Ecole Normale Supérieure
Gérard Biau
FIMFA — Statistique

CONTRÔLE TERMINAL — 9/01/2014

Durée 3h — Aucun document autorisé

Exercice 1

On considère un échantillon $X_1, \dots, X_n$ de variables aléatoires i.i.d., chacune de densité $f_{\theta,\ell}(x) = \frac{1}{\ell}\mathbf{1}_{[\theta,\theta+\ell]}(x)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, où $\theta > 0$ et $\ell > 0$ sont deux paramètres inconnus.

1. Calculer l'estimateur du maximum de vraisemblance $(\hat{\theta}_n, \hat{\ell}_n)$ de (θ, ℓ).
2. Montrer la convergence de cet estimateur.
3. 3.1 Déterminer la densité du couple
$$\left(\min_{i=1,\dots,n} X_i, \max_{i=1,\dots,n} X_i\right).$$
   3.2 En déduire la fonction de répartition de $\hat{\ell}_n$.
4. Soit $\alpha \in]0,1[$.
   4.1 Construire un intervalle de confiance bilatère de niveau $1-\alpha$ pour le paramètre ℓ.
   4.2 Proposer un test de niveau α pour l'hypothèse $\ell = \ell_0$ contre $\ell \neq \ell_0$ ($\ell_0 > 0$, fixé).

Exercice 2

Soit $X_1, \ldots, X_n$ des variables aléatoires i.i.d., admettant chacune la densité uniforme sur $[0, \theta]$, où $\theta > 1$ est un paramètre inconnu. On suppose que l'on n'a pas accès à ces observations, mais seulement à leurs versions tronquées

$$Y_i = \begin{cases} X_i & \text{si } X_i \geq 1 \\ 1 & \text{si } X_i < 1, \end{cases}$$

pour $i = 1, \ldots, n$.

1. Proposer un estimateur $\hat{\theta}_n$ de θ par la méthode des moments.
2. Donner la loi asymptotique de $\hat{\theta}_n$.

Exercice 3

On considère le modèle linéaire défini par les n équations

$$Y_i = \alpha W_i + \gamma Z_i + \delta + \varepsilon_i, \quad i = 1, \ldots, n,$$

où les W_i, Z_i sont des réels fixés, les ε_i sont des variables aléatoires réelles indépendantes de même loi $\mathcal{N}(0, \sigma^2)$, et α, γ, δ, σ^2 sont des paramètres réels inconnus ($\sigma^2 > 0$). On notera W, Y, Z les vecteurs de $\mathbb{R}^n$ de coordonnées respectives W_i, Y_i, Z_i, et $\bar{W} = \frac{1}{n}\sum_{i=1}^n W_i$, $\langle W, Z\rangle = \sum_{i=1}^n W_i Z_i$, $\|W\|^2 = \langle W, W\rangle$, et ainsi de suite pour les autres variables.

On suppose dans tout ce qui suit que $\bar{W} = \bar{Z} = 0$, $\|W\| = \|Z\| = a > 0$, $\langle W, Z\rangle = a^2 \sin\theta$, $\theta \in]-\frac{\pi}{2}, \frac{\pi}{2}[$.

1. Ecrire le modèle sous la forme matricielle classique $Y = X\beta + \varepsilon$, en précisant bien ce que sont X et β par rapport aux hypothèses.
2. Calculer l'estimateur du maximum de vraisemblance $\hat{\beta} = (\hat{\alpha}, \hat{\gamma}, \hat{\delta})^T$ de β en fonction de a, θ et de produits scalaires du type ci-dessus.
3. Donner la loi de $\hat{\beta}$.

   Dans la suite, on note $\hat{s}^2$ l'estimateur classique sans biais de σ^2.

4. Soit $\alpha \in]0, 1[$. Donner un intervalle de confiance bilatère pour γ au niveau $1-\alpha$. Quelle est l'espérance du carré de sa longueur ? Comment varie-t-elle en fonction de θ ?

5. Trouver la racine carrée positive de la matrice
$$\begin{pmatrix} 1 & \sin\theta \\ \sin\theta & 1 \end{pmatrix}.$$
6. Exhiber une matrice carrée B telle que si l'on définit
$$\begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} = B \begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix}$$
alors
$$\mathrm{Cov} \begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} = \sigma^2 I_{\mathrm{d}}.$$
7. Donner la loi de $t^2 = (\hat{u} - u)^2 + (\hat{v} - v)^2$, où
$$\begin{pmatrix} u \\ v \end{pmatrix} = B \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}.$$
8. Trouver une constante h telle que $ht^2/\hat{s}^2$ ait une loi classique que l'on précisera.
9. En utilisant ce qui précède, obtenir un domaine de confiance pour le vecteur $\begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$ dans $\mathbb{R}^2$ au niveau $1 - \alpha$.
10. Que devient le domaine précédent si $\theta = 0$? Que se passe-t-il si $\theta = \pi/2$?

Mini-problème 1

On considère dans ce problème des observations i.i.d. $X_1, \ldots, X_n$ admettant toutes la même densité f sur l'intervalle $[0, 1]$, par rapport à la mesure de Lebesgue. On s'intéresse à l'estimation non paramétrique de f à l'aide de l'estimateur *histogramme* $\hat{f}_n$.

Pour construire $\hat{f}_n$, on procède de la façon suivante : on se donne une suite croissante de points dans $[0, 1]$: $z_0 = 0 < z_1 < \cdots < z_D = 1$ (D est un nombre entier, supérieur ou égal à 1) et on pose, pour $x \in [0, 1]$,
$$\hat{f}_n(x) = \sum_{j=1}^{D} \frac{N_j}{n(z_j - z_{j-1})} \mathbf{1}_{I_j}(x),$$

avec

$$I_j = [z_{j-1}, z_j[\quad \text{et} \quad N_j = \sum_{i=1}^{n} \mathbf{1}_{I_j}(X_i).$$

On suppose dans tout le problème que $f \in L^2([0,1])$, de norme donnée par

$$\|f\|^2 = \int_0^1 f^2(x)\mathrm{d}x.$$

1. Expliquer brièvement comment fonctionne l'estimateur $\hat{f}_n$ (faire un dessin).
2. On pose, pour $j = 1, \ldots, N$,

$$p_j = \int_{I_j} f(x)\mathrm{d}x.$$

   Donner, pour $x \in I_j$, l'expression de $\mathbb{E}\hat{f}_n(x)$ puis celle de $\mathbb{V}\hat{f}_n(x)$.
3. On désigne par S l'espace vectoriel des fonctions constantes par morceaux sur les éléments de la partition, c'est-à-dire

$$S = \left\{ g = \sum_{j=1}^{D} a_j \mathbf{1}_{I_j} : a_j \in \mathbb{R} \text{ pour } j = 1, \ldots, D \right\}.$$

   Déterminer $\bar{f}$, la projection orthogonale de f sur S dans $L^2([0,1])$.
4. On définit le risque quadratique intégré de l'estimateur $\hat{f}_n$ (MISE en anglais, pour *Mean Integrated Squared Error*) par

$$\mathrm{MISE}(\hat{f}_n) = \mathbb{E}\|\hat{f}_n - f\|^2.$$

   Montrer alors que

$$\mathrm{MISE}(\hat{f}_n) = \|f - \bar{f}\|^2 + \mathbb{E}\|\bar{f} - \hat{f}_n\|^2.$$

   Comment s'interprète chacun des termes de cette égalité ?
5. Montrer que si f est bornée par M, on a

$$\mathrm{MISE}(\hat{f}_n) \leq \frac{M(D-1)}{n} + \|f - \bar{f}\|^2.$$

Afin d'aller plus loin, **nous supposerons désormais que tous les intervalles I_j ont la même longueur** $1/D$ ou, ce qui revient au même, que $z_j = j/D$, $j = 0, \ldots, D$.

6. 6.1 Etablir que

$$\mathrm{MISE}(\hat{f}_n) \leq \frac{D-1}{n} + \|f - \bar{f}\|^2.$$

   6.2 Pourquoi cette inégalité est-elle toujours meilleure que celle obtenue à la question 5 ?

7. On suppose la fonction f continue sur $[0, 1]$. En permettant à D de dépendre de n, donner alors des conditions sur la suite $(D_n)_{n\geq 1}$ assurant que

$$\mathrm{MISE}(\hat{f}_n) \to 0 \quad \text{lorsque } n \to \infty.$$

8. On suppose la fonction f lipschitzienne sur $[0, 1]$, de constante de Lipschitz L.

   8.1 Etablir que

$$\mathrm{MISE}(\hat{f}_n) \leq \frac{D-1}{n} + \frac{L^2}{D^2}.$$

   8.2 En déduire que, pour un choix de $D = D_n$ approprié,

$$\mathrm{MISE}(\hat{f}_n) = \mathrm{O}(n^{-2/3}).$$

9. Quels sont, selon vous, les principaux avantages et les principales faiblesses de l'estimateur $\hat{f}_n$?

Mini-problème 2

On considère une observation aléatoire $\mathbf{X}$ à valeurs dans $\mathbb{R}^d$, admettant une densité $f_\theta(\mathbf{x})$ par rapport à la mesure de Lebesgue. Ici, $\theta \in \mathbb{R}$ désigne un paramètre inconnu que l'on souhaite estimer. On choisit d'utiliser la notation $f(\mathbf{x}|\theta)$ en lieu et place de $f_\theta(\mathbf{x})$.

Dans le paradigme bayésien, on raisonne comme si le paramètre était luimême une variable aléatoire Θ à valeurs dans $\mathbb{R}$, la densité $f(\mathbf{x}|\theta)$ devenant ainsi la densité conditionnelle de $\mathbf{X}$ lorsque $\Theta = \theta$. Pour simplifier, on suppose jusqu'à nouvel ordre que la loi de Θ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$, de densité $\pi(\theta)$ appelée *densité a priori*. On rappelle que la loi conditionnelle de Θ sachant $\mathbf{X} = \mathbf{x}$ admet alors une densité $g(\theta|\mathbf{x})$ sur $\mathbb{R}$.

1. Rappeler l'expression de $g(\theta|\mathbf{x})$. Pourquoi appelle-t-on cette densité la *densité a posteriori* ?

Si $\delta(\mathbf{X})$ est un estimateur de θ basé sur $\mathbf{X}$, on appelle *risque de Bayes* de $\delta(\mathbf{X})$ la quantité

$$R_B\left(\delta(\mathbf{X})\right) = \int_{\mathbb{R}} \pi(\theta)\mathrm{d}\theta \int_{\mathbb{R}^d} |\delta(\mathbf{x}) - \theta|^2 f(\mathbf{x}|\theta)\mathrm{d}\mathbf{x}.$$

2. Expliquer l'idée derrière $R_B(\delta(\mathbf{X}))$.

On appelle *estimateur de Bayes* l'estimateur $\hat{\theta}(\mathbf{X})$ qui minimise en $\delta(\mathbf{X})$ le risque de Bayes.

3. Montrer que $\hat{\theta}(\mathbf{x})$ minimise en $\delta(\mathbf{x})$ la quantité

$$\int_{\mathbb{R}} |\delta(\mathbf{x}) - \theta|^2 g(\theta|\mathbf{x})\mathrm{d}\theta.$$

4. Donner alors l'expression de $\hat{\theta}(\mathbf{X})$.
5. **Application**. On suppose que $\mathbf{X} = (X_1, \ldots, X_n)$, où $X_1, \ldots, X_n$ sont des variables aléatoires réelles i.i.d., de loi commune $\mathcal{N}(\theta, 1)$. On suppose en outre que $\Theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, où $\mu_0 \in \mathbb{R}$ et $\sigma_0^2 > 0$ sont deux paramètres connus.
   - 5.1 Trouver l'estimateur de Bayes $\hat{\theta}(\mathbf{X})$.
   - 5.2 Cet estimateur est-il convergent ?
6. On suppose maintenant que $\mathbf{X} = \bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n X_i$, où $X_1, \ldots, X_n$ sont des variables aléatoires réelles i.i.d., de loi commune $\mathcal{N}(\theta, \sigma^2)$. Le paramètre $\theta \in \mathbb{R}$ est inconnu et $\sigma^2 > 0$ est connu. On complexifie un peu le modèle en supposant désormais que Θ admet une densité $\pi(\theta)$ par rapport à une mesure σ-finie ν sur $\mathbb{R}$, **éventuellement différente de la mesure de Lebesgue**.
   - 6.1 Etablir que la moyenne a posteriori de Θ sachant $\bar{\mathbf{X}} = \mathbf{x}$ est de la forme
     $$\gamma(\mathbf{x}) = \mathbf{x} + \frac{\sigma^2}{n}\frac{d\log\left(p(\mathbf{x})\right)}{d\mathbf{x}},$$
     où $p(\mathbf{x})$ est la densité marginale de $\bar{\mathbf{X}}$ (inconditionnelle en θ).
   - 6.2 Trouver des expressions explicites pour $p(\mathbf{x})$ et $\gamma(\mathbf{x})$ lorsque la loi a priori est $\mathcal{N}(\mu_0, \sigma_0^2)$ avec probabilité $1 - \varepsilon$ et un Dirac en μ_1 avec probabilité ε, où μ_0, μ_1 et $\sigma^2 > 0$ sont des constantes connues.