Université d'Orléans - Licence Economie et Gestion
Statistique Mathématique

C. Hurlin. Correction du Contrôle de Décembre 2010

Exercice 1 *Barème : 17 points*

Partie I : Test d'hypothèses simples (10 points)

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

$$L\left(X_1,..,X_{10},a_0\right) = \left(\frac{1}{b\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{X_i - a_0}{b}\right)^2\right] \quad 0.5 \text{ point} \tag{1}$$

$$L\left(X_1,..,X_{10},a_1\right) = \left(\frac{1}{b\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{X_i - a_1}{b}\right)^2\right] \tag{2}$$

Il s'ensuit immédiatement que :

$$\frac{L\left(X_1,..,X_{10},a_0\right)}{L\left(X_1,..,X_{10},a_1\right)} = \exp\left[\frac{1}{2b^2}\left(\sum_{i=1}^{10}(X_i - a_1)^2 - \sum_{i=1}^{10}(X_i - a_0)^2\right)\right] \quad 0.5 \text{ point} \tag{3}$$

Question 2 (1 point) On sait que selon le lemme de Neyman Pearson, la région critique du test UPP de niveau α est telle que :

$$W = \left\{X_1,..,X_{10} \middle| \ \frac{L\left(X_1,..,X_{10},a_0\right)}{L\left(X_1,..,X_{10},a_1\right)} < K\right\} \tag{4}$$

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

$$\frac{L\left(X_1,..,X_{10},a_0\right)}{L\left(X_1,..,X_{10},a_1\right)} < K \tag{5}$$

$$\iff \exp\left[\frac{1}{2}\left(\sum_{i=1}^{N}(X_i - a_1)^2 - \sum_{i=1}^{N}(X_i - a_0)^2\right)\right] < K \quad 0.5 \text{ point} \tag{6}$$

$$\iff \sum_{i=1}^{N} X_i^2 - 2a_1X_i + a_1^2 - X_i^2 + 2a_0X_i - a_0^2 < 2\log(K) \tag{7}$$

$$\iff 2\sum_{i=1}^{N}(a_0 - a_1)\ X_i < 2\log(K) + N\left(a_0^2 - a_1^2\right) \tag{8}$$

Ici, on a $a_0 - a_1 = 2 - 3 < 0$, dès lors on obtient au final :

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

avec $B = \left[2\log(K) + N\left(a_0^2 - a_1^2\right)\right] / \left[2\left(a_0 - a_1\right)\right]$ ou encore

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

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

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

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

Question 3 (2 points) D'après les hypothèses,

$$\overline{X}_N \sim N\left(a, \frac{b^2}{N}\right) \quad \text{0.5 point} \tag{12}$$

Dès lors, le seuil critique A dépend de b par définition du risque de première espèce :

$$\alpha = \Pr\left[W \mid H_0\right] \tag{13}$$

Ou encore :

$$\alpha = \Pr\left[\overline{X}_N > A \mid a = a_0\right] \tag{14}$$

Par conséquent :

$$\alpha = \Pr\left[\frac{\overline{X}_N - a_0}{b/\sqrt{N}} > \frac{A - a_0}{b/\sqrt{N}} \;\middle|\; \frac{\overline{X}_N - a_0}{b/\sqrt{N}} \sim N(0,1)\right] \quad \text{0.5 point} \tag{15}$$

ce qui implique que :

$$1 - \alpha = \Phi\left(\frac{A - a_0}{b/\sqrt{N}}\right) \tag{16}$$

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

$$A = a_0 + \Phi^{-1}(1-\alpha)\frac{b}{\sqrt{N}} \quad \text{0.5 point} \tag{17}$$

Le seuil critique dépend d'un paramètre inconnu et ne peut donc pas être déterminé.(0.5 point). Il convient de construire le test à partir d'une autre statistique dont la loi est connue.

Question 4 (2 points) Sous H_0:

$$Z_N = \frac{\overline{X}_N - a_0}{S_N/\sqrt{N}} \sim t_{(N-1)} \quad \text{(0.5 point)} \tag{18}$$

où $t_{(N-1)}$ désigne la loi de Student à $N-1$ degrés de liberté. Par définition du risque de première espèce :

$$\alpha = \Pr\left[W \mid H_0\right] \tag{19}$$

Ou encore :

$$\alpha = \Pr\left[\overline{X}_N > A \mid a = a_0\right] \quad \text{(0.5 point)} \tag{20}$$

ou encore :

$$1 - \alpha = \Pr\left[\overline{X}_N < A \mid Z_N \sim t_{(N-1)}\right] \tag{21}$$

On obtient :

$$1-\alpha=\Pr\left[\frac{\overline{X}_N-a}{S_N/\sqrt{N}}<\frac{A-a}{S_N/\sqrt{N}}\;\middle|\; a=a_0\right] \quad (0.5\text{ point}) \tag{22}$$

Dès lors,

$$1-\alpha=G\left(\frac{A-a_0}{S_N/\sqrt{N}}\right) \tag{23}$$

et finalement

$$A=a_0+G^{-1}(1-\alpha)\frac{S_N}{\sqrt{N}} \quad (0.5\text{ point}) \tag{24}$$

où $G(.)$ désigne la fonction de répartition de la loi $t_{(N-1)}$.

Question 5 (1 point) AN : $G^{-1}(0.95)=1.734$ pour une loi $t_{(18)}$. Dès lors,

$$A=2+1.734\frac{2}{\sqrt{19}}=2.795 \quad (0.5\text{ point}) \tag{25}$$

Ainsi, si $\overline{X}_N=2.4$, au seuil de risque de 5%, on ne peut pas rejeter l'hypothèse nulle $a=2$.

Question 6 (2 points) Par définition de la puissance :

$$\mathit{Puissance}=\Pr\left[W \mid H_1\right] \tag{26}$$

Ou encore :

$$\mathit{Puissance}=\Pr\left[\overline{X}_N>A \;\middle|\; a=a_1\right] \quad 0.5\text{ point} \tag{27}$$

ou encore :

$$\mathit{Puissance}=\Pr\left[\frac{\overline{X}_N-a}{S_N/\sqrt{N}}>\frac{A-a}{S_N/\sqrt{N}}\;\middle|\; a=a_1\right] \quad 0.5\text{ point} \tag{28}$$

Donc finallement,

$$\mathit{Puissance}=1-G\left(\frac{A-a_1}{S_N/\sqrt{N}}\right) \quad 0.5\text{ point} \tag{29}$$

ou encore :

$$\mathit{Puissance}=1-G\left(\frac{a_0-a_1}{S_N/\sqrt{N}}+G^{-1}(1-\alpha)\right) \quad 0.5\text{ point} \tag{30}$$

Question 7 (1 point) On vérifie que la puissance croît avec l'écart (positif) a_1-a_0 et que pour $a_1=a_0$ la fonction puissance (évaluée à tort sous H0) correspond au risque de première espèce α (0.5 point).

$$\mathit{Puissance}=1-G\left(\frac{2-3}{2/\sqrt{19}}+G^{-1}(0.95)\right)=0.6693 \tag{31}$$

où $G(.)$ désigne la fonction de répartition de la loi $t_{(18)}$.

Partie II : Test d'hypothèse simple contre hypothèse multiple

Question 1 (1 point) Nous avons montré que la région critique du test UPP d'hypothèse simple contre hypothèse simple $H_0: a=2$ contre $H_1: a=3$ ne dépendait pas de la valeur explicite de a_1. Par conséquent, il existe un test UPP pour $H_0: a=2$ contre $H_1: a>2$ et la région critique de ce test est identique à celle du test d'hypothèse simple :

$$W=\left\{X_1,..,X_{10} \mid \overline{X}_N<2.795\right\} \tag{32}$$

Question 2 (2 points) 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 :

$$\overline{W}=\left\{X_1,..,X_{10}\mid A<\overline{X}_N<B\right\} \tag{33}$$

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

$$A=a_0+G^{-1}\left(\frac{\alpha}{2}\right)\frac{S_N}{\sqrt{N}} \tag{34}$$

$$B=a_0+G^{-1}\left(1-\frac{\alpha}{2}\right)\frac{S_N}{\sqrt{N}} \tag{35}$$

où $G(.)$ désigne la fonction de répartition de la loi $t_{(18)}$. AN (0.5 point) :

$$A=2-2.10*\frac{2}{\sqrt{19}}=1.036 \tag{36}$$

$$B=2+2.10*\frac{2}{\sqrt{19}}=2.964 \tag{37}$$

Donc la région critique du test bilatéral est :

$$W=\left\{X_1,..,X_{10}\mid \overline{X}_N<1.036\ ou\ \overline{X}_N>2.964\right\}\quad 0.5\text{ point} \tag{38}$$

Si on observe une moyenne empirique de 2.4, pour un risque de première espèce de 5%, on ne peut pas rejetter l'hypothèse nulle $a_0=2$ (0.5 point).

Question 3 (2 points) La puissance est définie par :

$$P(a)=\Pr\left[W\mid H_1\right]=1-\Pr\left[1.036<\overline{X}_N<2.964\mid H_1\right]\quad 0.5\text{ point} \tag{39}$$

D'où l'on tire que :

$$P(a)=1-\Pr\left[\frac{1.036-a}{S_N/\sqrt{N}}<\overline{X}_N<\frac{2.964-a}{S_N/\sqrt{N}}\;\middle|\; H_1:a>2\right]\quad 0.5\text{ point} \tag{40}$$

Finallement :

$$P(a)=1-G\left(\frac{2.964-a}{S_N/\sqrt{N}}\right)+G\left(\frac{1.036-a}{S_N/\sqrt{N}}\right)\quad 0.5\text{ point} \tag{41}$$

où $G(.)$ désigne la fonction de répartition de la loi $t_{(18)}$. AN :

$$P(3)=1-G\left(\frac{2.964-3}{2/\sqrt{19}}\right)+G\left(\frac{1.036-3}{2/\sqrt{19}}\right) \tag{42}$$

$$P(3)=0.5311\quad 0.5\text{ point} \tag{43}$$

On vérifie que la puissance est plus faible que le dans le cas du test unilatéral correspondant.

Question 4 (2 points) On sait que :

$$Z_N=\frac{\overline{X}_N-a}{S_N/\sqrt{N}}\sim t_{(N-1)}\quad 0.5\text{ point} \tag{44}$$

Dès lors,

$$IC_{\alpha\%} = \left[\overline{X}_N \pm G^{-1}(1-\alpha) \frac{S_N}{\sqrt{N}} \right] \quad (0.5 \text{ point}) \tag{45}$$

$$= \left[2.4 \pm 2.10 \frac{2}{\sqrt{19}} \right] \tag{46}$$

$$= [1.60; 3.19] \quad 1 \text{ point} \tag{47}$$

Exercice 2 *Analyse de retour de campagne marketing. Barème : 4 points.*

Une entreprise souhaite analyser l'impact d'une campagne marketing suivant les canaux de diffusion utilisés (emailing, courriers et appels téléphoniques). Pour cela, elle dispose du tableau de contingence des achats moyens en euros pour 1500 clients ayant été contactés par l'un des trois médias. Exemple : 120

$X \setminus Y$	Emailing	Courriers	Appels	Total
50-100 euros	220	200	50	470
100 - 200 euros	140	250	100	490
plus de 200 euros	140	50	350	540
Total	500	500	500	1500

On teste :

$$H_0 : X \text{ et } Y \text{ sont indépendants} \tag{48}$$

Tableau des effectifs théoriques (1.5 point avec explication de la construction) :

$X \setminus Y$	Emailing	Courriers	Appels	Total
50-100 euros	156.66	156.66	156.66	470
100 - 200 euros	163.33	163.33	163.33	490
plus de 200 euros	180	180	180	540
Total	500	500	500	1500

Tous les effectifs théoriques sont supérieurs à 4. Statistique de test (1 point) :

$$D_N = \sum_i \sum_j \frac{(N_i - N\widehat{p}_i)^2}{N\widehat{p}_i} = 447.42 \tag{49}$$

Sous H_0 :

$$D_N \underset{N\to\infty}{\overset{d/H_0}{\longrightarrow}} \chi^2(4) \tag{50}$$

La région critique du test s'écrit :

$$W = \left\{ X \middle| D_N > G^{-1}(1-\alpha) \right\} \tag{51}$$

où $G(.)$ désigne la cdf de la loi $\chi^2(4)$. Donc pour un risque de 10% :

$$W = \{ X | D_N > 7.77 \} \quad (1 \text{ point}) \tag{52}$$

Pour un niveau de risque de 10%, on rejette l'hypothèse nulle d'indépendance entre les médias et le montant moyen des achats. (0.5 point)