M2R Mathématiques appliquées
2010–2011
Modèlisation Statistique

Examen du 27 Janvier 2011 de 9h30 à 12h30

Les notes de cours sont autorisées. Les calculatrices sont autorisées.

1 Choix de Modèles

Pour $n \geq 2$, on considère le modèle linéaire gaussien de dimension $2n \times 3$, $Y = X\theta^* + \varepsilon$ où $\varepsilon \sim \mathcal{N}_{2n}(0, \sigma_*^2)$,

$$X := \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & (-1)^n & 1 \\ 1 & (-1)^{n+1} & -1 \\ 1 & (-1)^{n+2} & -1 \\ \vdots & \vdots & \vdots \\ 1 & (-1)^{2n} & -1 \end{pmatrix} \quad \text{et } \theta^* := \begin{pmatrix} \theta_1^* \\ \theta_2^* \\ \theta_3^* \end{pmatrix}.$$

Dans tout l'exercice on suppose que l'on a $\theta_1^* = \theta_3^* = 0$. Evidemment l'observateur des résultats du modèle, le dénomé Jack Istique ne le sait pas et va chercher à le découvrir par différents moyens.

1. Montrer que le modèle précédent rentre dans le cadre du modèle linéaire gaussien régulier. Calculer les estimateurs du maximum de vraisemblance $\widehat{\theta_n}$ et $\widehat{\sigma_n^2}$ de θ^* et σ_*^2.
2. Montrer que $\widehat{\theta_n}$ converge presque sûrement, quand n tend vers l'infini, vers θ^*.
3. Quelle est la loi jointe de $\left(\sqrt{2n}(\widehat{\theta_n} - \theta^*), 2n\frac{\widehat{\sigma_n^2}}{\sigma_*^2}\right)$? Décrire les régions de confiance pour θ^* et σ_*^2.
4. Pour $J \subset \{1,2,3\}$, on pose $\theta_J^* := (\theta_j^*)_{j\in J}$. Pour $J \subset \{1,2,3\}$ et $J \neq \emptyset$, décrire précisement la procédure de test de $H_0 \quad \theta_J^* = 0$ contre $H_1 \quad \theta_J^* \neq 0$.
5. On se propose de comparer les stratégies de choix de modèles du C_p de Mallows avec BIC. Rappelons tout d'abord les fonctions d'objectifs associées. Pour $J \subset \{1,2,3\}$, $|J|$ désigne le nombre d'éléments de J et $\widehat{\sigma_{n,J}^2}$ la variance estimée par maximum de vraisemblance dans le modèle à $|J|$ paramètres où le paramètre intervenant dans la moyenne des observations est θ_J. On a alors

$$C_p(J) := \frac{\widehat{\sigma_{n,J}^2}}{\widehat{\sigma_n^2}} + \frac{|J|}{n},$$
$$BIC(J) := \log(\widehat{\sigma_{n,J}^2}) + \frac{\log 2 + \log n}{2n}|J|.$$

Rappelons que l'on se place dans le cas où $J^* = \{2\}$. Les questions qui suivent seront traitées pour le BIC et C_p de Mallows

- Calculer la probabilité de sous-ajustement ainsi que sa limite quand n tend vers l'infini.
- Pour chacun des 3 sous-ensembles J de $\{1,2,3\}$ qui sur-ajustent J^* calculer les probabilités de se tromper de modèle en selectionnant le modèle J. Calculer les limites, quand n tend vers l'infini, de ces probabilités.
- On rappelle que si A et B sont des esembles mesurables on a $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. Donner un majorant de la probabilité de sur-ajustement et déterminer la limite, quand n tend vers l'infini, de cette borne.
- Que pensez-vous des critères BIC et C_p de Mallows pour le modèle étudié ici ?

2 Loi des Extrêmes

Soit $X_1, \ldots, X_n$ des variables i.i.d. de loi de Cauchy. On rappelle que la loi de Cauchy a pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad (x \in \mathbb{R}).$$

1. Calculer la fonction de répartition de X_1 notée F.
2. La fonction caratéristique de X_1 vaut pour $t \in \mathbb{R}$, $\exp(-|t|)$. Quelle est la loi de la moyenne empirique de $X_1, \ldots, X_n$? Quelle est la loi de l'inverse de la moyenne empirique de $X_1, \ldots, X_n$?
3. On considère la fonction
$$\psi(x) := \tan(x) + \tan\left(\frac{1}{x}\right),$$
définie sur $\mathbb{R}_*$. Calculer la dérivée de ψ sur $\mathbb{R}_*$ et en déduire que cette fonction ne prend que 2 valeurs distinctes. Soit $x \in \mathbb{R}$ et un entier naturel $n > x$, on pose $u_n(x) := (1 + \frac{x}{n})^n$. Calculer la limite, quand n tend vers l'infini, de $u_n(x)$.
4. On pose maintenant, pour $n \geq 1$, $X_{(n)} := \max_{i=1\ldots n} X_i$. Pour $t \in \mathbb{R}$, calculer la fonction de répartition de $X_{(n)}$. On note $F_{X_{(n)}}$ cette fonction de répartition. En utilisant la question précédente montrer que
$$\begin{aligned}\lim_{n\to\infty} F_{X_{(n)}}(nu) &= \exp\left(-\frac{1}{\pi u}\right), \quad (u > 0),\\ \lim_{n\to\infty} F_{X_{(n)}}(nu) &= 0, \quad (u \leq 0).\end{aligned}$$
En déduire que $n^{-1}X_{(n)}$ converge en loi vers une loi G à préciser.
5. Soit $Y_1 \ldots, Y_n$ des variables i.i.d. de loi exponentielle de paramètre 1. Montrer que $Y_{(n)} - \log n$ converge en loi vers une loi H à préciser.

3 Séries chronologiques

Soit $(\varepsilon_n)_{n\in\mathbb{Z}}$ une suite de variables i.i.d. de loi normale standard. Soit $a \in \mathbb{R}$, on considère le processus $(X_n)_{n\in\mathbb{Z}}$ défini par l'équation de récurrence :

$$X_{n+1} := \varepsilon_n + a\varepsilon_{n+1}, \quad (n \in \mathbb{N}).$$

1. Calculer pour $n, h \in \mathbb{Z}$ $\mathrm{cov}(X_n, X_{n+h})$. Montrer que $(X_n)_{n\in\mathbb{Z}}$ est un processus gaussien stationnaire.
2. Quelle est la densité spectrale de ce processus.
3. Soit N un entier naturel non nul. Exprimer $(X_1, \dots, X_N)$ en fonction de $(\varepsilon_0, \dots, \varepsilon_N)$ puis $(\varepsilon_0, \dots, \varepsilon_N)$ en fonction de $(\varepsilon_0, X_1, \dots, X_N)$. Déterminer la densité du vecteur aléatoire $(\varepsilon_n)_{n=0,\dots,N}$. En déduire la densité du vecteur aléatoire $(\varepsilon_0, (X_n)_{n=1,\dots,N})$ puis celle de $(X_n)_{n=1,\dots,N}$.
4. Soit Γ_n la matrice de variance covariance du vecteur $(X_n)_{n=1,\dots,N})$. En utilisant la question précédente, déterminer Γ_n^{-1}.
5. On observe $(X_n)_{n=1,\dots,N}$, déterminer très précisement la loi conditionnelle de X_{N+1} sachant cette observation. On désire estimer le paramètre a. Utiliser la covariance empirique $\widehat{r}_N(1) := (N-1)^{-1} \sum_{j=1}^{N-1} X_j X_{j-1}$ pour bâtir un estimateur empirique $\widehat{a}_N$ de a. Ecrire la vraisemblance associée aux observations $(X_n)_{n=1,\dots,N}$. L'estimateur du maximum de vraisemblance est-il calculable ?
6. Montrer que $\widehat{a}_N$ est un estimateur sans biais et qu'il converge en probabilité vers a.