

Modèle linéaire

Séance du 17 Avril 2014

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 :

$$\hat{a}_0 = \frac{1}{n} \sum_{j=1}^n Y_j$$

$$\hat{a}_i = \frac{1}{p} \sum_{j=1}^n Y_j \cos(2\pi n_i \frac{j}{n}), \quad i = 1, 2$$

$$\hat{b}_i = \frac{1}{p} \sum_{j=1}^n Y_j \sin(2\pi n_i \frac{j}{n}), \quad i = 1, 2$$

$$S^2 = \frac{1}{2p-5} \sum_{j=1}^n \left[Y_j - \left(\hat{a}_0 + \hat{a}_1 \cos(2\pi n_1 \frac{j}{n}) + \hat{a}_2 \cos(2\pi n_2 \frac{j}{n}) + \hat{b}_1 \sin(2\pi n_1 \frac{j}{n}) + \hat{b}_2 \sin(2\pi n_2 \frac{j}{n}) \right) \right]^2$$

c) Construire le test d'hypothèse 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 $S_{\text{obs}}^2 = 1,5$, $\hat{a}_1^{\text{obs}} = 0,52$, $\hat{b}_1^{\text{obs}} = 0,98$

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 $\psi_0, \psi_1, \dots, \psi_p$ polynômes fixés tels que :

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

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

- Estimer les paramètres de ce modèle.
- 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 $\psi_2(x) = x^2 - 21/4$.

- Vérifier brièvement que le choix ci-dessus de $x_j, \psi_0, \psi_1, \psi_2$ rentre dans le cadre décrit plus haut.
- En supposant que $Y_j = \sum_{i=0}^2 \lambda_i \psi_i(x_j) + \varepsilon_j, 1 \leq j \leq 8$ avec $\varepsilon_1, \dots, \varepsilon_8$ indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$, estimer $\lambda_0, \lambda_1, \lambda_2$ et σ^2 . Tester si $\lambda_2 = 0$.
- Faire une prévision pour le revenu net par action en 1965.

N.B. Pour faciliter les calculs, on indique les valeurs numériques suivantes :

$$\begin{aligned} \sum_j \psi_1^2(x_j) &= \sum_j \psi_2^2(x_j) = 168 & \sum_j Y_j^2(\omega) &= 13,65 & \sum_j \psi_0(x_j) Y_j(\omega) &= 10,28 \\ \sum_j \psi_1(x_j) Y_j(\omega) &= 6,86 & \sum_j \psi_2(x_j) Y_j(\omega) &= -4,1. \end{aligned}$$

3 Modèle linéaire et test d'hypothèses non linéaires

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

- Quels sont les estimateurs du maximum de vraisemblance \hat{m}_i de $m_i (i = 1, \dots, k)$?
- 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

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

Montrer que T_n peut s'écrire sous la forme suivante

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

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

- 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.
- On suppose que H_1 est vérifiée. En utilisant la question précédente, montrer que presque sûrement $\lim_{n \rightarrow \infty} |T_n| = +\infty$.
- Déduire des questions précédentes une procédure de test pour décider entre H_0 et H_1 .