Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2007 – 2008

Statistiques I

Examen terminal, février 2008

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. Un magasin veut savoir s'il lui faut engager des vendeurs en plus le samedi. Pour ce faire, il dispose du nombre total de clients servis chaque samedi durant toute l'année qui précède, soit $(N_1, \dots, N_n)$ avec $n = 52$. Les capacités du magasin sont telles qu'il peut accueillir 1000 clients dans la journée au maximum. Le responsable du magasin se dit prêt accueillir un nouveau vendeur si la probabilité $p = P(N_i > 1000)$ (on suppose ici que les N_i ont toutes la même loi) de recevoir plus de 1000 clients est supérieure à 2.5%.

   (a) Dans un premier temps, et de façon naturelle, on modélise $(N_1, \dots, N_n)$ par une suite de variable aléatoires indépendantes identiquement distribuées (v.a.i.i.d.) suivant une loi de Poisson de paramètre θ. Déterminer une formule reliant p et θ. Il convient donc d'estimer θ pour savoir si $p > 2.5\%$. Après avoir précisé le modèle statistique et la mesure dominante, déterminer un estimateur $\widehat{\theta}_1$ efficace et sans biais de θ. Déterminer la variance de $\widehat{\theta}_1$.

   (b) Cependant (et contrairement à vous...) le responsable du magasin n'est pas adepte des logiciels de calcul numérique, et il est incapable de savoir à partir de cette estimation de θ si $p > 2.5\%$. Aussi a-t-il recours à une autre modélisation: $(N_1, \dots, N_n)$ est une suite de v.a.i.i.d. de loi gaussienne ayant le même couple (moyenne,variance) que la loi de Poisson précédente, soit (θ, θ). Préciser le modèle statistique et la mesure dominante. Montrer que le modèle est exponentiel et montrer que θ ne peut pas être estimé efficacement et sans biais. Montrer que l'estimateur du maximum de vraisemblance $\widehat{\theta}_2$ de θ est unique et vaut:

   $$\widehat{\theta}_2 := \sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n} N_i^2} - \frac{1}{2}.$$

   Montrer que $\widehat{\theta}_2$ vérifie un théorème de la limite centrale et en déduire une approximation du biais et de la variance de $\widehat{\theta}_2$ (on rappelle que $n = 52$ est... un grand nombre). Comparer les risques quadratiques de $\widehat{\theta}_1$ et de $\widehat{\theta}_2$ lorsque le modèle est gaussien. Quel estimateur faut-il préférer? Pouvait-on s'attendre à l'avance à une telle réponse?

   (c) Dans le cas de ce modèle gaussien, exprimer p en fonction de θ. En déduire une condition numérique sur θ pour que $p \geq 2.5\%$ (comme vous ne disposez pas de calculatrice vous pouvez approcher 1.96.. par 2 et faire d'autres approximations).

   (d) Déduire de ce qui précède, dans le cas gaussien, un test de niveau de confiance 95%, permettant de trancher entre l'hypothèse H_0: $p \geq 2.5\%$ contre H_1: $p < 2.5\%$.

2. On considère une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ définies sur le même espace de probabilité, et telle que pour tout $n \in \mathbf{N}^*$, la mesure de probabilité de $(X_1 \dots, X_n)$ admet une densité par rapport à la mesure de Lebesgue sur $]0, \infty[^n$ qui est:

   $$f_{X_1,\dots,X_n}(x_1, \dots, x_n) := \frac{1}{2}\left(\alpha^n e^{-\alpha\sum_{i=1}^n x_i} + \beta^n e^{-\beta\sum_{i=1}^n x_i}\right)$$

   où $\theta = (\alpha, \beta)$ est un couple de réels inconnus.

(a) Donner l'ensemble Θ des valeurs possibles pour θ.

(b) Montrer que $(X_1 \ldots, X_n)$ sont des variables aléatoires identiquement distribuées dont vous donnerez la densité marginale. Sont-elles indépendantes?

(c) On peut montrer (ne pas le faire!) que le modèle statistique induit par $(X_1 \ldots, X_n)$ n'est pas exponentiel. Donner une statistique $\widehat{T}$ exhaustive pour ce modèle.

(d) Soit h une fonction de $\mathbf{R}$ dans $\mathbf{R}$ telle que $\mathbb{E}_{\alpha,\beta}[h((\widehat{T})] = 0$ pour tout $(\alpha, \beta) \in \Theta$. Avec la décomposition naturelle $f_{X_1,\ldots,X_n} = \frac{1}{2}\,(f_\alpha^{(1)} + f_\beta^{(2)})$, montrer qu'alors

$$\int_{]0,\infty[^n} h(\widehat{T}(x)) f_\alpha^{(1)}(x) d\lambda(x) = - \int_{]0,\infty[^n} h(\widehat{T}(x)) f_\beta^{(2)}(x) = C$$

pour tout $(\alpha, \beta) \in \Theta$, avec C une constante réelle. Montrer que $f_\alpha^{(1)}$ et $f_\beta^{(2)}$) sont des densités appartenant à la famille exponentielle. En utilisant les résultats du cours, montrer que $C = h = 0$ p.s. Qu'en déduisez vous sur $\widehat{T}$?

(e) Déterminer $\mathbb{E}[\widehat{T}]$. Peut-on en déduire un estimateur non biaisé de (α, β)?

(f) Montrer qu'il existe un unique estimateur du maximum de vraisemblance de θ. Montrer que cet estimateur de θ n'est pas convergent.

(g) Pour $j = 1$ ou 2, soit $\widehat{m}_j := \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i^j$. Calculer l'espérance de $\widehat{m}_j$ pour $j = 1$ et 2. En déduire un estimateur $\widehat{\theta}$ de θ construit à partir de $\widehat{m}_1$ et $\widehat{m}_2$. Calculer la variance de $\widehat{m}_1$. L'estimateur $\widehat{\theta}$ est-il convergent?