Ecole Normale Supérieure
Gérard Biau
FIMFA — Statistique

CONTRÔLE TERMINAL — 26/11/2009

Durée 3h — Aucun document autorisé

Exercice 1

Soient $X_1, \ldots, X_n$ des variables aléatoires réelles i.i.d. de densité

$$f_\theta(x) = x\theta^{-x^2/2} \log\theta \, \mathbf{1}_{\mathbb{R}^+}(x),$$

avec $\theta > 1$ (log désigne le logarithme népérien). On **admettra** que

$$\mathbb{E}(X_i^2) = \frac{2}{\log\theta} \quad \text{et} \quad \mathbb{E}(X_i^4) = \frac{8}{(\log\theta)^2}.$$

On note

$$T_n = \sum_{i=1}^n X_i^2.$$

1. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ (on se contentera de montrer que la dérivée s'annule).
2. 2.a Montrer la convergence et la normalité asymptotique de la suite de variables aléatoires $T_n/(2n)$ (on précisera bien la limite et la loi limite).

   2.b En déduire la convergence et la normalité asymptotique de l'estimateur $\hat{\theta}_n$.
3. Calculer l'information de Fisher du modèle (à une observation) et vérifier ainsi que $\hat{\theta}_n$ est asymptotiquement efficace.
4. Soit X une variable aléatoire de densité $f_\theta(x)$. Calculer $\mathbb{E}(X)$ et en déduire un estimateur $\tilde{\theta}_n$ de θ fondé sur la méthode des moments.

Exercice 2

On suppose que l'on observe $X_1, \ldots, X_n$, i.i.d. et de même loi $\mathcal{N}(\mu, 1)$. On veut tester $H_0 : \mu = 0$ contre $H_1 : \mu = m < 0$.

1. Rappeler (sans démonstration) la forme du test de Neyman-Pearson de niveau $\alpha \in]0, 1[$ pour ce problème.
2. 2.a Calculer la fonction puissance $m \in \mathbb{R}^- \mapsto \pi_n(m)$ de ce test et tracer son graphe.

   2.b Etudier la convergence simple de π_n lorsque n tend vers $+\infty$. Peut–on parler de convergence uniforme sur $\mathbb{R}^-$?
3. On considère l'alternative dépendant de n

$$H_1 : \mu = -Cn^{-\gamma},$$

   avec $C > 0$ et $\gamma \in \mathbb{R}$. Etudier le comportement de la puissance du test sur H_1 en fonction de γ lorsque n tend vers $+\infty$.

Exercice 3

On admet dans cet exercice que si $X_1, \ldots, X_n$ sont des variables aléatoires réelles i.i.d. de fonction de répartition F et fonction de répartition empirique F_n, on a, pour $t > 0$,

$$\mathbb{P}\Big[\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| > t\Big] \leq 2\exp(-2nt^2).$$

1. On considère des observations i.i.d. $X_1, \ldots, X_n$ de densité

$$f_\theta(x) = \theta e^{-\theta x}\mathbf{1}_{[0,+\infty[}(x),$$

   et on note $F_\theta(x)$ leur fonction de répartition. Soit $h > -1$ et $\theta' = (1 + h)\theta$.

   1.a Calculer **soigneusement**

$$g(h) = \sup_{x\in\mathbb{R}} |F_\theta(x) - F_{\theta'}(x)| = \|F_\theta - F_{\theta'}\|_\infty$$

   en fonction de h. On prendra soin de distinguer les trois cas $h = 0$, $h > 0$ et $h < 0$.

   1.b Montrer que g est équivalente à $|h|/e$ lorsque $h \to 0$.

2. On appelle F_n la fonction de répartition empirique des X_i et on définit un estimateur $\hat{\theta}_n$ de θ par

$$\|F_{\hat{\theta}_n} - F_n\|_\infty = \inf_\theta \|F_\theta - F_n\|_\infty.$$

Soit θ_0 la vraie valeur du paramètre et $\mathbb{P}_{\theta_0}$ la probabilité correspondante. On pose

$$\frac{\hat{\theta}_n}{\theta_0} = 1 + \hat{h}_n$$

et on admet que g est une fonction strictement décroissante pour $h < 0$ et strictement croissante pour $h > 0$.

2.a Montrer que si $g(\hat{h}_n) > t$, alors $\|F_{\theta_0} - F_n\|_\infty > t/2$.

2.b Déduire de la question précédente une majoration explicite de

$$\mathbb{P}_{\theta_0}[g(\hat{h}_n) > t].$$

2.c Déduire alors des propriétés de g que la suite $\hat{\theta}_n$ est convergente en probabilité.

3. Donner finalement un intervalle de confiance de niveau $1 - \alpha$ pour θ_0 sous forme implicite. Donner la forme explicite approchée de cet intervalle lorsque n est grand.

Exercice 4

Rappels. Une variable aléatoire Y suit une loi bêta $\beta(a, b)$ ($a > 0$ et $b > 0$) si sa densité est donnée par

$$g(y) = \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}\mathbf{1}_{[0,1]}(y),$$

avec

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x.$$

On rappelle que

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

où

$$\Gamma(u) = \int_0^{+\infty} x^{u-1}e^{-x}\mathrm{d}x \quad (u > 0).$$

En outre, $\Gamma(u+1) = u\Gamma(u)$.

Enoncé. Une chaîne de télévision a programmé une **série** de K émissions hebdomadaires, stables dans la grille de programmes, c'est-à-dire passant toujours le même jour de la semaine et à la même heure.

Pour suivre l'audience, on dispose d'un panel de téléspectateurs indépendants entre eux, de taille n supposée élevée, de l'ordre de quelques milliers d'individus (en pratique, $n = 6000$). Ce panel permet, pour l'émission numéro k ($k = 1, \ldots, K$) de la série, de connaître le nombre (aléatoire) N_k d'individus ayant regardé cette émission k.

On désigne par p_k la probabilité de regarder l'émission numéro k de la série.

Première partie. Dans cette partie, le paramètre p_k est supposé **constant**, égal à $p \in]0, 1[$, sur l'ensemble du panel et pendant les K émissions de la série.

1. Pour k fixé entre 1 et K, on désigne par N_k la variable aléatoire qui compte le nombre de téléspectateurs de l'émission k. **On considère dans les trois questions qui suivent le modèle statistique à une observation** N_k.
   - 1.a Quelle est la loi de N_k ? Que valent $\mathbb{E}(N_k)$ et $\mathbb{V}(N_k)$?
   - 1.b Donner l'information de Fisher du modèle.
   - 1.c Calculer $\hat{p}_k$, l'estimateur du maximum de vraisemblance de p.
2. **On considère maintenant le modèle à** K **observations** $N_1, \ldots, N_K$, supposées **indépendantes**.
   - 2.a Calculer $\hat{p}$, l'estimateur de p par la méthode du maximum de vraisemblance.
   - 2.b Exprimer $\hat{p}$ en fonction de $\hat{p}_1, \ldots, \hat{p}_K$.
   - 2.c Calculer l'information de Fisher de ce modèle à K observations.
   - 2.d Donner la loi asymptotique de $\hat{p}$ lorsque n tend vers l'infini, K restant fixé. En déduire un intervalle de confiance approché, de niveau $1 - \alpha = 95\%$, pour p.

Seconde partie. Dans la première partie, on a supposé que tous les individus du panel avaient la même probabilité p de regarder une émission donnée de la série.

Cette hypothèse d'école est irréaliste. En effet, tous les panélistes n'ont pas les mêmes habitudes de comportement (présence au domicile, environnement familial, etc.). Nous faisons donc désormais, dans cette partie, les hypothèses suivantes :

H_1 : Le paramètre p est une variable aléatoire à valeurs dans $[0, 1]$, de loi bêta $\beta(a, b)$ ($a > 0$ et $b > 0$).

H_2 : La loi de probabilité de p ne varie pas d'une émission à l'autre. Quel que soit le rang k de l'émission ($k = 1, \ldots, K$), p suit donc toujours une loi $\beta(a, b)$.

3. Préciser la loi conditionnelle de N_k sachant p.
4. 4.a Calculer la loi de N_k.

   4.b Si Y suit une loi $\beta(a, b)$, que vaut $\mathbb{E}(Y)$?

   4.c Déduire de la question précédente la valeur de $\mathbb{E}(N_k)$.

   Dans les deux questions qui suivent, on désigne par X_1 le nombre d'émissions vues par le premier panéliste.
5. 5.a Déterminer la loi conditionnelle de X_1 sachant p, puis la loi de X_1.

   5.b Calculer $\mathbb{E}(X_1)$.
6. On note $p_K(x) = \mathbb{P}(X_1 = x)$. Etablir la relation de récurrence

$$p_{K-1}(x) = \frac{K-x}{K} p_K(x) + \frac{x+1}{K} p_K(x+1).$$

   On constate que cette relation de récurrence ne dépend pas du choix fait pour la loi a priori de p.
7. Etablir une relation de récurrence dépendant cette fois-ci du choix fait pour la loi de p.