Université d'Orléans - Licence Economie et Gestion
Statistique Appliquée à l'Economie et la Gestion

C. Hurlin. Correction de l'Examen de Décembre 2011

Exercice 1 *Barème : 18 points. Test de validation d'intervalle de confiance*

Partie I : Test d'hypothèses simples (13 points)

Question 1 (1 point) La variable I_t est une variable de Bernouilli :

$$I_t(\alpha) = \begin{cases} 1 & \text{avec } \Pr[I_t(\alpha) = 1] = \Pr[X_t \notin C_t(\alpha)] = \alpha \\ 0 & \text{avec } \Pr[I_t(\alpha) = 0] = \Pr[X_t \in C_t(\alpha)] = 1 - \alpha \end{cases}. \tag{1}$$

Dès lors

$$\mathbb{E}[I_t(\alpha)] = 1 \times \Pr[I_t(\alpha) = 1] + 0 \times \Pr[I_t(\alpha) = 0] = \alpha \tag{2}$$

L'hypothèse nulle H_0 est l'hypothèse correspond à la validité de l'intervalle de confiance : la probabilité d'obtenir une réalisation en dehors de l'intervalle est égal au risque α.

Question 2 (1 point) La probabilité d'obtenir une réalisation en dehors de l'intervalle est supérieure au risque α (0.5 point). Par exemple, on construit un intervalle de confiance à 95% et la probabilité d'obtenir une réalisation en dehors de l'intervalle est supérieure à 5%. Cela signifie que l'intervalle de confiance a une amplitude trop importante par rapport à ce qu'il devrait être s'il était correctement spécifié (0.5 point).

Question 3 (1 point) On admet que de façon générale la variable I_t suit une loi de Bernouilli de probabilité p, où p désigne la probabilité $\Pr(I_t = 1)$. La vraisemblance de l'échantillon $\{I_1, .., I_N\}$ s'écrit

$$\begin{aligned} L(I_1, .., I_N; p) &= \prod_{i=1}^{N} \Pr(I_t = 1)^{I_i} \times \Pr(I_t = 0)^{1-I_i} \quad \text{(0.5 point)} \\ &= p^H \times (1-p)^{N-H} \quad \text{(0.5 point)} \end{aligned} \tag{3}$$

où H désigne le nombre total de violations :

$$H = \sum_{t=1}^{N} I_t \tag{4}$$

Question 4 (1 point) Le rapport des vraisemblances de l'échantillon sous les hypothèses nulle et alternative peut s'écrire sous la forme suivante :

$$\begin{aligned} \frac{L(I_1, .., I_N; \alpha)}{L(I_1, .., I_N; \beta)} &= \frac{\alpha^H \times (1-\alpha)^{N-H}}{\beta^H \times (1-\beta)^{N-H}} \\ &= \left(\frac{\alpha}{\beta}\right)^H \left(\frac{1-\alpha}{1-\beta}\right)^{N-H} \end{aligned} \tag{5}$$

Question 5 (2 points) On sait que selon le lemme de Neyman Pearson, la région critique du test UPP de niveau α est telle que :

$$W = \left\{ X_1, .., X_{10} | \; \frac{L(I_1, .., I_N; \alpha)}{L(I_1, .., I_N; \beta)} < K \right\} \quad \text{(0.5 point)} \tag{6}$$

Donc :

$$\left(\frac{\alpha}{\beta}\right)^H \left(\frac{1-\alpha}{1-\beta}\right)^{N-H} < K \tag{7}$$

$$\Longleftrightarrow H \ln\left(\frac{\alpha}{\beta}\right) + (N-H)\ln\left(\frac{1-\alpha}{1-\beta}\right) < \ln(K) \tag{8}$$

$$\Longleftrightarrow H \left[\ln\left(\frac{\alpha}{\beta}\right) - \ln\left(\frac{1-\alpha}{1-\beta}\right)\right] < K' = \ln(K) - N\ln\left(\frac{1-\alpha}{1-\beta}\right) \quad \text{(0.5 point)} \tag{9}$$

Puisque $\alpha < \beta$, on a

$$\ln\left(\frac{\alpha}{\beta}\right) - \ln\left(\frac{1-\alpha}{1-\beta}\right) < 0 \tag{10}$$

Dès lors, on a une région critique de la forme :

$$H > K" = \left[\ln\left(\frac{\alpha}{\beta}\right) - \ln\left(\frac{1-\alpha}{1-\beta}\right)\right]^{-1} \left[\ln(K) - N\ln\left(\frac{1-\alpha}{1-\beta}\right)\right] \quad \text{(0.5 point)} \tag{11}$$

ou encore

$$\frac{H}{N} > \frac{K"}{N} = A \tag{12}$$

La région critique du test UPP au sens de Neyman Pearson est de la forme :

$$W = \left\{ I_1, .., I_N | \; \overline{I_N} > A \right\} \quad \text{(0.5 point)} \tag{13}$$

où $\overline{I_N} = (1/N)\sum_{t=1}^{N} I_t$ désigne la fréquence empirique des violations et où A est une constante déterminée par le niveau de risque de première espèce.

Question 6 (1 point) On admet que sous l'hypothèse nulle $\mathbb{E}(I_t) = \alpha$, on a $\mathbb{V}(I_t) = \alpha(1-\alpha)$. Donc (0.5 point):

$$\mathbb{E}\left(\overline{I_N}\right) = \mathbb{E}\left(\sum_{t=1}^{N} I_t\right) = \alpha \tag{14}$$

$$\mathbb{V}\left(\overline{I_N}\right) = \mathbb{V}\left(\sum_{t=1}^{N} I_t\right) = \frac{\alpha(1-\alpha)}{N} \tag{15}$$

Sous l'hypothèse nulle H_0, (0.5 point)

$$\overline{I_N} \sim N(\alpha, \alpha(1-\alpha)) \tag{16}$$

Question 7 (2 points) Le seuil critique A du test UPP pour un niveau de risque de première espèce $\gamma\%$ vérifie :

$$\gamma = \Pr\left[W \mid H_0\right] \text{ (0.5 point)} \tag{17}$$

Ou encore :

$$\gamma = \Pr\left[\overline{I_N} > A \mid \mathbb{E}(I_t) = \alpha\right] \tag{18}$$

On sait que sous H_0, la fréquence empirique suit une loi normale centrée sur α :

$$\overline{I_N} \sim N\left(\alpha, \frac{\alpha(1-\alpha)}{N}\right) \quad (0.5 \text{ point}) \tag{19}$$

Par conséquent :

$$\gamma = 1 - \Pr\left[\frac{\overline{I_N} - \alpha}{\sqrt{\alpha(1-\alpha)}/\sqrt{N}} < \frac{A - \alpha}{\sqrt{\alpha(1-\alpha)}/\sqrt{N}} \middle| \; \frac{\overline{I_N} - \alpha}{\sqrt{\alpha(1-\alpha)}/\sqrt{N}} \sim N(0,1) \right] \;\; (0.5 \text{ point}) \tag{20}$$

ce qui implique que :

$$1 - \gamma = \Phi\left(\frac{A - \alpha}{\sqrt{\alpha(1-\alpha)}/\sqrt{N}}\right) \tag{21}$$

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

$$A = \alpha + \Phi^{-1}(1-\gamma)\sqrt{\frac{\alpha(1-\alpha)}{N}} \;\; (0.5 \text{ point}) \tag{22}$$

Question 8 (2 points) On teste $H_0 : \mathbb{E}(I_t) = 0.1$ contre $H_1 : \mathbb{E}(I_t) = 0.15$. Donc, la région critique est définie par :

$$W = \left\{ I_1, .., I_N | \; \overline{I_N} < \alpha + \Phi^{-1}(1-\gamma)\sqrt{\frac{\alpha(1-\alpha)}{N}} \right\} \tag{23}$$

AN :

$$A = 0.1 + \Phi^{-1}(0.95)\sqrt{\frac{0.1 \times 0.9}{10}} = 0.2560 \;\; (0.5 \text{ point}) \tag{24}$$

$$W = \left\{ I_1, .., I_N | \; \overline{I_N} > 0.2560 \right\} \tag{25}$$

Sur le Figure 1, on observe une violation pour 10 prévision donc

$$\overline{I_N} = \frac{1}{10} = 0.1 \;\; (0.5 \text{ point}) \tag{26}$$

Donc, pour un risque de première espèce γ de 5%, on ne peut pas rejeter l'hypothèse nulle que l'intervalle de confiance est valide. (1 point)

Question 9 (2 points) La puissance du test est :

$$Puissance = \Pr[W \mid H_1] \;\; (0.5 \text{ point}) \tag{27}$$

D'où l'on tire que :

$$Puissance = \Pr\left[\frac{\overline{I_N} - \beta}{\sqrt{\beta(1-\beta)}/\sqrt{N}} > \frac{A - \beta}{\sqrt{\beta(1-\beta)}/\sqrt{N}} \;\middle|\; \frac{\overline{I_N} - \beta}{\sqrt{\beta(1-\beta)}/\sqrt{N}} \sim N(0,1) \right] \tag{28}$$

$$\iff Puissance = 1 - \Phi\left(\frac{A - \beta}{\sqrt{\beta(1-\beta)}/\sqrt{N}}\right) \;\; (1 \text{ point}) \tag{29}$$

AN :

$$\begin{aligned} Puissance &= 1-\Phi\left(\frac{0.2560-0.15}{\sqrt{0.15\,(1-0.15)}/\sqrt{10}}\right) && (30)\\ &= 1-\Phi\,(0.9391) && (31)\\ &= 0.1738 \quad (0.5 \text{ point}) && (32)\end{aligned}$$

On a, à peu près, une chance sur cinq de détecter un intervalle de confiance non valide.

Partie II : Test d'hypothèse simple contre hypothèse multiple (5 points)

Question 1 (1 point) Nous avons montré que la région critique du test UPP d'hypothèse simple contre hypothèse simple $H_0 : \mathbb{E}(I_t) = \alpha$ contre $H_1 : \mathbb{E}(I_t) = \beta$, avec $\alpha < \beta$ ne dépendait pas de la valeur explicite de β. Par conséquent, il existe un test UPP pour $H_0 : \mathbb{E}(I_t) = \alpha$ contre $H_1 : \mathbb{E}(I_t) > \alpha$ et la région critique de ce test est identique à celle du test d'hypothèse simple :

$$W = \left\{ I_1, .., I_N |\ \overline{I_N} > \alpha + \Phi^{-1}(1-\gamma)\sqrt{\frac{\alpha(1-\alpha)}{N}} \right\}, \tag{33}$$

Question 2 (2 points) Dans le cas de ce test unilatéral, la puissance est une fonction de la valeur de $\mathbb{E}(I_t) = \beta$ sous H_1 :

$$Puissance = 1-\Phi\left(\frac{A-\beta}{\sqrt{\beta\,(1-\beta)}/\sqrt{N}}\right) \quad (0.5\ \mathit{point}) \tag{34}$$

Pour $\beta = \alpha = 0.1$, par définition on retrouve le niveau de risque I :

$$P\,(0.1) = \alpha = 0.05 \quad 0.5 \text{ point} \tag{35}$$

Pour $\beta = 0.15$, nous avons montré que :

$$P\,(0.15) = 0.1738 \tag{36}$$

Enfin, on montre que :

$$P\,(0.2) = 1-\Phi\left(\frac{0.2560-0.2}{\sqrt{0.2\,(1-0.2)}/\sqrt{10}}\right) = 1-\Phi\,(0.4430) = 0.3289 \quad 0.5 \text{ point} \tag{37}$$

$$P\,(0.5) = 1-\Phi\left(\frac{0.2560-0.5}{\sqrt{0.5\,(1-0.5)}/\sqrt{10}}\right) = \Phi\,(1.5429) = 0.9386 \quad 0.5 \text{ point} \tag{38}$$

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

$$\overline{W} = \left\{ I_1, .., I_N |\ A < \overline{I_N} < B \right\}, \tag{39}$$

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

$$A = \alpha + \Phi^{-1}\left(\frac{\gamma}{2}\right)\sqrt{\frac{\alpha(1-\alpha)}{N}} \tag{40}$$

$$B = \alpha + \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)\sqrt{\frac{\alpha\,(1-\alpha)}{N}} \tag{41}$$

AN (0.5 point) :

$$A = 0.1 + \Phi^{-1}(0.025)\sqrt{\frac{0.1 \times 0.9}{10}} = -0.0859 \tag{42}$$

$$B = 0.1 + \Phi^{-1}(0.975)\sqrt{\frac{0.1 \times 0.9}{10}} = 0.2859 \tag{43}$$

Donc la région critique du test bilatéral est :

$$W = \left\{ I_1, .., I_N \mid \ \overline{I_N} < -0.0859 \ ou \ \overline{I_N} > 0.2859 \right\} \tag{44}$$

ou encore

$$W = \left\{ I_1, .., I_N \mid \ \overline{I_N} > 0.2859 \right\} \tag{45}$$

puisque $\overline{I_N}$ ne peut pas être négatif (ce nombre négatif provient de l'approximation normale de la loi binomiale, non justifiée dans ce cas). Si on observe une fréquence empirique de violation de 1/10, pour un risque de première espèce de 5%, on ne peut pas rejetter l'hypothèse nulle de validité de l'intervalle de confiance (0.5 point).

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

On teste :

$$H_0 : X \text{ et } Y \text{ sont indépendants} \tag{46}$$

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

$X \setminus Y$	canal A	canal B	canal C	Total
150-300 euros	230	230	230	690
300 - 600 euros	123.33	123.33	123.33	370
plus de 600 euros	146.67	146.67	146.67	440
Total	500	500	500	1500

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

$$D_N = \sum_i \sum_j \frac{\left(N_i - N\widehat{p}_i\right)^2}{N\widehat{p}_i} = 362.34 \tag{47}$$

Sous H_0 :

$$D_N \underset{N\to\infty}{\overset{d/H_0}{\longrightarrow}} \chi^2(4) \tag{48}$$

La région critique du test s'écrit :

$$W = \left\{ X \mid D_N > G^{-1}(1-\alpha) \right\} \tag{49}$$

où $G(.)$ désigne la cdf de la loi $\chi^2(4)$. Donc pour un risque de 5% :

$$W = \{ X \mid D_N > 9.49 \} \ \text{(1 point)} \tag{50}$$

Pour un niveau de risque de 5%, on rejette l'hypothèse nulle d'indépendance entre le canal de distribution et le montant moyen des achats. (0.5 point)