Cursus Intégré
2004-2005

Rappels de statistique mathématique

Enoncé des travaux dirigés n°7

Guillaume Lacôte

Bureau **E03**

Guillaume.Lacote@ensae.fr

http://ensae.no-ip.com/SE222/

Version du 20050421-15h38, révisée le 21 avril 2005

Enoncé de l'exercice 1

Pour étudier l'arrivée des appels dans un central téléphonique, on comptabilise lors de 200 observations consécutives, le nombre d'appels observés par seconde, ce qui produit les résultats suivants :

Nombre d'appels par seconde	Effectifs observés
0	6
1	15
2	40
3	42
4	37
5	30
6	10
7	9
8	5
9	3
10	2
11	1

On suppose les arrivées des appels indépendantes, et en outre que la probabilité élémentaire λdt qu'il arrive un appel entre les instants t et $t + dt$ est indépendante de la date t.

Autrement dit

- le nombre N_t d'appels observés sur l'intervalle de temps $[0, t]$ suit une loi de Poisson de paramètre λt.
- plus généralement $N_{t+s} - N_t$ est indépendant de N_t et suit une loi de Poisson de paramètre λs.

☞ Q1 (a) Tester l'adéquation de la loi empirique à la famille des lois de Poisson.

(b) Quel est le niveau limite permettant d'accepter l'adéquation à la loi de Poisson ?

☞ Q2 A une date antérieure le nombre d'appels sur un intervalle de temps $[0, t]$ suivait une loi de Poisson de paramètre 4.
Tester la stabilité du comportement entre les deux dates.

☞ Q3 Que donnerait un test d'adéquation à une loi de Poisson de paramètre $3,7$? Conclure.

Table de la loi de Poisson :

	$\mathbb{P}_{3,7}(X = k)$	$\mathbb{P}_4(X = k)$
0	0.0247	0,0183
1	0.0915	0,0733
2	0.1684	0,1465
3	0.2087	0,1957
4	0.1930	0,1954
5	0.1428	0,1563
6	0.0881	0,1042
7	0.0465	0,0595
8	0.0215	0,0298
9	0.0099	0,0132
10	0.0046	0,0053
11	0.0021	0,0019
12	0.0001	0,0006
13	< 0.0001	0,0002
14	< 0.0001	0,0001

Enoncé de l'exercice 2

On s'intéresse à la proportion des ménages équipés d'un magnétoscope. Pour cela, on tire de manière équiprobable avec remise un échantillon de n ménages, et on observe pour chaque ménage $i \in [\![1, n]\!]$ la variable

$$y_i = \begin{cases} 1 \text{ si le ménage } i \text{ est équipé} \\ 0 \text{ sinon.} \end{cases}$$

☞ Q1 Réaliser un test de l'hypothèse nulle : “la proportion des ménages équipés n'excède pas 20 %”.

☞ Q2 On se demande si la probabilité qu'un ménage i soit équipé n'est pas fonction d'une variable (scalaire) x_i donnée (revenu, âge du chef de famille…).

On définit à cet effet un modèle statistique conditionnel à X_i de la façon suivante :

$$\mathbb{P}(y_i = 1 \mid x_i) = \frac{e^{a+bx_i}}{1 + e^{a+bx_i}}$$

où a et b sont deux paramètres réels inconnus.

On cherche alors à tester $\underline{H_0 :}\ b = 0$.

(a) Calculer le score et la matrice d'information de Fisher du modèle pour n observations.

(b) Expliciter le modèle contraint par H_0.
Calculer les estimateurs du maximum de vraisemblance $\widehat{a^0}$ et $\widehat{b^0}$ de a et b dans ce modèle, puis évaluer le score en $\left(\widehat{a^0}, \widehat{b^0}\right)$.

(c) Donner un estimateur $\widehat{I_1^{H_0}}$ de I_1, convergent sous H_0 et fonction des estimateurs contraints des paramètres.

(d) Exprimer la statistique du test du score de l'hypothèse H_0.
Quelle est sa loi asymptotique sous H_0 ? Sur quelle corrélation repose le test ?

Enoncé de l'exercice 3

Un industriel reçoit K lots de n pièces, avec la garantie que la proportion p de pièces défectueuses est la même dans chaque lot et inférieure à 5%. Pour vérifier que la garantie est exacte, on tire avec remise des pièces dans chaque lot, jusqu'à obtenir une pièce défectueuse par lot.
Soit Y_k, la variable aléatoire désignant le nombre de tirages nécessaires dans le lot k.

☞ Q1 Calculer la loi de Y_k.
Calculer $\mathbb{E}(Y_k)$ et $\mathbb{V}(Y_k)$.

☞ Q2 Proposer un test de Wald de l'hypothèse :

$$\underline{H_0 :} \, p = 5\%$$

☞ Q3 Soit $A_k = \left\{\bar{Y} < k_\alpha\right\}$, pour k_α donné ; calculer $\mathbb{P}_p(A_k)$ en fonction de $p = p_0$.
Montrer que $\sup_{p \leq 5\%} \left(\lim_{K\infty} \mathbb{P}_p(A_k)\right) = \lim_{K\infty} \mathbb{P}_{p=5\%}(A_k)$.
En déduire un test asymptotique de l'hypothèse :

$$\underline{H_0 :} \, p \leq 5\%$$

Enoncé de l'exercice 4

On dispose de n observations i.i.d. d'un couple de variables aléatoires positives scalaires (W_i, X_i).
Le but de l'exercice est de suggérer une procédure pour tester l'hypothèse H_0 selon laquelle la loi conditionnelle de W_i sachant X_i est une loi de Pareto, de densité :

$$f(W|X, b) = \frac{bX}{W_0} \left(\frac{W_0}{W}\right)^{bX+1} \mathbb{1}_{W \geq W_0}$$

pour $b > 0$ et $W_0 > 0$

☞ Q1 On suppose dans un premier temps que les X_i sont tous égaux à $X \in \mathbb{R}$ connu, et que W_0 est connu, de sorte que le seul paramètre inconnu du modèle est $\alpha = bX + 1$.

(a) Calculer l'espérance et la variance de W, notées m et v.
A quelle condition sur α, supposée vérifiée par la suite, les moments m et v existent-ils ?

(b) Déterminer l'estimateur du maximum de vraisemblance $\widehat{\alpha}$ de α.

(c) Déterminer le score et l'information de Fisher du modèle.
En déduire $\mathbb{E}(\ln W)$ et $\mathbb{E}((\ln W)^2)$.

(d) Soit $s = \frac{1}{n}\sum_{i=1}^{n}\left(\ln(W_i) - \overline{\ln W}\right)^2$, où $\overline{\ln W}$ est la moyenne empirique des logarithmes de W_i.
On se propose de fonder un test de H_0 sur la différence :

$$s - \frac{1}{(\widehat{\alpha}-1)^2}$$

Justifier un tel test.

(e) Ecrire le Théorème Central Limite pour $\left(\ln W_i, (\ln W_i)^2\right)_{i \in [\![1,N]\!]}$, et tester l'hypothèse H_0 (On donnera la forme de la matrice de variance-covariance sans pour autant la calculer explicitement).

☞ Q2 On suppose toujours W_0 connu, mais les X_i prennent désormais des valeurs a priori distinctes.

(a) Calculer l'estimateur du maximum de vraisemblance de b.

(b) Montrer que : $\mathbb{E}\left(X^2\left(\ln\frac{W_0}{W}\right)^2 - \frac{2X}{b}\ln\frac{W_0}{W}\right) = 0$.
Comment testeriez-vous alors l'hypothèse H_0 ?

☞ Q3 On suppose enfin que W_0 est inconnu.

(a) Calculer l'estimateur du maximum de vraisemblance de W_0.

(b) Peut-on adapter la procédure de test précédemment mise-en-œuvre ?