Faculté des sciences Année universitaire 2020/2021

Département de mathématiques

Master 1: Statistique et Probabilités Approfondies

Module: Régression

Epreuve finale

Exercice 1: Considérons un modèle de régression multiple

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \varepsilon_i \;\; ; i = 1, \dots n$$

On suppose que les (ε_i) sont centrées non corrélée et de même variance $\sigma^2 = 10$.

Soit $\widehat{\beta_n}$ l'estimateur des moindres carrées de β, $\widehat{Y}$ le vecteur des prédictions de Y et $\hat{\varepsilon}$ =Y-$\widehat{Y}$ le vecteur des erreurs de la prévision.

1) On note $e = (1,1, \dots ,1)'$.
   Calculer $\prec \hat{\varepsilon} \, , e \succ$. Que peut-on déduire ?
2) Simplifier $\| X\widehat{\beta_n} \|^2 + \| \hat{\varepsilon} \|^2$.
3) On se donne

$$X^t X = \begin{pmatrix} 25 & 0 & 0 \\ ? & 9{,}3 & 5{,}4 \\ ? & ? & 12{,}7 \end{pmatrix}, \quad (X^t X)^{-1} = \begin{pmatrix} 0{,}04 & 0 & 0 \\ 0 & 0{,}1428 & -0{,}0607 \\ 0 & -0{,}0607 & 0{,}1046 \end{pmatrix}$$

   A) Donner les valeurs manquantes. Déterminer les valeurs de n et k.
   B) Calculer la matrice de covariance de $\widehat{\beta_n}$.

Exercice 2: Soit $(\boldsymbol{\varepsilon_t})_{t\in\mathbb{Z}}$ un bruit blanc fort.
Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stochastique défini par :

$$X_t = -\tfrac{1}{4} X_{t-1} + \tfrac{1}{8} X_{t-2} + \varepsilon_t \text{ , t} \in \mathbb{Z}$$

1) Soit M_4 la moyenne mobile arithmétique d'ordre 4. Ecrire $M_4(X_t)$.
2) Ecrire X_t en utilisant les opérateurs retard et avance.
3) $(X_t)_{t\in\mathbb{Z}}$ est -il faiblement stationnaire ?
4) Calculer $E(X_t)$.
5) Calculer $V(X_t)$.

Bon courage

Université Abou Bakr Belkaid - Tlemcen

Faculté des Sciences
Département de Mathématiques
Année universitaire : 2020-2021.

Master I : Probabilités-Statistiques
Module : Régression
Durée : 1h30

Corrigé de l'épreuve finale

Exercice 1. Le modèle est :

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ... + \beta_k x_{k,i} + \varepsilon_i, \quad i = 1, \ldots, n.$$

L'écriture matricielle de ce modèle est :

$$Y = X\beta + \varepsilon$$

où

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & \ldots & x_{k,1} & \\ 1 & x_{1,1} & x_{2,1} & \ldots & \ldots & x_{n,1} & \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \ldots & \ldots & x_{k,n} & \end{pmatrix}; \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_k . \end{pmatrix}$$

1) Soit $\mathcal{E}(X)$ l'espace engendré par les vecteurs colonnes de X. Comme $\hat{y} = X\hat{\beta_n}$ est la projection orthogonale de Y sur $\mathcal{E}(X)$ et $e = (1, 1, \ldots, 1)^t \in \mathcal{E}(X)$ alors

$$\langle \hat{\varepsilon}, e \rangle = \left\langle \hat{Y} - Y, e \right\rangle = 0.$$

02 points

Donc

$$\langle Y, e \rangle = \left\langle \hat{Y}, e \right\rangle$$

Par suite

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y_i}$$

En divisant par n, On en déduit que $\bar{Y_n} = \bar{\hat{Y_n}}$.

0,5 point

2) Puisque $\hat{Y} = X\hat{\beta_n}$ alors

$$X\hat{\beta_n} + \hat{\varepsilon} = X\hat{\beta_n} + Y - \hat{Y} = Y.$$

Donc

$$\left\| X\hat{\beta_n} + \hat{\varepsilon} \right\|^2 = \|Y\|^2,$$

d'autre part

$$\hat{Y} = X\hat{\beta_n} = P_{\mathcal{E}(X)}(Y).$$

Ainsi

$$\begin{aligned}\left\|X\hat{\beta_n}\right\|^2 + \|\hat{\varepsilon}\|^2 &= \left\|X\hat{\beta_n} + \hat{\varepsilon}\right\|^2 \\ &= \|Y\|^2 .\end{aligned}$$

02,5 points

3) A) La matrice X^tX est symétrique donc

$$X = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 9,3 & 5,4 \\ 0 & 5,4 & 12,7 \end{pmatrix}$$

01,5 point

La première colonne de la matrice X est le vecteur e qui a n lignes. Si on écrit $X^tX = (a_{i,j})$ alors

$$n = a_{1,1} = 25.$$

02 points

Le nombre de colonnes de la matrice X est $k+1$, donc le nombre de lignes de la matrice X^tX^{-1} est $k+1$. Par conséquent $k+1 = 3$, soit $k = 2$.

01 point

B) $C_{\hat{B_n}} = \sigma^2 X^tX^{-1}$. On obtient

$$C_{\hat{B_n}} = \begin{pmatrix} 0,4 & 0 & 0 \\ 0 & 1,428 & -0,607 \\ 0 & -0,607 & 1,046 \end{pmatrix}$$

01,5 point

Exercice 2. $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc fort de variance σ^2 et

$$X_t = -\frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2} + \varepsilon_t, \quad t \in \mathbb{Z}. \tag{1}$$

1) Comme 4 est pair alors

$$M_4(X_t) = \frac{1}{4}(\frac{1}{2}X_{t-2} + X_{t-1} + X_t + X_{t+1} + \frac{1}{2}X_{t+2}), \quad t \in \mathbb{Z}.$$

01,5 point

2) Soit B l'opérateur retard. On a

$$X_t = -\frac{1}{4}BX_t + \frac{1}{8}B^2X_{t-2} + \varepsilon_t = (-\frac{1}{4}B + \frac{1}{8}B^2)X_t + \varepsilon_t \;\; t \in \mathbb{Z}.$$

01 point

3) Le modèle $(X_t)_{t\in\mathbb{Z}}$ est un autorégressif d'ordre 2. Considérons le polynôme caractéristique

$$P(z) = z^2 + \frac{1}{4}z - \frac{1}{8}.$$

Les racines de P sont $z_1 = -\frac{1}{2}$ et $z_2 = \frac{1}{4}$. Puisque $|z_1| < 1$ et $|z_2| < 1$ alors le processus $(X_t)_{t\in\mathbb{Z}}$ est faiblement stationnaire.

02 points

4) Puisque $(X_t)_{t\in\mathbb{Z}}$ est faiblement stationnaire alors $E(X_t) = E(X_{t-1}) = E(X_{t-2})$. En remplaçant dans l'équation , on obtient $E(X_t) = 0$.

01,5 point

5) En utilisant La relation 1 on peut écrire

$$X_t^2 = \frac{1}{16}X_{t-1}^2 + \frac{1}{64}X_{t-2}^2 + \varepsilon_t^2 - \frac{1}{16}X_{t-1}X_{t-2} - \frac{1}{2}\varepsilon_t X_{t-1} + \frac{1}{4}\varepsilon_t X_{t-2}$$

Les conditions faites sur le modèle donnent

$$E(\varepsilon_t X_{t-1}) = E(\varepsilon_t X_{t-2}) = 0.$$

D'autre part, puisque le processus est faiblement stationnaire alors $E(X_{t-1}X_{t-2}) = \gamma(1)$ où γ est la fonction d'autocovariance. Ainsi

$$E(X_t^2) = \frac{1}{16}E(X_t^2) + \frac{1}{64}E(X_t^2) + \sigma^2 - \frac{1}{16}\gamma(1) \tag{2}$$

Par la relation 1, on a aussi

$$X_t X_{t-1} = -\frac{1}{4}X_{t-1}^2 + \frac{1}{8}X_{t-2}X_{t-1} + \varepsilon_t X_{t-1}$$

En prenant les espérances, on obtient

$$\gamma(1) = -\frac{1}{4}\gamma(0) + \frac{1}{8}\gamma(0),$$

il en découle que

$$\frac{\gamma(1)}{\gamma(0)} = -\frac{1}{4} + \frac{1}{8}\frac{\gamma(1)}{\gamma(0)},$$

d'où

$$\rho(1) = -\frac{1}{4} + \frac{1}{8}\rho(1),$$

où ρ est la fonction d'autocorrélation. On tire que

$$\rho(1) = -\frac{2}{7}$$

Or

$$\gamma(1) = \rho(1)\gamma(0) = \rho(1)Var(X_t).$$

En replaçant dans l'équation 2, on obtient

$$Var(X_t) = \frac{1}{16}Var(X_t) + \frac{1}{64}Var(X_t) + \sigma^2 - \frac{1}{16}\rho(1)Var(X_t),$$

donc

$$(1 - \frac{1}{16} - \frac{1}{64} - \frac{1}{16}\frac{2}{7})Var(X_t) = \sigma^2,$$

soit

$$Var(X_t) = \frac{448}{405}\sigma^2$$

03 points