Révisions

Séance du 29 Novembre 2012

1 Modèle linéaire

1.1 Cuatro Cuaranta

Soit n_1 et n_2 deux entiers strictement positifs fixés et $n = 2p > 2\max(n_1, n_2)$. On considère pour $j = 1 \dots n$, le modèle de régression périodique :

$$Y_j = a_0 + a_1 \cos(2\pi n_1 \frac{j}{n}) + a_2 \cos(2\pi n_2 \frac{j}{n}) + b_1 \sin(2\pi n_1 \frac{j}{n}) + b_2 \sin(2\pi n_2 \frac{j}{n}) + \varepsilon_j.$$

a) Montrer ou admettre pour $i, i' = 1, 2$, les relations suivantes :

$$\begin{aligned} \sum_{j=1}^{n} \cos(2\pi n_i \frac{j}{n}) \cos(2\pi n_{i'} \frac{j}{n}) &= 0 \quad \text{si } i \neq i' \\ &= p \quad \text{si } i = i' \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^{n} \sin(2\pi n_i \frac{j}{n}) \sin(2\pi n_{i'} \frac{j}{n}) &= 0 \quad \text{si } i \neq i' \\ &= p \quad \text{si } i = i' \end{aligned}$$

$$\sum_{j=1}^{n} \cos(2\pi n_i \frac{j}{n}) = \sum_{j=1}^{n} \sin(2\pi n_i \frac{j}{n}) = \sum_{j=1}^{n} \cos(2\pi n_i \frac{j}{n}) \sin(2\pi n_{i'} \frac{j}{n}) = 0.$$

b) Montrer que les estimateurs (du maximum de vraisemblance) des paramètres sont :

$$\begin{aligned} \widehat{a}_0 &= \frac{1}{n} \sum_{j=1}^{n} Y_j \\ \widehat{a}_i &= \frac{1}{p} \sum_{j=1}^{n} Y_j \cos(2\pi n_i \frac{j}{n}), \ i = 1, 2 \\ \widehat{b}_i &= \frac{1}{p} \sum_{j=1}^{n} Y_j \sin(2\pi n_i \frac{j}{n}), \ i = 1, 2 \\ S^2 &= \frac{1}{2p-5} \sum_{j=1}^{n} \left[Y_j - \left(\widehat{a}_0 + \widehat{a}_1 \cos(2\pi n_1 \frac{j}{n}) + \widehat{a}_2 \cos(2\pi n_2 \frac{j}{n}) + \widehat{b}_1 \sin(2\pi n_1 \frac{j}{n}) + \widehat{b}_2 \sin(2\pi n_2 \frac{j}{n}) \right) \right]^2 \end{aligned}$$

c) Construire le test d'hypthèse de $a_1 = a_2 = b_1 = b_2 = 0$, puis celui de $a_1 = b_1 = 0$.

d) Application numérique pour le second test de la question précédente. Cas de la note LA, $n_2 = 440$ $n = 1000$. Tester si la note est un LA pur (c'est-à-dire $a_1 = b_1 = 0$) avec $S^2_{\text{obs}} = 1,5$, $\widehat{a}_1^{\text{obs}} = 0,52$, $\widehat{b}_1^{\text{obs}} = 0,98$

1.2 Régression polynômiale

Soient p et n deux entiers tels que $0 \leq p < n$. Pour tout $1 \leq j \leq n$ on pose $x_j = j - \frac{n+1}{2}$. On considère $\psi_0, \psi_1, \ldots, \psi_p$ polynômes fixés tels que :

- degré$(\psi_i) = i, \ \forall 0 \leq i \leq p$
- $\sum_{j=1}^n \psi_i(x_j)\psi_k(x_j) = 0, \ \forall 0 \leq i \neq k \leq p$

On a en vue d'étudier le modèle de régression

$$Y_j = \sum_{i=0}^p \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq n$$

où $\varepsilon_1, \ldots, \varepsilon_n$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ et $\lambda_0, \ldots, \lambda_p, \sigma^2$ sont des paramètres inconnus.

a) Estimer les paramètres de ce modèle.

b) Ecrire un test de niveau α de "$\lambda_p = 0$" contre "$\lambda_p \neq 0$".

Le revenu net par action de la compagnie Gillette pour les années 57 à 64 est le suivant :

Année (z_j)	:	1957	1958	1959	1960	1961	1962	1963	1964
Revenu en \$ ($Y_j(\omega)$)	:	0,93	0,99	1,11	1,33	1,52	1,60	1,47	1,33

On pose $x_j = z_j - 1960,5$ puis $\psi_0(x) = 1, \psi_1(x) = 2x$ et $\psi_2(x) = x^2 - 21/4$.

c) Vérifier brièvement que le choix ci-dessus de $x_j, \psi_0, \psi_1, \psi_2$ rentre dans le cadre décrit plus haut.

d) En supposant que $Y_j = \sum_{i=0}^2 \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq 8$ avec $\varepsilon_1, \ldots, \varepsilon_8$ indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$, estimer $\lambda_0, \lambda_1, \lambda_2$ et σ^2. Tester si $\lambda_2 = 0$.

e) Faire une prévision pour le revenu net par action en 1965.

N.B. Pour faciliter les calculs, on indique les valeurs numériques suivantes :
$\sum_j \psi_1^2(x_j) = \sum_j \psi_2^2(x_j) = 168$ $\sum_j Y_j^2(\omega) = 13,65$ $\sum_j \psi_0(x_j)Y_j(\omega) = 10,28$
$\sum_j \psi_1(x_j)Y_j(\omega) = 6,86$ $\sum_j \psi_2(x_j)Y_j(\omega) = -4,1$.

1.3 Modèle linéaire et test d'hypothèses non linéaires

Soit $\varepsilon_{ij}, \quad i = 1, \ldots, k \ \ j = 1 \ldots n$, des variables i.i.d., de loi normale centrée réduite. Soit $m_1, \ldots, m_k \in \mathbb{R}$, on pose $Y_{ij} = m_i + \varepsilon_{ij} \ i = 1, \ldots, k \ \ j = 1 \ldots n$.

a) Quels sont les estimateurs du maximum de vraisemblance $\widehat{m}_i$ de m_i $(i = 1, \ldots, k)$?

b) On se propose maintenant de construire une statistique de test pour les hypothèses non linéaires :

$$H_0 \ \sum_{i=1}^k m_i^2 = 1 \quad H_1 \sum_{i=1}^k m_i^2 \neq 1.$$

Pour cela, on se propose d'utiliser la statistique de test

$$T_n = \sqrt{n}\left(\sum_{i=1}^k \widehat{m}_i^2 - 1\right).$$

Montrer que T_n peut s'écrire sous la forme suivante

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^k (\xi_i^n)^2 + 2\sum_{i=1}^k (\xi_i^n) m_i + \sqrt{n}(\sum_{i=1}^k m_i^2 - 1),$$

o $(\xi_i^n)_{i=1\ldots k}$ sont i.i.d. de loi $\mathcal{N}(0, 1)$.

c) On suppose que H_0 est vérifiée. Montrer que $\frac{1}{\sqrt{n}} \sum_{i=1}^k (\xi_i^n)^2$ converge en probabilité vers 0. En déduire que T_n converge en loi. Préciser la loi limite.

d) On suppose que H_1 est vérifiée. En utilisant la question précédente, montrer que presque sûrement $\lim_{n\to\infty} |T_n| = +\infty$.

e) Déduire des questions précédentes une procédure de test pour décider entre H_0 et H_1.

2 Théorèmes asymptotiques

2.1 Loi asymptotique des extrêmes

Pour $n \geq 1$, soit $X_1, \ldots, X_n$ des variables i.i.d. de densité f.

- On suppose que le support de f est $[0,1]$ et que cette fonction admet le développement :
$$f(x) = c_0 + c_1(1-x)^\alpha(1+\varepsilon(x)),$$
où $\alpha > -1$ et $\lim_{x\to 1^-} \varepsilon(x) = 0$. Discuter, en fonction de α de la normalisation stabilisant la loi de $X_{(n)}$ et déterminer la loi limite.
- Reprendre la question précédente dans le cas de la loi de Pareto de densité
$$f(x) = c_\alpha x^{-\alpha} 1_{\{x>1\}}.$$
On a ici $\alpha > 1$

2.2 Grandes déviations précises

On considère le modéle de régression
$$Y_i = a^* X_i + \varepsilon_i,\ i = 1, \cdots, N.$$
Où les vecteurs $(X_i, Y_i), i = 1, \ldots, n$ sont toutes i.i.d.. On suppose de plus que, pour tout $i = 1, \ldots, n$, X_i et ϵ_i sont des variables aléatoires indépendantes qui possèdent des moments d'ordre 4 finis.

1. Calculer $\widehat{a}$ l'estimateur des moindres carrès de a^*.
2. Montrer que, en général, $\widehat{a}$ est un estimateur biaisé.
3. Montrer que $\sqrt{n}(\widehat{a} - a^*)$ converge en loi vers une gaussienne centrée dont on précisera la variance.
4. On suppose que les variables (X_i) et (ε_i) sont des gaussiennes standard. Donne un équivalent, pour n grand et $t > 0$, de $\mathbb{P}(\widehat{a} - a^* \geq t)$ et de $\mathbb{P}(\widehat{a} - a^* \leq -t)$.
5. Reprendre la question précédente lorsque les (ϵ_i) sont des gaussiennes standard ou de loi de Laplace et que les variables (X_i) sont uniformément distribuées sur $\{-1, 1\}$ ou sur $\{-1, 0, 1\}$.

3 Processus gaussiens stationnaires

3.1 AR(1)

On considère le processus gaussien AR(1) causale
$$X_{n+1} = \theta^* X_n + \epsilon_{n+1}, (n \in \mathbb{Z}),$$
où $|\theta^*| < 1$ et (ϵ_n) est une suite i.i.d. de loi gaussienne standard. Calculer l'estimateurs de θ^* obtenu par la méthode du maximum de vraisemblance puis par la méthode de Yule Walker. Comparer leurs propriétés asymptotiques.

3.2 AR(2)

On considère le processus gaussien AR(2) causale
$$X_{n+1} = (\theta_1^* + \theta_2^*)X_n - (\theta_1^*\theta_2^*)X_{n-1} + \epsilon_{n+1}, (n \in \mathbb{Z}),$$
où $|\theta_i^*| < 1, (i = 1, 2)$ et (ϵ_n) est une suite i.i.d. de loi gaussienne standard. Exprimer X_n comme une fonction des (ϵ_i) de son passé. En déduire la fonction de covariance du processus et sa densité spectrale.