phi\left(\frac{1.0451-2}{\sqrt{2/1332}}\right) \approx \Phi(-24.643) \approx 0 \tag{25}$$

Ce risque est quasi nul.

Question 6 (1.5 points) On a vu que la RC du Test UPP d'hypothèse simple $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1 > \theta_0$ ne dépend pas de la valeur explicite de θ sous H_1, donc la RC du test UPP de niveau α $H_0 : \theta = \theta_0$ contre $H_0 : \theta > \theta_0$ est équivalente à la RC du test de la question 3. (1 point)

$$W = \left\{ X_1, .., X_N \middle| \overline{X}_N > 1.0451 \right\} \tag{26}$$

Dans l'échantillon, on observe $\overline{X}_N = 0.9565$. Pour un risque de 5%, on ne peut pas rejetter H_0, c'est à dire l'hypothèse selon laquelle les clients de la banque A sont relativement peu risqués (0.5 point).

Question 7 (2 points) On souhaite enfin tester l'hypothèse :

$$H_0 : \theta = 1 \tag{27}$$

$$H_1 : \theta \neq 1 \tag{28}$$

Construisez la région d'accpetation du test bilatéral correspond à l'intersection des RA des deux test unilatéraux associés $H_0 : \theta = 1$ contre $H_1 : \theta > 1$ et $H_0 : \theta = 1$ contre $H_1 : \theta < 1$ définies pour des risque de première espèce de $\alpha/2 = 2.5\%$. Soient K_1 et K_2 le deux seuils critiques correspondant (1 point) :

$$K_1 = \theta_0 + \Phi^{-1}(\alpha)\sqrt{\theta_0/N} = 0.9549 \tag{29}$$

$$K_1 = \theta_0 + \Phi^{-1}(1-\alpha)\sqrt{\theta_0/N} = 1.0451 \tag{30}$$

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

$$W = \left\{ X_1, .., X_N \middle| \overline{X}_N \notin [0.9549; 1.0451] \right\} \tag{31}$$

Question 8 (1 point) Puissance :

$$Puissance(\theta) = \Pr\left[\overline{X}_N \notin [0.9549; 1.0451] \,\middle|\, \frac{\overline{X}_N - \theta}{\sqrt{\theta/N}} \underset{N\to\infty}{\overset{d}{\longrightarrow}} N(0,1)\right] \tag{32}$$

D'où l'on tire :

$$Puissance(\theta) \approx 1 - \Phi\left(\frac{1.0451-\theta}{\sqrt{\theta/N}}\right) + \Phi\left(\frac{0.9549-\theta}{\sqrt{\theta/N}}\right) \tag{33}$$

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 :

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

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

Exercice 3 *Test d'Ajustement (Université Paris IX Dauphine, avec l'autorisation de Mme Bessec). Barème : 4 points.*

Question 1(2 points) Soit $\{X_1, .., X_N\}$, avec $N = 100$, l'échantillon des ventes quotidiennes supposées *i.i.d.* et de même loi que X où X suit une loi de poisson de paramètre λ. Ecrivons la log-vraisemblance de cet échantillon :

$$L(X_1, .., X_N, \lambda) = \Pi_{i=1}^N P(X = X_i) = \Pi_{i=1}^N e^{-\lambda} \frac{\lambda^{X_i}}{X_i!} \tag{34}$$

Ce qui peut se réecrire sous la forme d'une log-vraisemblance (0.5 point) :

$$\log L(X_1, .., X_N, \lambda) = -N\lambda + \log(\lambda) \ \Sigma_{i=1}^N X_i - \log\left(\Pi_{i=1}^N X_i!\right) \tag{35}$$

La condition nécessaire de la maximisation de la vraisemblance est alors :

$$\frac{\partial L(X_1, .., X_N, \lambda)}{\partial \lambda} = -N + \frac{1}{\lambda} \Sigma_{i=1}^N X_i = 0 \tag{36}$$

D'où l'on déduit l'estimateur du MV (0.5 point) :

$$\widehat{\lambda} = \frac{1}{N} \Sigma_{i=1}^N X_i = \overline{X}_N \tag{37}$$

On vérifie qu'il s'agit d'un maximum (0.5 point) puisque :

$$\frac{\partial^2 L(X_1, .., X_N, \lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2} \Sigma_{i=1}^N X_i < 0 \tag{38}$$

On en déduit finalement une estimation ponctuelle du paramètre λ (0.5 point) :

$$\widehat{\lambda} = \overline{x}_N = 1.5 \tag{39}$$

Question 2 (1 point) Cf Tableau de la Figure 2

Question 3 (1 point) Si les ventes quotidiennes suivent effectivement une loi de Poisson de paramètre 1.5, la probabilité que le produit ne soit pas acheté est égale à 22,31%. (0.5 point). Mais cette hypothèse n'est pas nécessairement valable. Tout dépend de la puissance du test d'adéquation : la puissance correspond à la probabilité de rejeter l'hypothèse nulle d'une loi de Poission alors qu'effectivement les ventes suivent une autre loi que la loi de Poisson. Plus cette puissance est faible, plus la direction prend un risque important. (**1 point bonus**)

Figure 1:

2. Test d'ajustement à une loi de Poisson

- Avant regroupement des classes

xi	0	1	2	3	4	5	6	7	8 (et +)	Total
pi	0,2231	0,3347	0,251	0,1255	0,0471	0,0141	0,0035	0,0008	0,0002	1
ti	22,31	33,47	25,1	12,55	4,71	1,41	0,35	0,08	0,02	100

- Après regroupement des classes

xi'	0	1	2	3	4 et +	Total
pi'	0,2231	0,3347	0,251	0,1255	0,0657	1
ti'	22,31	33,47	25,1	12,55	6,57	100
oi'	25	31	22	16	6	100
chi'	0,32434	0,18228	0,38287	0,94841	0,04945	1,88735

stat	1,88735	ddl	3	vc	7,81

$\Rightarrow$ non rejet de H_0 au seuil $\alpha = 5\%$.