ENSAE

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

Cursus Intégré
2004-2005

Rappels de statistique mathématique
Enoncé des travaux dirigés n°1

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

On dispose d'observations Y_i relatives au comportement de remboursement ou de non-remboursement d'emprunteurs :

$$Y_i = \begin{cases} 1 & \text{si l'emprunteur } i \text{ rembourse son crédit} \\ 0 & \text{si l'emprunteur } i \text{ est défaillant} \end{cases}$$

Afin de modéliser ce phénomène, on suppose l'existence d'une variable aléatoire Y_i^* normale, d'espérance m et de variance σ^2, qu'on appellera "capacité de remboursement de l'individu i", telle que :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* \geq 0 \\ 0 & \text{si } Y_i^* < 0 \end{cases}$$

☞ Q1 On note Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$.
Exprimer la loi de Y_i en fonction de Φ.

☞ Q2 Les paramètres m et σ^2 sont-ils identifiables ?

Enoncé de l'exercice 2

Un système S fonctionne en utilisant deux machines de types différents. Les durées de vie X_1 et X_2 des deux machines suivent des lois exponentielles de paramètres λ_1 et λ_2. Les variables aléatoires X_1 et X_2 sont supposées indépendantes.

☞ Q1 Montrer que

$$X \rightsquigarrow \mathcal{E}(\lambda) \Leftrightarrow \forall x \in \mathbb{R},\ \mathbb{P}(X > x) = \exp(-\lambda x)$$

☞ Q2 Calculer la probabilité pour que le système ne tombe pas en panne avant la date t.
En déduire la loi de la durée de vie Z du système.
Calculer la probabilité pour que la panne du système soit due à une défaillance de la machine 1.

☞ Q3 On dispose de n systèmes $S_1, ..., S_n$ identiques dont on observe les durées de vie $Z_1, ...Z_n$.

(a) Ecrire le modèle statistique correspondant et la vraisemblance des observations.
A-t-on suffisamment d'information pour estimer λ_1 et λ_2 ?

(b) Si on observe à la fois les durées de vie des systèmes et la cause de la défaillance (machine 1 ou 2), écrire le modèle statistique correspondant et la vraisemblance des observations.
A-t-on alors suffisamment d'information pour estimer λ_1 et λ_2 ?

☞ Q4 Dans cette question, on considère un seul système S utilisant une machine de type 1 et une machine de type 2, mais on suppose que l'on dispose d'un stock de n_1 machines de type 1, de durées de vie $X_1^1, ..., X_1^{n_1}$, et d'un stock de n_2 machines de type 2, de durées de vie $X_2^1, ..., X_2^{n_2}$. Quand une machine tombe en panne, on la remplace par une machine du même type, tant que le stock de machines de ce type n'est pas épuisé. Quand cela arrive, on dit que le système S lui-même est en panne. On note toujours Z la durée de vie du système.

Le cas $n_1 = n_2 = 0$ correspond donc à la première question (pas de stock).

(a) Donner la loi de la somme de n variables indépendantes qui suivent une loi exponentielle de même paramètre λ.

(b) Ecrire Z en fonction des X_j^i et en déduire $P(Z \geq t)$ en fonction de certaines lois gamma, dont on précisera les paramètres.

On note alors N le nombre de machines (des deux types) sorties du stocks quand le système tombe en panne, et Z_0 la durée écoulée avant la première panne d'une machine. On note Z_i la durée écoulée entre la i-ème panne et la $(i+1)$-ème panne. La durée de vie totale du système est donc :

$$Z = \sum_{i=0}^{N} Z_i$$

La $(N+1)$-ème panne est donc la panne fatale au système.

(c) Montrer que les variables Z_i sont i.i.d. et donner leur loi.
On pourra utiliser (après l'avoir démontré) le résultat suivant :
Si X est une variable aléatoire de loi exponentielle de paramètre λ, alors

$$\mathbb{P}(X \geq s + t | X \geq s) = \mathbb{P}(X \geq t) = e^{-\lambda t}$$

(on dit que X est "sans mémoire").

(d) Préciser l'ensemble des valeurs possibles pour la variable N et en donner la loi.

(e) On admet que N et les Z_i sont indépendantes. Calculer $\mathbb{E}(Z|N)$ en fonction de N, λ_1 et λ_2.
Donner l'expression de $\mathbb{E}(Z)$ en fonction de $\mathbb{E}(N)$, λ_1 et λ_2.

Enoncé de l'exercice 3

Ecrire la vraisemblance et déterminer une statistique exhaustive pour un échantillon de n observations i.i.d. de lois :

☞ Q1 loi de Poisson de paramètre λ :

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}^*$$

☞ Q2 loi de Pareto de paramètres α et θ avec $\alpha > 1$, $\theta > 0$ de densité :

$$f(x) = \frac{\alpha - 1}{\theta} \left(\frac{\theta}{x}\right)^{\alpha} \mathbb{1}_{[\theta;+\infty[}(x)$$

☞ Q3 loi de Weibull de paramètre α et θ avec $\alpha > 0, \theta > 0$ de densité :

$$f(x) = \alpha\theta x^{\alpha-1} e^{-\theta x^{\alpha}} \mathbb{1}_{[0;+\infty[}(x)$$

☞ Q4 loi uniforme sur $[0, \theta]$ avec $\theta > 0$ inconnu.

Enoncé de l'exercice 4

On veut compter le nombre θ de poissons dans un lac fermé. Pour cela, on tire un poisson au hasard, on le marque et on le remet dans le lac. On tire un second poisson. S'il est déjà marqué, on en prend note et on le remet dans le lac. Sinon, on le marque à son tour et on le remet dans le lac. Et ainsi de suite.

On tire n poissons selon la procédure ci-dessus. Au n-ième tirage, l'observation consiste en une variable aléatoire Y_n qui vaut 1 si le poisson est déjà marqué, 0 sinon. Par définition, on a $Y_1 = 0$. Le but de l'exercice est de montrer que :

$$R_n = \sum_{i=1}^{n} Y_i$$

est une statistique exhaustive pour θ.

☞ Q1 Montrer que :

$$\mathbb{P}(Y_n = y_n, ..., Y_1 = y_1) = \begin{cases} & \mathbb{P}(Y_n = y_n | Y_{n-1} = y_{n-1}, \ldots, Y_1 = y_1) \\ \times & \mathbb{P}(Y_{n-1} = y_{n-1} | Y_{n-2} = y_{n-2}, ..., Y_1 = y_1) \\ \vdots & \\ \times & \mathbb{P}(Y_1 = y_1) \end{cases}$$

☞ Q2 Montrer que la loi conditionnelle de Y_n sachant $R_{n-1} = r_{n-1}$ est une Bernoulli de paramètre :

$$\frac{n - r_{n-1} - 1}{\theta}$$

et en déduire que la vraisemblance est proportionnelle à :

$$\prod_{i=1}^{n} \frac{(\theta - i + 1 + r_{i-1})^{1-y_i}}{\theta} \tag{1}$$

☞ Q3 Montrer que l'expression (1) se réécrit :

$$\frac{1}{\theta^n}\frac{(\theta-1)!}{(\theta-n-1+r_n)!}$$

☞ Q4 En déduire que R_n est une statistique exhaustive pour θ.