

Université d'Orléans - Licence Economie et Gestion

Statistique Mathématique

C. Hurlin. Examen Terminal Janvier 2008

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

Une banque souhaite mettre en place sur Internet une procédure systématique d'attribution de crédit à la consommation. A partir d'un fichier de clients existants, ayant déjà souscrit un prêt, elle définit plusieurs classes d'individus correspondants de profils de risque différents. Parmi celles-ci, elle distingue la catégorie des artisans, âgés de plus de 45 ans, célibataires, et travaillant dans le secteur du bâtiment. La question posée par la direction des risques de la banque est de savoir s'il convient ou non d'accorder systématiquement un prêt à la consommation à tous les individus de cette catégorie.

Pour répondre à cette question, vous disposez d'un N -échantillon aléatoire d'anciens clients appartenant à cette catégorie. On note Y_i , pour $i = 1, \dots, N$, la variable dichotomique telle que :

$$Y_i = \begin{cases} 1 & \text{si le client } i \text{ a connu un incident de paiement} \\ 0 & \text{sinon} \end{cases} \quad (1)$$

On suppose que les variables $\{Y_1, \dots, Y_N\}$ sont *i.i.d.* et telles que :

$$\Pr(Y_i = 1) = p \quad (2)$$

où $p \in]0, 1[$ désigne la probabilité de défaut (au sens incident de paiement), supposée homogène, des clients de cette catégorie. La banque souhaite mettre en place une règle de décision quant à l'attribution ou non du crédit à la consommation fondée sur cette probabilité de défaut.

Question 1 (1.5 point) Proposez à la banque (i) un test d'hypothèse simple contre hypothèse simple sur p , puis (ii) un test d'hypothèse simple contre hypothèse multiple unilatérale répondant à son attente. *Vous expliquerez la signification économique de l'hypothèse nulle et justifierez votre choix.*

Question 2 (1 point) On considère le test

$$H_0 : p = p_0 \quad (3)$$

$$H_1 : p = p_1 \quad (4)$$

avec $p_1 > p_0$. Expliquez la signification économique du risque de première espèce et de la puissance de ce test au responsable du département des risques.

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 $\{Y_1, \dots, Y_N\}$. On admet que la région critique du test UPP de niveau α de l'hypothèse $H_0 : p = p_0$ contre $H_1 : p = p_1 > p_0$ est de la forme :

$$W = \{Y_1, \dots, Y_N \mid Z_N > C\} \quad (5)$$

où C désigne une constante déterminée par le risque de première espèce. En utilisant le théorème central limite, démontrez que dans le cas d'un échantillon de taille asymptotique ($N \rightarrow \infty$) le seuil critique C est défini par :

$$C = p_0 + \Phi^{-1}(1 - \alpha) \sqrt{\frac{p_0(1 - p_0)}{N}} \quad (6)$$

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

Question 4 (1 point) La banque vous demande un premier diagnostic quant à la décision qu'elle doit prendre d'accorder ou non de façon systématique un prêt à la consommation à tous les individus de la catégorie étudiée. La banque souhaite limiter son risque de première espèce à 10%. Pour cela, elle vous fournit les éléments suivants :

$$p_0 = 0.2 \quad p_1 = 0.3 \quad N = 100 \quad \sum_{i=1}^{100} Y_i = 23 \quad (7)$$

Quelle décision conseillez vous à la banque concernant l'octroi de crédits à la consommation aux clients appartenant de cette catégorie ?

Question 5 (1 point) A partir des éléments de la question 4, déterminez la puissance du test proposé. Expliquez la signification économique de ce résultat. Pensez vous que ce résultat satisfiera les responsables de la banque ?

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 . Montrez lui que pour cela la nouvelle règle de décision doit être fondée sur un échantillon de taille N vérifiant :

$$N = \left[\frac{\Phi^{-1}(1 - \gamma) \sqrt{p_1(1 - p_1)} - \Phi^{-1}(1 - \alpha) \sqrt{p_0(1 - p_0)}}{p_0 - p_1} \right]^2 \quad (8)$$

Quelle taille d'échantillon convient-il d'adopter pour garantir 90% de puissance pour un niveau de risque de première espèce de 10% ?

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

$$H_0 : p = p_0 \quad (9)$$

$$H_1 : p > p_0 \quad (10)$$

Pour un échantillon de taille $N = 100$, quelle est la région critique du test UPP de niveau $\alpha = 10\%$ correspondant ?

Question 8 (1 point) Déterminez la fonction puissance de ce test en fonction des valeurs de p et des paramètres α , p_0 et N . 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%, quelle est la puissance de votre test ?

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

$$H_0 : p = 0.2 \quad (11)$$

$$H_1 : p \neq 0.2 \quad (12)$$

Si pour un échantillon de taille $N = 100$, on observe $\sum_{i=1}^{100} Y_i = 23$, que concluez vous pour un risque de première espèce de 5% ?

Exercice 2 *Test d'Indépendance et Relation Salaire / Diplôme. Barème : 4 points.*

On considère un échantillon établi par le CEREQ (Centre d'Etudes et de Recherches sur les Qualifications) et constitué de 705 jeunes peu diplômés sortis du système scolaire en juin 1989. On s'intéresse à la liaison entre le niveau de salaire des jeunes en euros (noté X) et leur niveau de formation (variable Y). On vous demande de tester au seuil de 10% l'indépendance du niveau de salaire des jeunes à leur niveau de formation. *Vous détaillerez précisément votre démarche.*

$X \setminus Y$	Bac	BEP-CAP	Sixième	Total
600-750	115	284	30	429
750-900	65	109	11	185
900-1650	45	44	2	91
Total	225	437	43	705

Exercice 3 *Maximum de Vraisemblance. Barème : 4 points.*

On suppose qu'au cours de l'année écoulée, le chiffre d'affaires des entreprises dont le chiffre d'affaire est inférieur à c , peut être correctement représenté par une variable aléatoire X absolument continue dont la densité de probabilité est définie, pour tout $X \in [0, c]$ comme suit :

$$f(X) = \frac{1}{\theta c^{\frac{1}{\theta}}} X^{\frac{1}{\theta}-1} \quad \text{pour } 0 \leq X \leq c \quad (13)$$

où c et θ désignent deux paramètres réels strictement positifs. La valeur du paramètre c est connue, mais non celle du paramètre θ . On a constitué au hasard et avec remise un échantillon de N entreprises, parmi celles dont le chiffre d'affaires est inférieur à c , et déterminé leur chiffre d'affaires au cours de l'année écoulée. Soit X_i pour $i = 1, \dots, N$ le chiffre d'affaire de l'entreprise i . On suppose que les variables X_1, X_2, \dots, X_N sont *i.i.d.*

Question 1 (1 point) Déterminez la fonction de vraisemblance $L(X_1, X_2, \dots, X_n; \theta)$.

Question 2 (1 point) Montrez que l'estimateur $\hat{\theta}$ de θ obtenu par la méthode du maximum de vraisemblance est :

$$\hat{\Theta}_n = \ln(c) - \frac{1}{n} \sum_{i=1}^n \ln(X_i) \quad (14)$$

Question 3 (2 points) On pose $Y = \ln(c) - \ln(X)$ et l'on admet que $E(Y) = \theta$ et $V(Y) = \theta^2$. L'estimateur $\hat{\Theta}_n$ est-il sans biais ? Est-il convergent ?