Université d'Orléans - Licence Economie et Gestion Statistique

C. Hurlin. Correction de l'Examen de Décembre 2008

Exercice 1 *Barème : 15 points. Test Paramétrique et Lemme de Neyman Pearson*

Partie I : Test d'hypothèses simples (8 points)

Question 1 (1 point) C'est l'hypothèse dans laquelle le fabricant a le plus confiance et qu'il lui importe le plus.

Question 2 (1 point) On sait que, pour $N = 10$:

$$L(X_1, .., X_{10}, m_0) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{X_i - m_0}{\sigma}\right)^2\right] \quad \text{0.5 point} \tag{1}$$

$$L(X_1, .., X_{10}, m_1) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{X_i - m_1}{\sigma}\right)^2\right] \tag{2}$$

Il s'ensuit immédiatement que :

$$\frac{L(X_1, .., X_{10}, m_0)}{L(X_1, .., X_{10}, m_1)} = \exp\left[\frac{1}{2\sigma^2}\left(\sum_{i=1}^{10}(X_i - m_1)^2 - \sum_{i=1}^{10}(X_i - m_0)^2\right)\right] \quad \text{0.5 point} \tag{3}$$

Question 3 (1 point) On sait que selon le lemme de Neyman Pearson, la région critique du test UPP de niveau α est telle que :

$$W = \left\{X_1, .., X_{10} \middle| \; \frac{L(X_1, .., X_{10}, m_0)}{L(X_1, .., X_{10}, m_1)} < K\right\} \tag{4}$$

où K désigne une constante déterminée par le risque de première espèce α. Or, :

$$\frac{L(X_1, .., X_{10}, m_0)}{L(X_1, .., X_{10}, m_1)} < K \tag{5}$$

$$\Longleftrightarrow \exp\left[\frac{1}{2}\left(\sum_{i=1}^{N}(X_i - m_1)^2 - \sum_{i=1}^{N}(X_i - m_0)^2\right)\right] < K \quad \text{0.5 point} \tag{6}$$

$$\Longleftrightarrow \sum_{i=1}^{N} X_i^2 - 2m_1X_i + m_1^2 - X_i^2 + 2m_0X_i - m_0^2 < 2\log(K) \tag{7}$$

$$\Longleftrightarrow 2\sum_{i=1}^{N}(m_0 - m_1)\; X_i < 2\log(K) + N\left(m_0^2 - m_1^2\right) \tag{8}$$

Ici, on a $m_0 - m_1 = 300 - 290 > 0$, dès lors on obtient au final :

$$\sum_{i=1}^{N} X_i < B \tag{9}$$

avec $B = \left[2\log(K) + N\left(m_0^2 - m_1^2\right)\right] / \left[2\left(m_0 - m_1\right)\right]$ ou encore

$$\overline{X}_N < A \quad 0.5 \text{ point} \tag{10}$$

où $A = B/N$. La région critique W du test UPP au sens de Neyman Pearson est de la forme

$$W = \left\{ X_1, .., X_{10} | \ \overline{X}_N < A \right\} \tag{11}$$

Remarque : attribuer les points même si les termes de droite des inégalités ne sont pas spécifiés.

Question 4 (1 point) Par définition du risque de première espèce :

$$\alpha = \Pr\left[W \mid H_0\right] \tag{12}$$

Ou encore :

$$\alpha = \Pr\left[\overline{X}_N < A \mid m = m_0\right] \tag{13}$$

On sait que sous H_0, la moyenne empirique suit une loi normale centrée sur m_0 :

$$\overline{X}_N \sim N\left(m_0, \frac{\sigma^2}{N}\right) \tag{14}$$

Par conséquent :

$$\alpha = \Pr\left[\frac{\overline{X}_N - m_0}{\sigma/\sqrt{N}} < \frac{A - m_0}{\sigma/\sqrt{N}} \;\middle|\; \frac{\overline{X}_N - m_0}{\sigma/\sqrt{N}} \sim N(0,1)\right] \tag{15}$$

ce qui implique que :

$$\alpha = \Phi\left(\frac{A - m_0}{\sigma/\sqrt{N}}\right) \tag{16}$$

où $\Phi(.)$ désigne la fonction de répartition de loi normale centrée réduite. On en déduit donc au final que :

$$A = m_0 + \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{N}} \quad 0.5 \text{ point} \tag{17}$$

AN :

$$A = 300 - 1.64 * \frac{5}{\sqrt{10}} = 297.39 \quad 0.5 \text{ point} \tag{18}$$

Question 5 (1 point) La région critique du test UPP au seuil de 5% est définie par :

$$W = \left\{ X_1, .., X_{10} | \ \overline{X}_N < 297.39 \right\} \tag{19}$$

Dans le cas de ce lot, la moyenne empirique égale à 296 grammes figure dans la région critique. Donc pour un seuil de risque de première espèce de 5% on rejette l'hypothèse nulle que $m = 300$ grammes. La marchandise est donc non conforme et doit être détruite.

Question 6 (1.5 point) La puissance du test est :

$$Puissance = 1 - \beta = \Pr\left[W \mid H_1\right] \tag{20}$$

D'où l'on tire que :

$$Puissance = \Pr\left[\frac{\overline{X}_N - m_1}{\sigma/\sqrt{N}} < \frac{A - m_1}{\sigma/\sqrt{N}} \;\middle|\; \frac{\overline{X}_N - m_1}{\sigma/\sqrt{N}} \sim N(0,1)\right] \tag{21}$$

$$\Longleftrightarrow Puissance = \Phi\left(\frac{A - m_1}{\sigma/\sqrt{N}}\right) \quad \text{0.5 point} \tag{22}$$

AN :

$$Puissance = \Phi\left(\frac{297.39 - 295}{5/\sqrt{10}}\right) = \Phi(1.51) = 0.93 \;\; \text{0.5 point} \tag{23}$$

Ce qui signifie que sur 100 lots défectueux, ce test permet de rejeter l'hypothèse nulle de validité du lot dans 93 cas (0.5 point).

Question 7 (1.5 points) On a donc :

$$Puissance = \Phi\left(\frac{A - m_1}{\sigma/\sqrt{N}}\right) = 0.95 \tag{24}$$

$$\Longleftrightarrow \frac{A - m_1}{\sigma/\sqrt{N}} = \Phi^{-1}(0.95) \tag{25}$$

$$\Longleftrightarrow N = \left[\frac{\Phi^{-1}(0.95)\ \sigma}{(A - m_1)}\right]^2 \quad \text{1 point} \tag{26}$$

AN :

$$N = \left[\frac{\Phi^{-1}(0.95) * 5}{(297.39 - 295)}\right]^2 = 11.75 \tag{27}$$

Donc N doit être supérieure à 12 unités (0.5 point).

Partie II : Test d'hypothèse simple contre hypothèse multiple (7 points)

Question 1 (1 point) Nous avons montré que la région critique du test UPP d'hypothèse simple contre hypothèse simple $H_0 : m = 300$ contre $H_1 : m = 295$ ne dépendait pas de la valeur explicite de m_1. Par conséquent, il existe un test UPP pour $H_0 : m = 300$ contre $H_1 : m < 300$ et la région critique de ce test est identique à celle du test d'hypothèse simple :

$$W = \left\{ X_1, .., X_{10} \mid \overline{X}_N < 297.39 \right\} \tag{28}$$

Question 2 (2 points) Dans le cas de ce test unilatéral, la puissance est une fonction de la valeur de m sous H_1 :

$$P(m) = \Phi\left(\frac{A - m}{\sigma/\sqrt{N}}\right) \quad m < 300 \tag{29}$$

Pour $m = m_0 = 300$, par définition on retrouve le niveau de risque I :

$$P(300) = \alpha = 0.05 \quad \text{0.5 point} \tag{30}$$

Pour $m = 295$, nous avons montré que :

$$P(295) = 0.93 \tag{31}$$

Enfin, on montre que :

$$P(298) = \Phi\left(\frac{297.39 - 298}{5/\sqrt{10}}\right) = \Phi(-0.37) = 0.35 \quad \text{0.5 point} \tag{32}$$

$$P(290) = \Phi\left(\frac{297.39 - 290}{5/\sqrt{10}}\right) = \Phi(4.67) = 1 \quad \text{0.5 point} \tag{33}$$

plus 0.5 point pour le graphique de la courbe de puissance.

Question 3 (2 point) La région d'acceptation du test bilatéral de niveau α correpond à l'intersection des régions d'accpetation des tests unilatéraux de niveau $\alpha/2$ (0.5 point). Donc on a :

$$\overline{W} = \left\{ X_1, .., X_{10} \mid A < \overline{X}_N < B \right\} \tag{34}$$

où les constantes A et B vérifient (0.5 point) :

$$A = m_0 + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}} \tag{35}$$

$$B = m_0 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}} \tag{36}$$

AN (0.5 point) :

$$A = 300 + \Phi^{-1}\left(\frac{0.05}{2}\right)\frac{5}{\sqrt{10}} = 296.90 \tag{37}$$

$$B = 300 + \Phi^{-1}\left(1 - \frac{0.05}{2}\right)\frac{5}{\sqrt{10}} = 303.09 \tag{38}$$

Donc la région critique du test bilatéral est :

$$W = \left\{ X_1, .., X_{10} \mid \overline{X}_N < 296.90 \ ou \ \overline{X}_N > 303.09 \right\} \quad \text{0.5 point} \tag{39}$$

Si on observe une moyenne empirique de 296 grammes pour un lot, pour un risque de première espèce de 5%, on rejette l'hypothèse nulle de validité de la production.

Question 4 (1 point) La puissance est définie par :

$$\begin{aligned} P(m) &= \Pr\left[W \mid H_1\right] \\ &= 1 - \Pr\left[296.90 < \overline{X}_N < 303.09 \mid H_1\right] \quad \text{0.5 point} \end{aligned} \tag{40}$$

D'où l'on tire que :

$$P(m) = 1 - \Phi\left(\frac{303.09 - m}{\sigma/\sqrt{N}}\right) + \Phi\left(\frac{296.90 - m}{\sigma/\sqrt{N}}\right) \quad \text{0.5 point} \tag{41}$$

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 *Test d'Adéquation : Barème : 4 points.*

Effectifs théoriques (1 point) sous H_0 : $X_i \sim P(2)$

X_i	0	1	2	3	4	5	4	7 et +	Total
$N * p_i$	13.53	27.06	27.06	18.04	9.02	3.60	1.20	0.45	100

On doit regrouper les classes 4, 5 et 7. Après regroupement, les effectifs théoriques et empiriques (1 point) sont les suivants :

X_i	0	1	2	3	4	5 et +	Total
N_i	15	25	25	12	10	13	100
$N * p_i$	13.53	27.06	27.06	18.04	9.02	5.16	100

Par conséquent la statistique du chi-deux vaut :

$$D_{(F_N,F)} = \sum_{i=1}^{N} \frac{(N_i - N * pi)^2}{N * pi} = 14.54 \quad \text{(0.5 point)}$$

Sous H_0, on a :

$$D_{(F_N,F)} \sim \chi^2(6) \quad \text{(0.5 point)} \tag{42}$$

Le seuil critique à 95% étant égal à 12.59, pour un risque de première espèce de 5%, on rejette l'hypothèse nulle selon laquelle la loi de X est une loi de poisson de paramètre 2 (0.5 point).