

Remerciements

Je remercie Dieu tout puissant de m'avoir donné les forces qui m'ont permis de dépasser toutes les difficultés pour aboutir enfin à la réalisation de ce modeste travail.

Je remercie Monsieur A.LABBAS qui m'a dirigé dans cette étude et m'a confié le sujet de cette thèse.

Je remercie Monsieur T.MOURID pour son acceptation de présider le jury de thèse.

Ainsi que Monsieur F.BOUKHARI et Monsieur G.SENOUCI BREKSI et Monsieur D.BENMANSOUR d'avoir accepté d'examiner cette thèse. .

Mes vifs remerciements vont également à mes camarades qui m'ont fait profiter de leurs diverses expériences.

Ma reconnaissance va, enfin, à tous mes professeurs qui ont contribué à ma formation mathématique.

Table des Matières

Introduction	4
1 Généralités sur les Processus MA(q) - AR(p) et ARMA(p,q)	6
1.1 Processus Stationnaires	7
1.1.1 Fonction d'autocovariance et fonction d'autocorrelation	8
1.2 Processus linéaires	10
1.2.1 Processus ARMA (p,q)	12
1.2.2 Processus AR(p)	14
1.2.3 Processus MA(q)	20
2 Estimateurs des Moindres Carrés Et du Maximum de Vraisemblance	25
2.1 Estimateurs des moindres carrés	25
2.1.1 Théorème de Gauss-Markov	29
2.1.2 Les moindres carrés généralisé [GLS]	30
2.2 Méthode du Maximum de Vraisemblance	31
2.2.1 Théorème	35
2.2.2 Fonction de Log-vraisemblance Concentrée :	39
3 Estimateur de Maximum de Vraisemblance dans une Moyenne mobile :Cas	
Scalaire	40
3.1 Introduction	40
3.2 Fonction de vraisemblance	45
3.2.1 Evaluation de la fonction de vraisemblance $\ell^*(\theta)$	60

3.3	Difficultés de l'évaluation directe	60
3.4	Inversibilité	62
3.5	Inférence	65
3.6	Exemple	66
3.7	Conclusion	69
4	Estimateurs Exacte et Conditionnel du Maximum de Vraisemblance d'un	
	MA(q) Vectoriel	70
4.1	Introduction	70
4.2	Fonction de Vraisemblance Exacte	70
4.3	Approximation de la fonction de vraisemblance	78
4.4	Evaluation de la Vraisemblance	82
4.5	Conclusion	84
	Annexe	86
	Annexe	93
	Conclusion	97

Introduction

Les processus moyenne mobile (*MA* pour moving average) abordés dans ce mémoire trouvent leurs applications dans plusieurs domaines, économie, biologie, médecine ect...

Un processus (X_t) est une moyenne mobile d'ordre q $MA(q)$ si il est défini par

$$X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (t \in Z)$$

où (ε_t) est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne zéro et de variance σ^2 (bruit blanc) et θ_i sont des constantes en module inférieures à un.

Parmi les problèmes abordés dans cette classe de processus nous relevons l'importance de l'estimation des paramètres θ_i définissant le modèle $MA(q)$.

Dans ce mémoire nous développons les résultats traitants l'estimation des paramètres dans un modèle MA par la méthode suivante :

***Méthode du Maximum de Vraisemblance ML (Maximum Likelihood).**

Nous développons les résultats des articles suivants:

1_ Maximum likelihood estimation of moving average processes.

.DeniseR.Osborn(1976).Annals of Economic and Social Measurement 5/1,1976.

2_ Exact and approximate Maximum likelihood estimators for vector moving average processes .

DeniseR.Osborn(1976).National Institute of Economic and Social research ,london.

3_ The first-order moving average process.Charles R.Nelson(1973).Graduats

School of Business University of Chicago, III, 60637, USA.

Dans le chapitre 1, nous rappelons les modèles et quelques propriétés des processus $MA(q)$, $AR(p)$ et $ARMA(p, q)$.

Dans le chapitre 2, nous considérons deux classes d'estimateurs (estimateurs des moindres carrés et Maximum de vraisemblance) pour des modèles statistiques généraux.

Dans le chapitre 3, nous appliquons la méthode d'estimation du Maximum de vraisemblance ML (Maximum likelihood) pour une moyenne mobile, dans le cas scalaire. Elle est basée sur une maximisation de la fonction de vraisemblance. Dans une première étape on suppose que les valeurs initiales sont des nombres fixés. Dans la deuxième étape on suppose que les valeurs initiales sont des variables aléatoires, avec utilisation d'estimateurs des moindres carrés (LS).

Dans le chapitre 4, nous donnons la fonction exacte de vraisemblance pour les processus Moyenne mobile vectoriel. La méthode utilisée est une généralisation de celle du chapitre précédent.

En Annexe, nous présentons des résultats de simulations numériques sur le comportement de l'estimateur de Maximum de vraisemblance pour un processus moyenne mobile d'ordre un $MA(1)$ en utilisant le logiciel R .

Chapitre 1

Généralités sur les Processus $MA(q)$ - $AR(p)$ et $ARMA(p,q)$

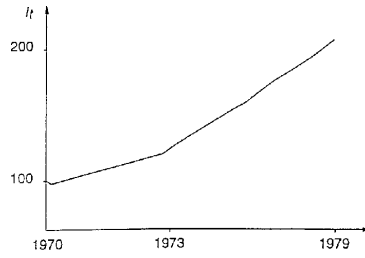
Dans ce chapitre nous présentons des définitions et propriétés des processus MA , AR et $ARMA$ qui seront utilisées par la suite.

Séries chronologiques: définitions et exemples :

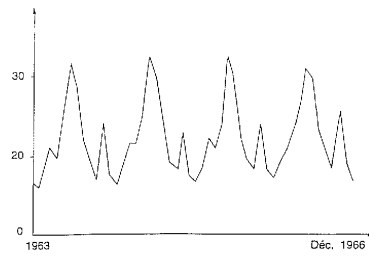
La théorie des séries temporelles (ou chronologiques) est appliquée de nos jours dans des domaines aussi divers que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie.

Une série chronologique est une suite de données formée d'observations au cours du temps: l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à des phénomènes naturels (comme le nombre de taches solaires).

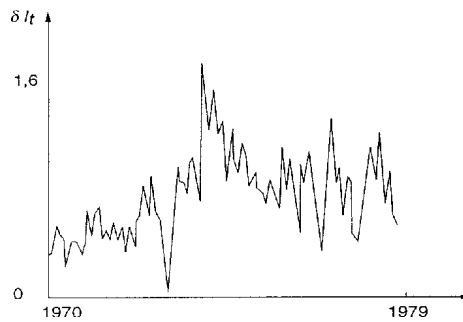
La définition mathématique d'une série chronologique est la donnée d'une famille de variables aléatoires $(X(t), t \in T)$ définie sur un espace de probabilité, où T est un intervalle de temps qui est discret (dans ce cas $T = \{1, 2, 3, \dots, n\}$).



(a) Indice mensuel des prix à la consommation I_t .



(b) Trafic voyageur de la SNCF en 2ième classe



(c) Accroissement relatif mensuel de l'indice des prix

Exemples de séries chronologiques

1.1 Processus Stationnaires

Définition 1.1

le processus $(X(t), t \in T)$ est dit strictement ou fortement stationnaire si pour toute partie définie $(t_1, t_2, \dots, t_n) \in T$ et pour tout $s > 0$ on a :

$$\mathcal{L}(X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s}) = \mathcal{L}(X_{t_1}, X_{t_2}, \dots, X_{t_n})$$

où $\mathcal{L}(Y)$ désigne loi de probabilité de Y .

Un processus $(X(t), t \in T)$ est dit faiblement stationnaire (ou stationnaire) si

i) $E(X_t) = m < \infty, \forall t$ (indépendant de t)

- ii) $E(X_t^2) < \infty \quad \forall t$
- iii) $cov(X_{s+h}, X_{t+h}) = cov(X_s, X_t), \quad \forall t, s, h$

Lemme 1.1

*Si (X_t) est fortement stationnaire et $E(X_t^2) < \infty$, alors X_t est faiblement stationnaire.
La réciproque est fautive en général.*

Preuve:

Si (X_t) est fortement stationnaire alors

$$\mathcal{L}(X_t) = \mathcal{L}(X_{t'}) \text{ et donc } E(X_t) = E(X_{t'})$$

D'autre part, on a aussi $\forall t, s, h \in T$

$$\mathcal{L}(X_t, X_s) = \mathcal{L}(X_{t+h}, X_{s+h}) \text{ .Donc } cov(X_t, X_s) = cov(X_{t+h}, X_{s+h})$$

Donc X_t est faiblement stationnaire .

■

1.1.1 Fonction d'autocovariance et fonction d'autocorrelation

Pour un processus (X_t) faiblement stationnaire on definit:

i) la fonction

$$\gamma_X(h) = cov(X_t, X_{t+h}) \quad h \in \mathbb{Z}$$

dite fonction d'autocovariance de processus (ACVF).

ii) la fonction

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

est dite fonction d'autocorrelation (ACF)

Propriétés:

- 1) $\gamma_X(0) = \sigma_X^2$ et $|\gamma_X(h)| \leq \gamma_X(0)$

2) $\gamma_X(h) = \gamma_X(-h)$

3) $\forall a_1, a_2, \dots, a_n \in \mathbb{R} \quad , \quad t_1, t_2, \dots, t_n \in \mathbb{R}$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_X(t_i - t_j) \geq 0$$

Preuve:

Supposons que $E(X_t) = 0$. On a :

1) $\gamma_X(0) = \text{var}(X_t) = \sigma_X^2 \geq 0$ et par l'inegalité de Cauchy-Schwartz :

$$|\gamma_X(h)| = |E(X_t X_{t+h})| \leq (E(X_t^2))^{1/2} (E(X_{t+h}^2))^{1/2} = E(X_t^2) = \gamma_X(0)$$

2) $\gamma_X(h) = \text{cov}(X_t, X_{t+h})$,on pose $t = t_0 - h$

$$\gamma_X(h) = \text{cov}(X_{t_0-h}, X_{t_0}) = \text{cov}(X_{t_0}, X_{t_0-h}) = \gamma_X(-h)$$

3) $0 \leq \text{var}\left(\sum_{j=1}^n a_j X_{t_j}\right) = E\left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j X_{t_i} X_{t_j}\right)$
 $= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_X(t_i - t_j).$

Dans la classe des processus stationnaires il existe des processus particuliers qui sont les processus bruit blanc (White noise).

Définition 1.2

Une suite de variable aléatoires (ε_t) constitue un bruit blanc faible (resp fort) si :

$$E(\varepsilon_t) = 0 \quad \forall t \in \mathbb{Z}$$

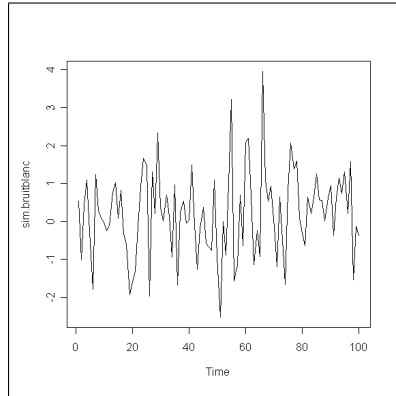
$$E(\varepsilon_t^2) = \sigma^2 \text{ est constante strictement positive}$$

$$\text{cov}(\varepsilon_t, \varepsilon_s) = 0 \text{ si } t \neq s \text{ (resp } \varepsilon_t \text{ sont iid)}$$

Notation :

Si (ε_t) un bruit blanc faible gaussien on notera $(\varepsilon_t) \hookrightarrow WN(0, \sigma^2)$

Si (ε_t) est bruit blanc fort gaussien $(\varepsilon_t) \hookrightarrow IID(0, \sigma^2)$



Réalisation d'un bruit blanc (ε_t)

avec $E(\varepsilon_t) = 0$ et $var(\varepsilon_t) = \sigma^2$

1.2 Processus linéaires

Deux exemples importants dans la classe de processus linéaires sont les processus autorégressifs (AR) et les processus moyennes mobiles (moving average MA).

Les processus ARMA sont des généralisations de MA et AR et sont très importants dans la modélisation des séries chronologiques.

Définition 1.3

Une série temporelle $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire si elle admet une représentation

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} \text{ pour tout } t, (\varepsilon_t) \hookrightarrow WN(0, \sigma^2) \text{ et } \sum_{j=0}^{\infty} |\Psi_j| < \infty$$

On définit B l'opérateur du retard sur une suite (Y_t) par:

$$BY_t = Y_{t-1} \quad \text{et} \quad B^j Y_t = Y_{t-j}$$

Le processus linéaire $(X_t)_{t \in \mathbb{Z}}$ peut s'écrire

$$X_t = \Psi(B)\varepsilon_t \quad \text{avec} \quad \Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j$$

Proposition 1.1

Soit (X_t) un processus stationnaire centré et de fonction d'autocovariance $\gamma_X(\cdot)$ et tel que

$$\sum_{j=0}^{\infty} |\Psi_j| < \infty$$

Alors

$$\Psi(B)X_t = \sum_{j=0}^{\infty} \Psi_j B^j X_t = \sum_{j=0}^{\infty} \Psi_j X_{t-j}$$

où la série converge absolument (en prob) et en moyenne quadratique .

Si $Y_t = \Psi(B)X_t$ Alors le processus (Y_t) est stationnaire de centré et de fonction d'auto-covariante

$$\gamma_Y(h) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \Psi_j \Psi_k \gamma_x(h - j + k)$$

Preuve:

Des conditions sur le processus (X_t) on a:

$$E(X_t) = 0 \quad , \quad \text{var}(X_t) = \gamma_x(0) = \sigma^2 = E(X_t^2).$$

D'autre part,

$$E\left(\sum_{j=0}^{\infty} |\Psi_j| |X_{t-j}|\right) = E\left(\lim_{n \rightarrow \infty} \sum_{j=0}^n |\Psi_j| |X_{t-j}|\right) \dots \dots (1)$$

D'après le théorème de convergence monotone

$$(1) = \lim_{n \rightarrow \infty} \sum_{j=0}^n |\Psi_j| E(|X_{t-j}|) < \infty$$

* On va appliquer le critère de Cauchy

$$E \left| \sum_{j=0}^p \Psi_j X_{t-j} - \sum_{j=0}^q \Psi_j X_{t-j} \right|^2 = E \left| \sum_{j=p}^q \Psi_j X_{t-j} \right|^2 = \sum_{j=p}^q \sum_{k=p}^q \Psi_j \Psi_k E(X_{t-j} X_{t-k}) \dots (2)$$

car

$$\left(\sum_{i=0}^n a_j\right)^2 = \sum_{j=0}^n \sum_{i=0}^n a_j a_i$$

$$(2) \leq E(X_t^2) \underbrace{\left(\sum_{j=p}^q |\Psi_j|\right)^2}_{\substack{\rightarrow 0 \\ p, q \rightarrow \infty}}$$

car $\sum_{j=p}^q |\Psi_j| \prec \infty$.

$\Psi(B)Y_t$ converge en moyenne quadratique.

Si S désigne la limite de moyenne quadratique donc d'après le Lemme de FATTOU

$$E|S - \Psi(B)X_t|^2 = E \liminf_{n \rightarrow \infty} \left| S - \sum_{j=0}^n \Psi_j X_{t-j} \right|^2 \leq \liminf_{n \rightarrow \infty} E \left| S - \sum_{j=0}^n \Psi_j X_{t-j} \right|^2 = 0$$

On conclut que la limite S et $\Psi(B)X_t$ sont égales en probabilité

d'où la 1ere condition de la stationnarité ($E(Y_t^2) \prec \infty$) est vérifiée.

$$\text{ii) } E(Y_t) = E\left(\sum_{j=0}^{\infty} \Psi_j X_{t-j}\right) = \sum_{j=0}^{\infty} \Psi_j E(X_{t-j}) = 0$$

$$\begin{aligned} \text{iii) } E(Y_{t+h}Y_t) &= E \left[\left(\sum_{j=0}^{\infty} \Psi_j X_{t+h-j}\right) \left(\sum_{k=0}^{\infty} \Psi_k X_{t-k}\right) \right] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Psi_j \Psi_k E(X_{t+h-j} X_{t-k}) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Psi_j \Psi_k \gamma_x(h-j-k) = \gamma_Y(h) \quad \blacksquare \end{aligned}$$

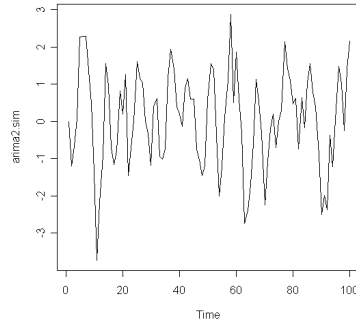
1.2.1 Processus ARMA (p,q)

Définition 1.4

Un processus linéaire stationnaire (X_t) centré est appelée ARMA(p,q) $p \geq 0, q \geq 0$ s'il existe des constantes $\alpha_1, \dots, \alpha_p$ ($\alpha_p \neq 0$) et $\theta_1, \dots, \theta_q$ ($\theta_q \neq 0$) et un processus (ε_t) un bruit blanc (faible)

tel que :

$$X_t - \sum_{k=1}^p \alpha_k X_{t-k} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$



Simulation d'un ARMA(2,2)

avec $\theta_1 = -0.22$, $\theta_2 = 0.24$
 $\alpha_1 = 0.88$, $\alpha_2 = -0.48$

Le processus (ε_t) est appelé processus des innovations.

Pour $p = q = 0$ on a

$$X_t = \varepsilon_t$$

Un processus ARMA(0,0) est un bruit blanc faible

Deux classes de processus ARMA que nous étudierons dans un premier temps sont ceux pour les quels un des deux paramètres p ou q est nul.

Si $p = 0$ et $q > 0$, un processus ARMA(0, q) noté MA(q) est de la forme

$$X_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

où $\varepsilon_t \hookrightarrow N(0, \sigma^2)$.

Ce processus est appelé processus moyenne mobile d'ordre q .

Si $p > 0$ et $q = 0$, un processus ARMA($p, 0$), noté AR(p), est de la forme

$$X_t = \varepsilon_t + \sum_{k=1}^p \alpha_k X_{t-k}$$

où $\varepsilon_t \hookrightarrow N(0, \sigma^2)$.

Ce processus est appelé processus autorégressif d'ordre p

1.2.2 Processus AR(p)

Les processus $AR(p)$ forment une classe de modèles pour de nombreux phénomènes observés. Ils sont définis implicitement par la relation :

$$X_t = \varepsilon_t + \sum_{k=1}^p \alpha_k X_{t-k}$$

Cette définition pose quelques problèmes :

Le processus est-t-il stationnaire ? Quelle est sa forme explicite ?

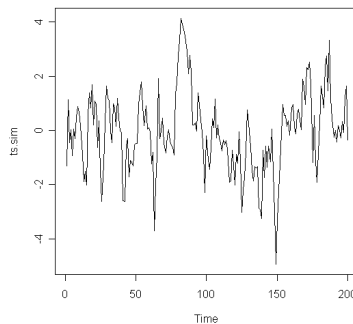
Soit B l'opérateur du retard, $BX_t = X_{t-1}$, $\forall t \in \mathbb{Z}$

$B^0 X_t = X_t$ et $B^j X_t = X_{t-j}$

Nous avons

$$X_t - \sum_{k=1}^p \alpha_k X_{t-k} = \varepsilon_t \Leftrightarrow (1 - \alpha_1 B - \dots - \alpha_p B^p) X_t = \varepsilon_t \Leftrightarrow \alpha(B) X_t = \varepsilon_t$$

où $\alpha(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)$



Simulation d'un AR(1)

avec $\alpha = 0.7$

Stationnarité et Inversibilité

Un processus (X_t) satisfaisant une représentation AR est toujours inversible par définition, ce résultat indique que tout processus AR peut être inversé .

On va étudier la stationnarité et donné des conditions sur les paramètres α_k .

***Cas AR(1) :**

Proposition 1.2 soit (X_t) un processus autorégressif d'ordre 1

$$X_t = \alpha X_{t-1} + \varepsilon_t \quad (*)$$

avec (ε_t) un bruit blanc fort gaussien , $E(X_s \varepsilon_t) = 0$.

Si

$$|\alpha| < 1$$

Alors il existe une solution unique stationnaire de l'équation (*) donnée par :

$$X_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

Preuve:

Existance :

Soit B l'opérateur du retard

$$(*) \iff X_t - \alpha X_{t-1} = \varepsilon_t \iff (1 - \alpha B)X_t = \varepsilon_t .$$

Si $|\alpha| < 1$ l'opérateur $(I - \alpha B)$ est inversible d'inverse $\sum_j \alpha^j B^j$:

$$(1 - \alpha B) \lim_{j \rightarrow \infty} (1 + \alpha B + \dots + \alpha^j B^j) = \lim_{j \rightarrow \infty} (1 + \alpha B + \dots + \alpha^j B^j - \alpha B - \alpha^j B^j - \alpha^{j+1} B^{j+1})$$

$$\begin{aligned} &= \lim_{j \rightarrow \infty} (1 - \alpha^{j+1} B^{j+1}) \\ &= I \quad \text{car} \quad \lim_{j \rightarrow \infty} \alpha^j = 0 \quad \text{quand} \quad |\alpha| < 1 \end{aligned}$$

$$\text{Donc } X_t = (1 - \alpha B)^{-1} \varepsilon_t \quad (**)$$

$$(1 - \alpha B)^{-1} \varepsilon_t = \lim_{j \rightarrow \infty} (1 + \alpha B + \dots + \alpha^j B^j) \varepsilon_t$$

$$(**) \iff X_t = \sum_{j=0}^{\infty} \alpha^j B^j \varepsilon_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

D'après la proposition 1.1

on a (ε_t) un processus stationnaire et $\sum_{j=0}^{\infty} |\alpha^j| < \infty$ c'est une série géométrique ($|\alpha| < 1$)

Finalement

$$X_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

est stationnaire et de fonction d'autocovariance

$$\gamma_X(h) = \sum_{j=0}^{\infty} \alpha^j \alpha^{j+h} \sigma^2 = \frac{\sigma^2 \alpha^h}{1 - \alpha^2}$$

Unicité : Supposons qu'il existe un autre processus (Y_t) tq :

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + \varepsilon_t = \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha^2 Y_{t-2} \\ &= \varepsilon_t + \alpha \varepsilon_{t-1} + \dots + \alpha^k \varepsilon_{t-k} + \alpha^{k+1} Y_{t-k-1} \end{aligned}$$

Or (Y_t) est stationnaire, alors $E(Y_t^2)$ est finie et indépendant de t

$$E\left(Y_t - \sum_{j=0}^k \alpha^j \varepsilon_{t-j}\right)^2 = \alpha^{2(k+1)} E(Y_{t-k-1}^2) \rightarrow 0 \text{ quand } k \rightarrow \infty$$

$$\begin{aligned} \text{Donc} \quad Y_t &= \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \\ \text{et alors} \quad Y_t &= X_t \text{ p.s } \blacksquare \end{aligned}$$

Proposition 1.3

(X_t) est un processus autorégressif d'ordre p stationnaire si et seulement si les racines du polynôme

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^k$$

sont en module supérieur strictement à 1.

Preuve:

*cas $AR(1)$:

$$X_t = \alpha X_{t-1} + \varepsilon_t \Leftrightarrow (1 - \alpha B)X_t = \varepsilon_t$$

$Mq (X_t)$ est stationnaire si et seulement si la racine de $(1 - \alpha z)$ supérieure stricte à 1.

La racine de

$$A(z) = 1 - \alpha z \text{ est } z = \frac{1}{\alpha}$$

c.a.d (X_t) est stationnaire si et seulement si $|\alpha| < 1$

D'après la proposition 1.1 on a la 1ère implication, il nous reste la 2ème implication c.a.d: si (X_t) est stationnaire alors $|\alpha| < 1$

$$\text{On a:} \quad X_t = \alpha X_{t-1} + \varepsilon_t \Leftrightarrow X_t^2 = \alpha^2 X_{t-1}^2 + \varepsilon_t^2 + 2\alpha X_{t-1}\varepsilon_t$$

$$\text{Donc :} \quad E(X_t^2) = \alpha^2 E(X_{t-1}^2) + E(\varepsilon_t^2) \Leftrightarrow (1 - \alpha^2)\gamma_X(0) = \sigma^2$$

$$\Rightarrow (1 - \alpha^2) > 0 \Rightarrow \alpha^2 < 1 \Rightarrow |\alpha| < 1$$

*cas $AR(p)$:

$$\text{Soit} \quad X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t \Leftrightarrow \alpha(B)X_t = \varepsilon_t$$

$$\text{où} \quad \alpha(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p$$

$$\text{on écrit} \quad X_t = \alpha^{-1}(B)\varepsilon_t \text{ si } \alpha(B) \text{ est inversible}$$

Considérons $\alpha(z)$ tel que:

$$\alpha(z) = (1 - \lambda_1 z)(1 - \lambda_2 z)\dots(1 - \lambda_p z) = \prod_{i=1}^p (1 - \lambda_i z)$$

$$\alpha^{-1}(z) = \frac{1}{\alpha(z)} = \sum_{i=1}^p \frac{K_i}{(1 - \lambda_i z)}$$

$$\text{donc } X_t = \alpha^{-1}(B)\varepsilon_t = \sum_{i=1}^p K_i (1 - \lambda_i B)^{-1} \varepsilon_t$$

comme dans le cas de $AR(1)$ il faut et il suffit que $|\lambda_i| < 1$ pour la stationnarité.

Et $\frac{1}{\lambda_i}$ sont des racines de

$$A(z) = (1 - \lambda_1 z)(1 - \lambda_2 z)\dots(1 - \lambda_p z) = 1 - \sum_{k=1}^p \alpha_k z^k$$

Ce qui nous montre que

$$\left| \frac{1}{\lambda_i} \right| > 1.$$

■

Fonction d'autocorrelation : les équations de Yule-Walker

$$\text{De } X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t$$

on a, pour $h > 0$:

$$X_{t-h} X_t = \alpha_1 X_{t-h} X_{t-1} + \dots + \alpha_p X_{t-h} X_{t-p} + X_{t-h} \varepsilon_t$$

$$\text{Donc } E(X_{t-h} X_t) = \alpha_1 E(X_{t-h} X_{t-1}) + \dots + \alpha_p E(X_{t-h} X_{t-p}) + E(X_{t-h} \varepsilon_t)$$

$$\Leftrightarrow \gamma_h = \alpha_1 \gamma_{h-1} + \alpha_2 \gamma_{h-2} + \dots + \alpha_p \gamma_{h-p}$$

$$\text{avec } E(X_{t-h} \varepsilon_t) = 0 \text{ et } \rho_h = \frac{\gamma_h}{\gamma_0}$$

où ρ_h : la fonction d'autocorrection

γ_h : la fonction d'auto covariance.

Ainsi

$$\rho_h = \alpha_1 \rho_{h-1} + \alpha_2 \rho_{h-2} + \dots + \alpha_p \rho_{h-p} \text{ et } \gamma_0 = E(X_t X_t)$$

De même:

$$X_t X_t = \alpha_1 X_t X_{t-1} + \dots + \alpha_p X_t X_{t-p} + X_t \varepsilon_t$$

Donc

$$E(X_t X_t) = \gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \dots + \alpha_p \gamma_p$$

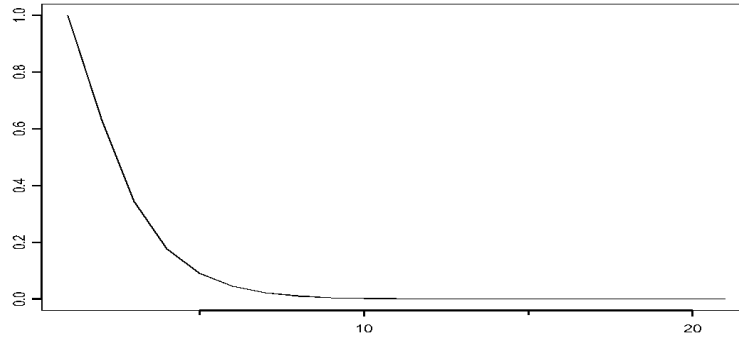
Les Equations de Yule-Walker :

$$\begin{aligned} \text{si} \quad h &= 1, 2, \dots, p \\ \rho_1 &= \alpha_1 + \alpha_2 \rho_1 + \dots + \alpha_p \rho_{p-1} \\ \rho_2 &= \alpha_1 \rho_1 + \alpha_2 + \dots + \alpha_p \rho_{p-2} \\ \\ \rho_h &= \alpha_1 \rho_{p-1} + \alpha_2 \rho_{p-2} + \dots + \alpha_p \end{aligned}$$

Ce système peut s'écrire sous forme matricielle

$$\begin{aligned} \alpha &= \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} & \mathfrak{b}_p &= \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} \\ \\ P_p &= \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{p-2} \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \rho_{p-1} & \rho_{p-2} & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \\ \alpha &= P_p^{-1} \mathfrak{b}_p \end{aligned}$$

Elles relient les paramètres α_p du processus autoregressif a la fonction d'autocorrection théorique, qui définit \mathfrak{b}_p et P_p



Fonction d'autocorrelation de AR(2), $\alpha_1 = 0.1$ et $\alpha_2 = 0.7$

1.2.3 Processus MA(q)

Les processus MA(q) apparaissent naturellement comme un cas particulier d'un processus linéaire.

Définition 1.5 *Un processus à moyenne mobile d'ordre q est défini par:*

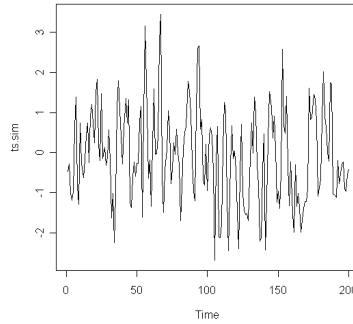
$$X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

où (ε_t) b.b.fort $\hookrightarrow N(0, \sigma^2)$.et $t \in \mathbb{Z}$, $\theta_0 = 1$, $E(X_t) = 0$.

On peut écrire :

$$\begin{aligned} X_t &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ &= \theta(B) \varepsilon_t \end{aligned} \tag{2}$$

où B est l'opérateur du retard $BY_t = Y_{t-1}$



Simulation d'un MA(1), $\theta = 0.5$

Stationnarité et inversibilité

Un processus (X_t) MA(q) est toujours stationnaire car il n'est autre qu'une somme pondérée de bruits blanc (v.a iid).

Définition 1.6

Si (X_t) est un processus MA(q), on dit que X_t est inversible lorsqu'il existe une suite Ψ_j avec $\sum_j |\Psi_j| < \infty$ tq

$$\varepsilon_t = \sum_{j=0}^{\infty} \Psi_j X_{t-j}.$$

Proposition 1.4

Un processus (X_t) MA(q) est inversible lorsque toutes les racines du polynôme $\theta(B)$ notées $\alpha_j \in \mathbb{C}$, $\forall j \leq q$ sont en module strictement supérieur à 1.

$$\theta(B) = \prod_{j=1}^q \left(1 - \frac{1}{\alpha_j} B\right) \text{ et } |\alpha_j| > 1 \forall j$$

Preuve:

La démonstration est faite de la même façon que dans la section précédente (stationnaire de AR) en échangeant uniquement les ε_t par X_t alors la relation (2) implique sous des conditions sur θ discutées.

Cas $MA(1)$:

Soit

$$X_t = \theta(B)\varepsilon_t = (1 - \theta B)\varepsilon_t$$

Par inversion du polynôme $\theta(B)$, on peut exprimer X_t sous la forme d'un $AR(\infty)$:

$$\varepsilon_t = \theta^{-1}(B)X_t$$

De la même façon que précédente on montre que si $|\theta| < 1, \theta(B)$ inversible et son inverse est :

$$(1 - \theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j$$

donc $\theta^j = \Psi_j$

$$\sum_{j=0}^{\infty} |\Psi_j| < \infty \Leftrightarrow \sum_{j=0}^{\infty} |\theta^j| < \infty$$

C'est une série géométrique convergente ssi $|\theta| < 1$.Donc

$$\varepsilon_t = \theta^{-1}(B)X_t = \sum_{j=0}^{\infty} \theta^j X_{t-j}$$

$\alpha_1 = \frac{1}{\theta}$ est la racine du polynôme $\theta(z) = (1 - \theta z)$.

Cas $MA(q)$:

$$\theta(B) = \prod_{j=1}^q \left(1 - \frac{1}{\alpha_j} B\right)$$

Posons

$$\frac{1}{\alpha_j} = \lambda_j \implies \theta(B) = \prod_{j=1}^q (1 - \lambda_j B)$$

Donc si $|\lambda_j| < 1$, $\theta(B)$ inversible et son inverse

$$\theta^{-1}(B) = \sum_{j=1}^q M_j (1 - \lambda_j B)^{-1}$$

$\alpha_j = \frac{1}{\lambda_j}$ les racines de $\theta(z)$ sont strictement supérieures à 1 en module alors

$$\varepsilon_t = \sum_{j=1}^q M_j (1 - \lambda_j B)^{-1} X_t$$

■

Remarque 1.1

Ce résultat indique qu'un processus (X_t) de $MA(q)$ peut être exprimé sous la forme $AR(\infty)$ si le polynôme $\theta(B)$ est inversible c'est-à-dire si les racines de $\theta(z)$ sont toutes supérieures à 1 en module.

Fonction d'autocorrelation

La fonction d'autocovariance de $MA(q)$ est :

$$\gamma_h = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-h} - \theta_1 \varepsilon_{t-h-1} - \dots - \theta_q \varepsilon_{t-h-q})]$$

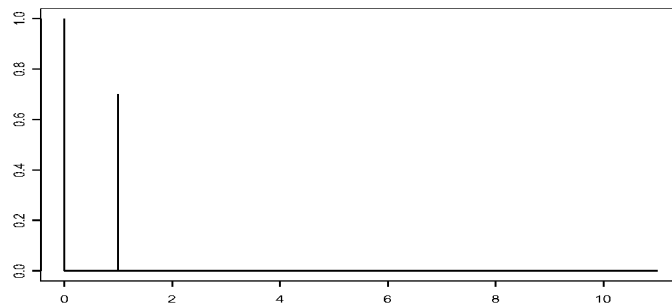
$$\text{avec } E(\varepsilon_t \varepsilon_{t'}) = 0, \quad t \neq t'$$

$$\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2$$

$$\gamma_1 = (-\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \dots + \theta_{q-1}\theta_q) \sigma_\varepsilon^2$$

$$\gamma_h = (-\theta_h + \theta_1\theta_{h+1} + \theta_2\theta_{h+2} + \dots + \theta_{q-h}\theta_q) \sigma_\varepsilon^2$$

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \begin{cases} \frac{-\theta_h + \theta_1\theta_{h+1} + \theta_2\theta_{h+2} + \dots + \theta_{q-h}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & h \leq q \\ 0 & h > q \end{cases}$$



Fonction d'autocorrélation d'un MA(1)

Chapitre 2

Estimateurs des Moindres Carrés Et du Maximum de Vraisemblance

2.1 Estimateurs des moindres carrés

Soit $(z_1, y_1), \dots, (z_n, y_n)$ un nuage de points dans \mathbb{R}^2 . IL s'agit de trouver *la meilleure droite*

$$y = \theta_0 + \theta_1 z$$

pour approximer les observations y_1, \dots, y_n d'une variable dépendante y prise aux valeurs fixes z_1, \dots, z_n de la variable indépendante z .

Les estimateurs des moindres carrés ordinaires (**OLS**) $\hat{\theta}_0, \hat{\theta}_1$ de θ_0, θ_1 sont définis par la minimisation de la somme :

$$S(\theta_0, \theta_1) := \sum_{t=1}^n (y_t - \theta_0 - \theta_1 z_t)^2$$

[la quantité $S(\theta_0, \theta_1)$ est identique à la distance carré euclidienne entre y et $\theta_0 I + \theta_1 z_t$ ie:

$$S(\theta_0, \theta_1) = \|y - \theta_0 I + \theta_1 z\|^2$$

où

$$\begin{aligned}z &= (z_1, \dots, z_n)' \\I &= (1, \dots, 1)' \\y &= (y_1, \dots, y_n)'\end{aligned}$$

Posons les dérivées partielles de S par rapport à θ_0 et θ_1 égales à zéro, nous trouvons le vecteur $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1)$ qui satisfait « **les équations normales** » :

$$Z'Z\hat{\theta} = Z'y$$

où Z est la matrice de dimension $(n \times 2)$: $Z = (I, z)$.

On a :

$$S(\theta) \geq 0 \text{ et } S(\theta) \rightarrow \infty \text{ qd } \|\theta\| \rightarrow \infty$$

Donc les équations normales ont au moins une solutions.

En effet:

Si $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}$ sont deux solutions de les équations normales alors un calcul simple montre cela :

$$(\hat{\theta}^{(1)} - \hat{\theta}^{(2)})Z'Z(\hat{\theta}^{(1)} - \hat{\theta}^{(2)}) = 0$$

ie:
$$Z\hat{\theta}^{(1)} = Z\hat{\theta}^{(2)}$$

La solution des équations normales est unique si et seulement si la matrice $Z'Z$ est non singulière.

Mais les calculs précédents montrent cela même si $Z'Z$ est singuliere le vecteur $\hat{y} = Z\hat{\theta}$ des valeurs adaptées est le même pour n'importe quel $\hat{\theta}$ solution des équations normales.

L'argument juste donné s'applique également bien à l'estimation des moindres carrées pour le modèle linéaire générale. Donnée un ensemble de point .

$$(z_{t1}, \dots, z_{tm}, y_t) \quad t = 1, \dots, n \text{ tq } m \leq n$$

Les estimateurs des moindres carrés

$$\widehat{\theta} = (\widehat{\theta}_1, \dots, \widehat{\theta}_m) \text{ de } \widehat{\theta} = (\theta_1, \dots, \theta_m)'$$

rendent maximum

$$\begin{aligned} S(\theta) &= \sum_{t=1}^n (y_t - \theta_1 z_{t1} - \dots - \theta_m z_{tm})^2 \\ &= \left\| y - \theta_1 z^{(1)} - \dots - \theta_m z^{(m)} \right\|^2 \end{aligned}$$

où

$$\begin{aligned} y &= (y_1, \dots, y_n)' \\ z^{(j)} &= (z_{1j}, \dots, z_{nj})' \\ j &= 1, \dots, m \end{aligned}$$

Comme dans le cas spécial précédent le $\widehat{\theta}$ satisfait les équations :

$$Z'Z\widehat{\theta} = Z'y$$

où Z la matrice $(n \times m)$, $Z = (z^{(1)}, \dots, z^{(m)})$. La solution de cette équation est unique si et seulement si le $X'X$ est non singulière. Dans ce cas :

$$\widehat{\theta}_{OLS} = (Z'Z)^{-1}Z'y$$

Si $Z'Z$ est singulier, il ya infinité de solutions $\widehat{\theta}$ mais le vecteur des valeurs adaptée $Z\widehat{\theta}$ est le même pour tous.

Exemple 2.1 Pour illustrer le cas général, adaptons la fonction quadratique

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

Aux données

$$\begin{cases} x = 0 & 1 & 2 & 3 & 4 \\ y = 1 & 0 & 3 & 5 & 8 \end{cases}$$

On pose

$$z_{t1} = x_t \text{ et } z_{t2} = x_t^2$$

donc

$$y_t = \theta_0 + \theta_1 z_{t1} + \theta_2 z_{t2}, \quad t = 1, \dots, 5$$

avec

$$z = (z^{(1)}, z^{(2)})' \text{ tq } \begin{cases} z^{(1)} = (0, 1, 2, 3, 4, 5) \\ z^{(2)} = (0, 1, 4, 9, 16) \end{cases}$$

La matrice Z pour ce problème est

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} = (I, z)$$
$$(Z'Z)^{-1} = \frac{1}{140} \begin{pmatrix} 124 & -108 & 20 \\ -108 & 174 & -40 \\ 20 & -40 & 10 \end{pmatrix}$$

l'estimation des moindres carrés $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)'$ est donc unique et donné par

$$\hat{\theta} = (Z'Z)^{-1}Z'y = \begin{pmatrix} 0.6 \\ -0.1 \\ 0.5 \end{pmatrix}$$

le vecteur des valeurs est donnée par :

$$\hat{y} = Z\hat{\theta} = (0.6, 1, 2.4, 4.8, 8.2)$$

par rapport aux valeurs observées $y = (1, 0, 3, 5, 8)'$.

2.1.1 Théorème de Gauss-Markov

Supposons que les observations y_1, \dots, y_n sont des réalisations des variables aléatoires Y_1, \dots, Y_n

:

$$Y_t = \theta_1 x_{t1} + \dots + \theta_m x_{tm} + V_t$$

tq $V_t \hookrightarrow WN(0, \sigma^2)$ un bruit blanc fort

Posons:

$$Y = (Y_1, \dots, Y_n)'$$
$$V = (V_1, \dots, V_n)'$$

Nous pouvons écrire ces équations comme :

$$Y = Z\theta + V$$

Supposons pour la simplicité que la matrice $Z'Z$ est non singulière (pour le cas général voir ; ex : SILVEY 1975) .

Alors l'estimation des moindres carrés de θ est :

$$\hat{\theta}_{OLS} = (Z'Z)^{-1}Z'Y$$

L'estimateur des moindres carrés de paramètre σ^2 est l'estimateur sans biais

$$\hat{\sigma}^2 = \frac{1}{n-m} \|Y - Z\hat{\theta}\|^2$$

Il est facile de voir que le $\hat{\theta}$ est également sans biais ie:

$$E(\hat{\theta}) = \theta$$

Il suit que si $C'\theta$ est une combinaison linéaire des paramètres $\theta_i, i = 1, \dots, m$, alors $C'\hat{\theta}$ est un estimateur sans biais de $C'\theta$.

Le Théorème de Gauss Markov indique que tous les estimateurs sans biais de $C'\theta$ de la forme $\sum_{t=1}^n a_t Y_t$ de la forme $C'\hat{\theta}$ ont la plus petite variance.

Dans le cas particulier où V_1, \dots, V_n sont $IIDN(0, \sigma^2)$, l'estimateur des moindres carrés $\hat{\theta}$ carrés suit une loi $N(\theta, \sigma^2(Z'Z)^{-1})$ et $\frac{(n-m)\widehat{\sigma^2}}{\sigma^2}$ Soit une loi χ^2 de degré de liberté $(n-m)$.

2.1.2 Les moindres carrés généralisé [GLS]

Le Théorème de Gauss Markov utilise l'hypothèse que le vecteur $V = (V_1, \dots, V_n)'$ est centré et de matrice de covariance non singulière $\sigma^2\Sigma$; $\Sigma \neq I$, et les V_i sont non corrélées.

Considérons le vecteur transformé d'observation

$$U = R^{-1}Y$$

où R est une matrice non singulière tel que

$$RR^{-1} = \Sigma$$

Alors

$$\begin{aligned} U &= R^{-1}Z\theta + R^{-1}V = M\theta + W \\ \text{où } M &= R^{-1}Z, W = R^{-1}V \end{aligned}$$

et W centré de matrice de covariance σ^2I car:

$$\begin{aligned} cov(W) &= E(WW') = E(R^{-1}V(R^{-1}V)') \\ &= R^{-1}E(VV')R^{-1'} = R^{-1}\sigma^2RR'R^{-1} \\ &= \sigma^2I \quad \text{on utilise } (A^{-1})' = A^{-1'} \end{aligned}$$

Théorème de Gauss Markov implique maintenant que la meilleure estimation linéaire de n'importe quelle combinaison linéaire $C'\theta$ est $C'\hat{\theta}$.

Où $\hat{\theta}$ est l'estimateur des moindres carrés généralisé (GLS) qui réduit au minimum

la quantité:

$$\|U - M\theta\|^2$$

Dans le cas où V_1, \dots, V_n sont non-corrélées et V_i centré, $\text{var}(V_i) = \sigma^2 r_i^2$, l'estimateurs des moindres carrés généralisé (**GLS**) minimisent la somme suivante:

$$\sum_{t=1}^n \frac{1}{r_i^2} (Y_t - \theta_1 z_{t1} - \dots - \theta_m z_{tm})^2$$

Dans le cas général, si $Z'Z$ et Σ sont tous les deux non singulière. L'estimateur des moindres carré généralisé (**GLS**) est donné par :

$$\hat{\theta}_{GLS} = (M'M)^{-1}M'U$$

encore:

$$\begin{aligned} \hat{\theta}_{GLS} &= (M'M)^{-1}M'U = ((R^{-1}Z)' R^{-1}Z)^{-1} (R^{-1}Z)' R^{-1}Y \\ &= (Z'(RR')^{-1}Z)^{-1} Z'(RR')^{-1}Y \quad \text{car} \quad (A^{-1})' = A^{-1} \\ \hat{\theta}_{GLS} &= (Z'\Sigma^{-1}Z)^{-1} Z'\Sigma^{-1}Y \end{aligned}$$

2.2 Méthode du Maximum de Vraisemblance

Les techniques d'estimations comme les moindres carrés, sont applicable aux modèles de régression. Dans quelques cas les moindres carrés et autres estimateurs sont tout simplement pas appropriés.

Nous introduisons une méthode d'estimation qui est beaucoup plus largement applicable que les moindres carrés mais qui nécessite également d'assez fortes hypothèses.

Il s'agit de l'estimation par la méthode **du maximum de vraisemblance (ML)**.

L'idée fondamentale de l'estimation par maximum de vraisemblance est de trouver un ensemble d'estimation de paramètres $\hat{\theta}$, telle que la vraisemblance de l'échantillon que nous utilisons soit maximale.

La densité de probabilité jointe ou vraisemblance pour le modèle que l'on estime est évaluée

aux valeurs observées de la (des) variable(s) dépendante(s) et traitée comme une fonction de paramètres du modèle.

Le vecteur $\hat{\theta}$ des estimateurs ML donne alors le maximum de cette fonction. Ce principe d'estimation est très largement applicable : si nous pouvons écrire la densité jointe de l'échantillon, nous pouvons en principe utilisé l'estimateur ML, sous certaines conditions de régularité.

Soit y un échantillon de taille n , la densité jointe sera désigné sous le nom de fonction de vraisemblance est noté $L(y, \theta)$ tel que :

$$L : A \times \Theta \rightarrow \mathbb{R} \quad y \in A \quad \text{et} \quad \theta \in \Theta \quad , y = (y_1, \dots, y_n).$$

avec A : un sous ensemble de \mathbb{R}^n

Θ : Espace paramétrique et est un sous ensemble de \mathbb{R}^k

Si les observations y_t sont indépendantes et de densité de probabilité $L_t(y_t, \theta)$, **la fonction de vraisemblance** est :

$$L(y, \theta) = \prod_{t=1}^n L_t(y_t, \theta) \tag{*}$$

Il est usage de maximiser $\log L(y, \theta) := \ell(y, \theta)$, car le log est une fonction croissante. $\ell(y, \theta)$ est appelé fonction de log- vraisemblance.

Si $\hat{\theta}$ maximise $\ell(y, \theta)$, il maximise $L(y, \theta)$.

Si on ne peut pas écrire $L(y, \theta)$ sous la forme (*), il est parfois possible de la factoriser sous :

$$\begin{aligned} L(y, \theta) &= \prod_{t=1}^n L_t(y_t \setminus y_{t-1}, \dots, y_1, \theta) \\ &= L_1(y_1, \theta) L_2(y_2 \setminus y_1, \theta) \dots L_n(y_n \setminus y_{n-1}, \dots, y_1, \theta) \end{aligned}$$

Donc

$$\ell(y, \theta) = \sum_{t=1}^n \ell_t(y_t \setminus y_{t-1}, \dots, y_1, \theta)$$

où

$$\ell_t(y_t \setminus y_{t-1}, \dots, y_1, \theta) = \log L_t(y_t \setminus y_{t-1}, \dots, y_1, \theta)$$

Nous donnons la définition de l'estimation par maximum de vraisemblance.

Nous disons que $\hat{\theta} \in \Theta$ est une estimation par maximum de vraisemblance (**ML**) pour les

observations y si

$$\ell(y, \hat{\theta}) = \max_{\theta \in \Theta} \ell(y, \theta)$$

c'est à dire:

$$\ell(y, \hat{\theta}) \geq \ell(y, \theta) \quad \forall \theta \in \Theta$$

Si $\ell(y, \hat{\theta}) > \ell(y, \theta), \forall \theta \in \Theta$ alors $\hat{\theta}$ est unique .

Dans ce cas $\hat{\theta}$ est dit **estimateur ML de type1**.

Il est souvent commode d'utiliser une autre définition de ML ; lorsque la fonction de vraisemblance L est deux fois continûment différentiable par rapport à θ et la matrice

Hessienne $H(y, \theta)$ d'éléments:

$$H_{ij}(y; \theta) = \frac{\partial^2 \ell(y, \theta)}{\partial \theta_i \partial \theta_j}$$

est une matrice définie négative.

Dans ce cas, $\forall \theta \in \Theta$, $\hat{\theta}$ est l'estimateur de maximum de vraisemblance est **solution du système suivant**:

$$g(y, \hat{\theta}) = 0 \quad (**)$$

et $\hat{\theta}$ est de **type2**

où g est le vecteur gradient, $g \in \mathbb{R}^k$, défini par:

$$g^T(y, \theta) \equiv D_{\theta} \ell(y, \theta) = \sum_{t=1}^n D_{\theta} \ell_t(y, \theta)$$

Puisque $D_{\theta} \ell$ est un vecteur ligne, g est le vecteur colonne des dérivées partielles de la fonction de log-vraisemblance ℓ par rapport aux paramètres θ .

Nous pouvons définir une matrice $G(y, \theta)$ de dimension $n \times k$, d'éléments

$$G_{ti}(y, \theta) := \frac{\partial \ell_t(y, \theta)}{\partial \theta_i}$$

$G(y, \theta)$: Matrice des contributions au gradient (CG) tel que g^T est un vecteur de G .

La matrice **Hessienne** $H(y, \theta)$ associée à la fonction de vraisemblance $\ell(y, \theta)$ est la matrice de dimension $k \times k$ d'éléments :

$$H_{ij}(y, \theta) := \frac{\partial^2 \ell(y, \theta)}{\partial \theta_i \partial \theta_j}$$

L'espérance de la **Hessienne** pour un échantillon de taille n est définie par :

$$]^{-}[^n(\theta) := E_{\theta}(n^{-1}H(y, \theta))$$

La limite de la Hessienne ou **Hessienne asymptotique** si elle existe est définie par :

$$]^{-}[(\theta) := \lim_{n \rightarrow \infty}]^{-}[^n(\theta)$$

Cette matrice symétrique, et en générale semi définie négative.

Nous définissons la matrice $I_t(\theta)$ d'**information contenue dans** y_t la matrice $k \times k$ d'element :

$$(I_t(\theta))_{ij} := E_{\theta}(G_{ti}(\theta)G_{tj}(\theta))$$

La matrice d'**information moyenne** pour un échantillon y de taille n est définie par :

$$\mathfrak{S}^n(\theta) := \frac{1}{n} \sum_{t=1}^n I_t(\theta) = n^{-1}I^n$$

La matrice d'information asymptotique si elle existe est définie par :

$$\mathfrak{S}(\theta) := \lim_{n \rightarrow \infty} \mathfrak{S}^n(\theta)$$

La matrice $I_t(\theta)$ mesure la quantité d'information contenue dans l' observation y_t .

$I^n = n\mathfrak{S}^n$: mesure la quantité d'information contenue dans l'échantillon y .

Nous commençons notre analyse par le lemme suivant ,en démontrant un résultat simple mais fondamental concernant le gradient g et la matrice G

Lemme 2.1

Supposons que y_t à une densité $L_t(y_t, \theta)$, et $\ell_t(y_t, \theta) = \log L_t(y_t, \theta)$ et $G_{ti}(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta_i}$.

Alors:

$$E_{\theta}(G_{ti}(\theta)) = E_{\theta}\left(\frac{\partial \ell_t(\theta)}{\partial \theta_i}\right) = 0$$

Ceci implique que

$$E_{\theta}(g(\theta)) = 0 \text{ et } E_{\theta}(G(\theta)) = 0$$

Preuve: On a

$$\begin{aligned} E_{\theta}(G_{ti}(y_t, \theta)) &= \int \frac{\partial \log L_t(y_t, \theta)}{\partial \theta_i} L_t(y_t, \theta) dy_t \\ &= \int \frac{1}{L_t(y_t, \theta)} \frac{\partial L_t(y_t, \theta)}{\partial \theta_i} L_t(y_t, \theta) dy_t \\ &= \int \frac{\partial L_t(y_t, \theta)}{\partial \theta_i} dy_t = \frac{\partial}{\partial \theta_i} \int L_t(y_t, \theta) dy_t \\ &= \frac{1}{\partial \theta_i}(1) = 0 \end{aligned}$$

■

Remarque 2.1

Ce Lemme permet d'appliquer le Théorème central limite à la quantité

$$n^{-\frac{1}{2}}g(\theta)$$

Le résultat suivant donne des propriétés asymptotiques des estimateurs de maximum de vraisemblance (**ML**).

2.2.1 Théorème

Sous les conditions :

- i) $\ell_t(y, \theta)$ sont au moins deux fois continûment différentiables en $\theta \in \Theta$ et les dérivés sont intégrables.
- ii) $E\left(\frac{\partial^2 \ell_t(y, \theta)}{\partial \theta_i \partial \theta_j}\right) < 0$

Nous avons:

- 1) $\hat{\theta}_n \rightarrow \theta$ p.s
- 2) $n^{1/2}(\hat{\theta}_n - \theta) \rightarrow N(0, \mathfrak{I}^{-1}(\theta))$

Preuve:

On admet 1).

Pour 2) nous commençons par le développement de Taylor des équations de vraisemblance

(**) autour de θ pour obtenir (avec $g(y, \theta) = g(\theta)$):

$$0 = g(\hat{\theta}) = g(\theta) + H(\tilde{\theta})(\hat{\theta} - \theta)$$

où $\tilde{\theta} \in (\hat{\theta}_n, \theta)$

ce qui implique
$$n^{1/2}(\hat{\theta}_n - \theta) = - \underbrace{(n^{-1}H(\tilde{\theta}))^{-1}}_{\text{matrice } (k \times k)} (n^{1/2}g(\theta))_{\text{vecteur } (k)}$$

Asymptotiquement la matrice est non aléatoire et le vecteur suit une loi normale.

Ce qui implique que $n^{1/2}(\hat{\theta}_n - \theta)$ asymptotiquement normal .

Montrons que $n^{-1}H(\tilde{\theta})$ tend vers une matrice déterministe quand n tend vers l'infini.

Rappelons que le (i,j)ème élément de $n^{-1}H(\tilde{\theta})$ est :

$$(n^{-1}H(\tilde{\theta}))_{ij} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \quad \text{évalué en } \theta = \tilde{\theta}$$

Posons

$$Z_t = \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j}$$

Donc

$$\begin{aligned} (n^{-1}H(\tilde{\theta}))_{ij} &= \frac{1}{n} \sum_{t=1}^n Z_t \\ &= \overline{Z_n} \text{ et } E(Z_t) < \infty \end{aligned}$$

On applique la loi forte des grands nombre pour avoir:

$$\overline{Z_n} \rightarrow \mu = E(Z)$$

Donc

$$\frac{1}{n}H(\tilde{\theta}) \xrightarrow[n \rightarrow \infty]{}]-(\tilde{\theta})$$

et on a par 1)

$$\hat{\theta} \rightarrow \theta \text{ et } \hat{\theta} \leq \tilde{\theta} \leq \theta$$

Donc

$$\frac{1}{n}H(\tilde{\theta}) \xrightarrow[n \rightarrow \infty]{}]-(\theta)$$

Alors

$$n^{1/2}(\hat{\theta} - \theta) \approx -]^{-1}(\theta)(n^{1/2}g(\theta)) \quad (***)$$

L'élément stochastique dans le membre droite est $n^{1/2}g(\theta)$ donné par :

$$n^{-1/2} \sum_{t=1}^n \frac{\partial \log L_t(y_t, \theta)}{\partial \theta_i} = n^{-1/2} \sum_{t=1}^n G_{ti}(\theta)$$

Nous avons d'après le Lemme 2.1 : $E(G_{ti}(\theta)) = 0$. On applique le T.C.L($EG_{ti}^2(\theta) < \infty$)

$n^{1/2}g(\theta)$ converge en loi vers une loi normale.

Par suite $n^{1/2}(\hat{\theta} - \theta)$ est asymptotique normale.

La matrice de covariance asymptotique de $n^{1/2}(\hat{\theta} - \theta)$ est l'espérance asymptotique de $n(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T$:

$$V(n^{1/2}(\hat{\theta} - \theta)) = E_{\theta}(n(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T)$$

En utilisant (***), nous avons

$$V(n^{1/2}(\hat{\theta} - \theta)) = (-]^{-1}(\theta)) \left(\frac{1}{n} E_{\theta}(g(\theta)g^T(\theta)) \right) (-]^{-1}(\theta))^T$$

$$\text{on a } (-]^{-1}(\theta))^T =]^{-1}(\theta) \quad \text{car }]^{-1}(\theta) \text{ symétrique}$$

Un élément de l'espérance dans le facteur central est

$$\frac{1}{n} E_{\theta} \left(\sum_{t=1}^n G_{ti}(\theta) \right) \left(\sum_{s=1}^n G_{sj}(\theta) \right)$$

$$\text{et } G_{ti}(\theta)G_{sj}(\theta) = \frac{\partial \log L_t \partial \log L_s}{\partial \theta_i \partial \theta_j}$$

$$E_{\theta}(G_{ti}(\theta)G_{sj}(\theta)) \begin{cases} = 0 & \text{si } t \neq s \\ \neq 0 & \text{si } t = s \end{cases}$$

Donc

$$\begin{aligned} & \frac{1}{n} E_{\theta} \left(\left(\sum_{t=1}^n G_{ti}(\theta) \right) \left(\sum_{s=1}^n G_{sj}(\theta) \right) \right) \\ = & \frac{1}{n} E_{\theta} \left(\sum_{t=1}^n G_{ti}(\theta) G_{tj}(\theta) \right) \dots \dots \dots (+) \end{aligned}$$

on a

$$\left\{ \begin{array}{l} I_t(\theta)_{ij} = E_{\theta}(G_{ti}(\theta)G_{tj}(\theta)) \\ \mathfrak{I}^n(\theta) := \frac{1}{n} \sum_{t=1}^n I_t(\theta) = n^{-1} I^n \end{array} \right.$$

Nous voyons que le membre droite de (+) correspond à $\mathfrak{I}^n(\theta)$ et $\mathfrak{I}^n(\theta) \rightarrow \mathfrak{I}(\theta)$

$$V^{\infty}(n^{1/2}(\hat{\theta} - \theta)) =]-[-^{-1}(\theta)\mathfrak{I}(\theta)]-[-^{-1}(\theta) \quad (++)$$

On va simplifier l'expression (++) et on a:

$$]-[-(\theta) = -\mathfrak{I}(\theta)$$

En effet

$$\left\{ \begin{array}{l} E_{\theta} \left(\frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j} \right) = -E_{\theta} \left(\frac{\partial \ell_t \partial \ell_t}{\partial \theta_i \partial \theta_j} \right) \\ H_{ij}(y, \theta) = \frac{\partial^2 \ell(y, \theta)}{\partial \theta_i \partial \theta_j} \end{array} \right.$$

$$\begin{aligned}]-[-(\theta) &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n E_{\theta} \left(\frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \right) \right) \\ &= - \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n E_{\theta} \left(\frac{\partial \ell_t \partial \ell_t}{\partial \theta_i \partial \theta_j} \right) \right) \\ &= - \lim_{n \rightarrow \infty} \mathfrak{I}^n(\theta) \\ &= - \mathfrak{I}(\theta) \end{aligned}$$

Conclusion :

$$V(n^{1/2}(\hat{\theta} - \theta)) = \mathfrak{I}^{-1}(\theta) = -]-[-^{-1}(\theta)$$

2.2.2 Fonction de Log-vraisemblance Concentrée :

Souvent les paramètres dans une fonction de logvraisemblance sont partitionnés en deux ensemble de façon à rendre l'écriture de l'estimation **ML** d'un groupe de paramètres comme une fonction des valeurs de l'autre groupe. Nous rencontrons un exemple en connexion avec l'estimation **ML dans** des modèles de régression. Dans la suite il peut être très pratique d'exprimer **la fonction de logvraisemblance** comme une fonction d'un seul des deux groupe de paramètre.

Soit $\ell(y, \theta) = \ell(y, \theta_1, \theta_2)$ donc $\hat{\theta}_1$, $\hat{\theta}_2$ sont solution de :

$$\begin{cases} D_1 \ell(y, \theta_1, \theta_2) = \frac{\partial \ell(y, \theta_1, \theta_2)}{\partial \theta_1} = 0 \\ D_2 \ell(y, \theta_1, \theta_2) = \frac{\partial \ell(y, \theta_1, \theta_2)}{\partial \theta_2} = 0 \end{cases}$$

Supposons qu'il soit possible de résoudre la seconde équation afin de pouvoir écrire $\theta_2 = \tau(y, \theta_1)$ ceci implique que identiquement en θ_1

$$D_2 \ell(y, \theta_1, \tau(y, \theta_1)) = 0$$

Donc **la fonction logvraisemblance concentrée** :

$$\ell^c(y, \theta_1) := \ell(y, \theta_1, \tau(y, \theta_1))$$

Si $\hat{\theta}_1$ maximise celle ci alors on a :

$$\hat{\theta}_2 = \tau(y, \hat{\theta}_1)$$

et de plus

$$\ell^c(y, \hat{\theta}_1) \equiv \ell(y, \hat{\theta})$$

Chapitre 3

Estimateur de Maximum de Vraisemblance dans une Moyenne mobile : Cas Scalaire

3.1 Introduction

Un processus moyenne mobile d'ordre q est défini par :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.1.1)$$

où (ε_t) est un bruit blanc $\varepsilon_t \hookrightarrow N(0, \sigma^2)$

le processus (3.1.1) est écrit de manière compacte en utilisant l'opérateur retard L par :

$$X_t = \theta(L)\varepsilon_t \quad (3.1.2)$$

où $Lx_t = x_{t-1}$ et $\theta(L)$ est définie par $(1 + \theta_1 L + \dots + \theta_q L^q)$:

$$\theta(L) = \prod_{i=1}^q (1 - \alpha_i L) \quad (3.1.4)$$

De plus si les racines de $\theta(L)$ sont en module strictement supérieur à 1

c-a-d : $\frac{1}{\alpha_i} > 1 \Leftrightarrow |\alpha_i| < 1$ alors (1,1) est inversible (Voir la proposition 1.4)

Souvent l'estimation des paramètres $\theta' = (\theta_1, \dots, \theta_q)$ par le maximum de vraisemblance procède généralement par un calcul numérique [en pratique, les estimateurs ML sont déterminés numériquement par des algorithmes itératifs : la méthode de Newton-Raphson] .

Durbin (1959) et Walker (1961) ont considéré d'autres méthodes que le maximum de vraisemblance, parce que la maximisation de la fonction de vraisemblance ne pourrait pas être effectuée par des méthodes linéaires.

Une classe d'estimateur non linéaire (les estimateurs des moindres carrés (LS)) est considéré asymptotiquement comme des estimateur (ML).

Le calcul, des estimateurs (LS), considère les conditions initiales $\varepsilon^{*'} = (\varepsilon_{1-q}, \dots, \varepsilon_0)$ déterministes fixés. Dans ce cas l'expression de la fonction de log-vraisemblance est donné dans le lemme suivant :

Lemme 3.1

Soit $(X_t)_t$ un processus $MA(q)$: $X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$ avec $\varepsilon_t \hookrightarrow N(0, \sigma^2)$;

On suppose que $\varepsilon^{*'} = (\varepsilon_{1-q}, \dots, \varepsilon_0)$ sont des nombres fixes .

Alors le logarithme de la fonction de vraisemblance pour un échantillon de n observation $X = (X_1, \dots, X_n)$ est :

$$\ell(x, \theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\theta) \tag{3.1.5}$$

où

$$S(\theta) = \varepsilon' \varepsilon$$

Preuve:

Cas $q = 1$:

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad \text{où} \quad \varepsilon_t \hookrightarrow N(0, \sigma^2)$$

d'où la loi conditionnelle:

$$\mathcal{L}(X_t \mid \varepsilon_{t-1}) \hookrightarrow N(\theta \varepsilon_{t-1}, \sigma^2)$$

et sa densité est:

$$f(x_t \setminus \varepsilon_{t-1}; \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_t - \theta\varepsilon_{t-1})^2}{2\sigma^2}}$$

Supposons que ε_0 est fixé et $\varepsilon_1 = X_1 - \theta\varepsilon_0$

On a ε_1 en fonction de X_1 et ε_0 alors

$$\mathcal{L}(X_2 \setminus X_1, \varepsilon_0) = \mathcal{L}(X_2 \setminus \varepsilon_1)$$

De cette relation on obtient la densité: .

$$\begin{aligned} f(x_2 \setminus x_1, \varepsilon_0; \theta, \sigma^2) &= f(x_2 \setminus \varepsilon_1; \theta, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_2 - \theta\varepsilon_1)^2}{2\sigma^2}} \end{aligned}$$

Si ε_1 est fixé, ε_2 peut se calculer à partir de :

$$\varepsilon_2 = X_2 - \theta\varepsilon_1$$

Ainsi, en fixant ε_0 la séquence $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ peut se calculer par:

$$\begin{aligned} \varepsilon_0 &= \varepsilon_0 \\ \varepsilon_1 &= X_1 - \theta\varepsilon_0 \\ \varepsilon_2 &= X_2 - \theta\varepsilon_1 \\ &\cdot \\ &\cdot \\ \varepsilon_n &= X_n - \theta\varepsilon_{n-1} \end{aligned}$$

Pour $t = 1, 2, 3, \dots, n$ la densité conditionnelle de X_t sachant $X_{t-1}, \dots, X_1, \varepsilon_0$ est:

$$\mathcal{L}(X_t \setminus X_{t-1}, X_{t-2}, \dots, X_1, \varepsilon_0) = \mathcal{L}(X_t \setminus \varepsilon_{t-1})$$

d'où

$$f(x_t \setminus x_{t-1}, x_{t-2}, \dots, x_1, \varepsilon_0) = f(x_t \setminus \varepsilon_{t-1})$$

Par suite

$$L(\theta, \sigma^2) = f(x_n \setminus x_{n-1}, x_{n-2}, \dots, x_1, \varepsilon_0)$$

$$L(\theta, \sigma^2) = f(x_1 \setminus \varepsilon_0; \theta, \sigma^2) \prod_{t=2}^n f(x_t \setminus x_{t-1}, \dots, x_1, \varepsilon_0; \theta, \sigma^2)$$

ce résultat est obtenue par la factorisation de

$$f(x_1, \dots, x_n) = \prod_{t=1}^n f(x_t \setminus x_{t-1}, \dots, x_1)$$

Par exemple ($n = 3$)

$$f(x_1, x_2, x_3) = f(x_3 \setminus x_2, x_1) f(x_2 \setminus x_1) f(x_1)$$

$$= \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} \frac{f(x_1, x_2)}{f(x_1)} f(x_1)$$

Enfin:

$$L(\theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - \theta\varepsilon_0)^2}{2\sigma^2}} \prod_{t=2}^n f(x_t \setminus \varepsilon_{t-1}; \theta, \sigma^2)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - \theta\varepsilon_0)^2}{2\sigma^2}} \prod_{t=2}^n e^{-\frac{(x_t - \theta\varepsilon_{t-1})^2}{2\sigma^2}}$$

et comme $\varepsilon_t = X_t - \theta\varepsilon_{t-1}$, donc

$$L(\theta, \sigma^2) = \left[\frac{1}{2\pi\sigma^2} \right]^{n/2} e^{-\sum_{t=1}^n \frac{\varepsilon_t^2}{2\sigma^2}}$$

la fonction de log-vraisemblance (remplaçant $\ell(y, \theta, \sigma^2)$ par $\ell(\theta, \sigma^2)$ pour faciliter l'écriture).

$$\ell(\theta, \sigma^2) = \log L(\theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \sum_{t=1}^n \frac{\varepsilon_t^2}{2\sigma^2}$$

Cas : $q > 1$

$$X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Pour $\varepsilon^{*'} = (\varepsilon_{1-q}, \dots, \varepsilon_0)$ fixé et à l'aide de ces valeurs on peut itérer la relation:

$$\varepsilon_t = X_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad \text{pour } t = 1, \dots, n$$

Définissons ε_0 comme $(\varepsilon_{1-q}, \dots, \varepsilon_0)'$ on a le même résultat comme dans le cas précédent ($q = 1$)

$$\begin{aligned} \ell(\theta, \sigma^2) &= \log f(x_n, x_{n-1}, \dots, x_1 \mid \varepsilon_0, \dots, \varepsilon_{1-q}) \\ \ell(\theta, \sigma^2) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \sum_{t=1}^n \frac{\varepsilon_t^2}{2\sigma^2} \end{aligned}$$

■

Propriétés :

1) $\hat{\sigma}^2$ l'estimateur du maximum de vraisemblance de σ^2 est donné par:

$$\hat{\sigma}^2 = \frac{S(\theta)}{n} \tag{3.1.6}$$

2) $\ell^c(\theta)$ fonction de log-vraisemblance concentré est donné par:

$$\begin{aligned} \ell^c(\theta) &: = \ell(\theta, \tau(\theta)) = -\frac{n}{2} \log 2\pi - \frac{n}{2} - \frac{n}{2} \log \frac{S(\theta)}{n} \\ \text{avec } \tau(\theta) &= \frac{S(\theta)}{n} \end{aligned} \tag{3.1.7}$$

Preuve:

1) $\frac{\partial \ell}{\partial \sigma^2} = 0$ entraîne que

$$-\frac{n}{2\sigma^2} + \frac{S(\theta)}{2\sigma^2} = 0$$

Donc

$$\hat{\sigma}^2 = \frac{S(\theta)}{n}$$

2) en remplaçant σ^2 dans (3.1.5), la fonction de log-vraisemblance concentré s'exprime

uniquement en fonction du parametre θ par:

$$\begin{aligned}\ell^c(\theta) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \frac{S(\theta)}{n} - \frac{n}{2S(\theta)} S(\theta) \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} - \frac{n}{2} \log \frac{S(\theta)}{n}\end{aligned}$$

Remarque 3.1

1) Minimiser la somme de résidues $S(\theta)$ est équivalent a maximiser (3.1.5).

La formule (3.1.5), (et par conséquent les estimateurs des moindres carrés), dépend des conditions initiales ε^* (qui sont considérées des constantes).

Si ces conditions sont supposés des variables aléatoires, alors la fonction de vraisemblance (globale) est donnée dans la prochaine section .

2) Kang (1975) a prouvé que l'hypothèse que le processus soit inversible est nécessaire avant toute estimation par les moindres carrés. L'estimation de maximum de vraisemblance exige cette condition pour l'identification.

3.2 Fonction de vraisemblance

Ecrivant (3.1.1) en tant que système de $(n + q)$ équations en ε_t :

$$\left\{ \begin{array}{l} \varepsilon_{1-q} = \varepsilon_{1-q} \\ \cdot \\ \varepsilon_0 = \varepsilon_0 \\ \varepsilon_1 = X_1 - \theta_1 \varepsilon_0 - \dots - \theta_q \varepsilon_{1-q} \\ \cdot \\ \varepsilon_n = X_n - \theta_1 \varepsilon_{n-1} - \dots - \theta_q \varepsilon_{n-q} \end{array} \right. \quad (3.2.1)$$

Proposition 3.1

Posons $\varepsilon = (\varepsilon_{1-q}, \dots, \varepsilon_n)$, $X = (X_1, \dots, X_n)$, $\varepsilon^* = (\varepsilon_{1-q}, \dots, \varepsilon_0)$.

Alors le système (3.2.1) se réduit à:

$$\varepsilon = MX + W\varepsilon^* \quad (3.2.2)$$

où M et W sont deux matrices données dépendant de θ seulement (voir la démonstration pour leurs expressions).

Application: On applique l'expression (3.2.2) pour deux cas:

Cas $q = 1$

pour tout n on a:

$$(3.2.1) \Leftrightarrow \begin{cases} \varepsilon_0 = \varepsilon_0 \\ \varepsilon_1 = X_1 - \theta_1 \varepsilon_0 \\ \varepsilon_2 = X_2 - \theta_1 \varepsilon_1 \\ \\ \varepsilon_n = X_n - \theta_1 \varepsilon_{n-1} \end{cases}$$

$$\Leftrightarrow \begin{cases} \varepsilon_0 = \varepsilon_0 \\ \varepsilon_1 = X_1 - \theta_1 \varepsilon_0 \\ \varepsilon_2 = X_2 - \theta_1(X_1 - \theta_1 \varepsilon_0) = X_2 - \theta_1 X_1 + \theta_1^2 \varepsilon_0 \\ \varepsilon_3 = X_3 - \theta_1(X_2 - \theta_1 X_1 + \theta_1^2 \varepsilon_0) = X_3 - \theta_1 X_2 + \theta_1^2 X_1 - \theta_1^3 \varepsilon_0 \\ \varepsilon_4 = X_4 - \theta_1 X_3 + \theta_1^2 X_2 - \theta_1^3 X_1 + \theta_1^4 \varepsilon_0 \\ \\ \varepsilon_n = X_n - \theta_1 X_{n-1} + (-\theta_1)^2 X_{n-2} + \dots + (-\theta_1)^{n-1} X_1 + (-\theta_1)^n \varepsilon_0 \end{cases}$$

Donc

$$\varepsilon_i = X_i + \sum_{k=1}^{i-1} (-1)^k \theta_1^k X_{i-k} + (-1)^i \theta_1^i \varepsilon_0$$

alors le système (3.2.1) est donné par :

$$\underbrace{\begin{pmatrix} \varepsilon_{-1} \\ \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}}_{\varepsilon} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -\theta_1 & 1 & 0 & 0 \\ \theta_1^2 - \theta_2 & -\theta_1 & 1 & 0 \\ \theta_1(2\theta_2 - \theta_1^2) & \theta_1^2 - \theta_2 & -\theta_1 & 1 \end{pmatrix}}_M \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}}_X + \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\theta_2 & -\theta_1 \\ \theta_1\theta_2 & \theta_1^2 - \theta_2 \\ \theta_2(\theta_2 - \theta_1^2) & \theta_1(2\theta_2 - \theta_1^2) \\ -\theta_1\theta_2(2\theta_2 - \theta_1^2) & \theta_1^2(\theta_1^2 - 3\theta_2) + \theta_2^2 \end{pmatrix}}_W \underbrace{\begin{pmatrix} \varepsilon_{-1} \\ \varepsilon_0 \end{pmatrix}}_{\varepsilon^*}$$

Preuve:

Il suffit de remplacer successivement dans le système (3.2.1), ε_t $t \geq 1$ en fonction de (X_1, \dots, X_{n-1}) et de $(\varepsilon_{1-q}, \dots, \varepsilon_0)$ pour obtenir (3.2.2).

Montrons par récurrence la relation (3.2.2) :

$n = 1$

$$\begin{aligned}
 \varepsilon_{1-q} &= \varepsilon_{1-q} \\
 &\cdot \\
 &\cdot \\
 \varepsilon_0 &= \varepsilon_0 \\
 \varepsilon_1 &= X_1 - \theta_1\varepsilon_0 - \theta_2\varepsilon_{-1} - \dots - \theta_q\varepsilon_{1-q}
 \end{aligned}$$

⇔

$$\begin{pmatrix} \varepsilon_{1-q} \\ \cdot \\ \cdot \\ \varepsilon_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ \cdot \\ \cdot \\ X_n \end{pmatrix} + \begin{pmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & 0 & \cdot \\ \cdot & 0 & \cdot & 1 \\ -\theta_q & \cdot & \cdot & -\theta_1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1-q} \\ \cdot \\ \cdot \\ \varepsilon_0 \end{pmatrix}$$

⇔ $\varepsilon = MX + W \varepsilon^*$.

Donc la relation est vraie pour $n = 1$

On suppose que la relation est juste pour n . Montrons la relation pour $n + 1$:

Le système (3.2.2), s'écrit pour n :

$$M = \begin{pmatrix} m_{11} & m_{12} & \cdot & m_{1n} \\ m_{21} & m_{22} & \cdot & m_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ m_{q1} & \cdot & \cdot & m_{qn} \\ \cdot & \cdot & \cdot & \cdot \\ m_{n+q1} & m_{n+q2} & \cdot & m_{n+qn} \end{pmatrix} \text{ et } W = \begin{pmatrix} w_{11} & w_{12} & \cdot & w_{1q} \\ w_{21} & w_{22} & \cdot & w_{2q} \\ \cdot & \cdot & \cdot & \cdot \\ w_{q1} & \cdot & \cdot & w_{qq} \\ \cdot & \cdot & \cdot & \cdot \\ w_{n+q1} & w_{n+q2} & \cdot & w_{n+qq} \end{pmatrix}$$

$$(3.2.2) \Leftrightarrow \left\{ \begin{array}{l} \varepsilon_{1-q} = \sum_{i=1}^n m_{1,i} X_i + \sum_{t=1}^q w_{1,t} \varepsilon_{t-q} = \varepsilon_{1-q} \dots \dots \dots (3.2.2.1) \\ \varepsilon_{2-q} = \sum_{i=1}^n m_{2,i} X_i + \sum_{t=1}^q w_{2,t} \varepsilon_{t-q} = \varepsilon_{2-q} \dots \dots \dots (3.2.2.2) \\ \cdot \\ \varepsilon_0 = \sum_{i=1}^n m_{q,i} X_i + \sum_{t=1}^q w_{q,t} \varepsilon_{t-q} = \varepsilon_0 \dots \dots \dots (3.2.2.3) \\ \varepsilon_1 = \sum_{i=1}^n m_{q+1,i} X_i + \sum_{t=1}^q w_{q+1,t} \varepsilon_{t-q} = X_1 - \theta_1 \varepsilon_0 - \dots - \theta_q \varepsilon_{1-q} \\ \cdot \\ \varepsilon_n = \sum_{i=1}^n m_{q+n,i} X_i + \sum_{t=1}^q w_{q+n,t} \varepsilon_{t-q} \end{array} \right.$$

c.a.d:

$$\varepsilon_j = \sum_{i=1}^n m_{q+j,i} X_i + \sum_{t=1}^q w_{q+j,t} \varepsilon_{t-q}$$

Avant d'exprimer ε_{n+1} , nous développons des résultats utiles pour la suite.

Nous remarquons de (3.2.2.1) et (3.2.2.2) et (3.2.2.3) que

$$m_{j,i} = 0 \quad j = 1 \dots q \text{ et } i = 1 \dots n \quad (3.2.2.4)$$

$$\begin{cases} w_{t,t} = 1 & t = 1 \dots q \\ w_{j,t} = 0 & t \neq j \quad t, j = 1 \dots q \end{cases} \quad (3.2.2.5)$$

et

$$\left\{ \begin{array}{l} \varepsilon_1 \text{ depend seulement de } X_1 \Rightarrow m_{q+1,1} = 1 \text{ et } m_{q+1,i} = 0 \text{ pour } i = 2 \dots n \\ \varepsilon_2 \text{ depend seulement de } X_1, X_2 \Rightarrow m_{q+2,2} = 1 \text{ et } m_{q+2,i} = 0 \text{ pour } i = 3 \dots n \\ \quad \cdot \\ \quad \cdot \\ \varepsilon_{n-1} \approx X_1, \dots, X_{n-1} \Rightarrow m_{q+n-1,n-1} = 1 \text{ et } m_{q+n-1,i} = 0 \text{ pour } i = n \\ \varepsilon_n \approx X_1, \dots, X_n \Rightarrow m_{q+n,n} = 1 \end{array} \right.$$

par suite (3.2.2.6) $\begin{cases} m_{j,i} = 0 & i > j - q \quad j = q + 1, \dots, n + q \\ m_{j,i} = 1 & i = j - q \quad j = q + 1, \dots, n + q \end{cases}$

Donc

$$\begin{aligned}
\varepsilon_{n+1} &= X_{n+1} - \theta_1 \varepsilon_n - \dots - \theta_q \varepsilon_{n+1-q} \\
&= X_{n+1} - \theta_1 \left(\sum_{i=1}^n m_{q+n,i} X_i + \sum_{t=1}^q w_{q+n,t} \varepsilon_{t-q} \right) \\
&\quad - \dots - \theta_q \left(\sum_{i=1}^n m_{1+n,i} X_i + \sum_{t=1}^q w_{1+n,t} \varepsilon_{t-q} \right) \\
&= X_{n+1} + (-\theta_1 m_{q+n,1} - \theta_2 m_{q+n-1,1} \dots - \theta_q m_{n+1,1}) X_1 \\
&\quad + (-\theta_1 m_{q+n,2} - \theta_2 m_{q+n-1,2} - \dots - \theta_q m_{n+1,2}) X_2 \\
&\quad \cdot \\
&\quad + (-\theta_1 m_{q+n,i} - \theta_2 m_{q+n-1,i} - \dots - \theta_q m_{n+1,i}) X_i \\
&\quad + (-\theta_1 m_{q+n,n} - \theta_2 m_{q+n-1,n} - \dots - \theta_q m_{n+1,n}) X_n \\
&\quad + (-\theta_1 w_{q+n,1} - \theta_2 w_{q+n-1,1} - \dots - \theta_q w_{n+1,1}) \varepsilon_{1-q} \\
&\quad \cdot \\
&\quad + (-\theta_1 w_{q+n,q} - \theta_2 w_{q+n-1,q} - \dots - \theta_q w_{n+1,q}) \varepsilon_0
\end{aligned}$$

Nous remarquons aussi

$$\begin{cases} m_{j;i} = -\theta_1 m_{j-1,i} - \theta_2 m_{j-2,i} - \dots - \theta_q m_{j-q,i} & \dots (3.2.2.7) \\ i \prec j - q & j = q + 1, \dots, q + n \end{cases}$$

$$\begin{cases} w_{j;t} = -\theta_1 w_{j-1,t} - \theta_2 w_{j-2,t} - \dots - \theta_q w_{j-q,t} & \dots (3.2.2.8) \\ t = 1, \dots, q & j = q + 1, \dots, q + n \end{cases}$$

Donc l'expression de ε_{n+1} est donnée par:

$$\begin{aligned}
\varepsilon_{n+1} &= X_{n+1} + m_{q+n+1,n} X_n + \dots + m_{q+n+1,1} X_1 + \\
&\quad + w_{q+n+1,1} \varepsilon_{1-q} + \dots + w_{q+n+1,q} \varepsilon_0
\end{aligned}$$

D'où le résultat. :

$$\varepsilon = MX + W \varepsilon^*$$

■

Le resultat suivant permet d'estimer les valeurs initiales ε^* .

Proposition 3.2

Posons $\varepsilon^* = (\varepsilon_{1-q}, \dots, \varepsilon_0)$.

L'estimateur de maximum de vraisemblance $\widehat{\varepsilon}^*$ de ε^* est obtenu à partir de (3.2.2) par les moindres carrés ordinaires :

$$\widehat{\varepsilon}^* = -(W'W)^{-1}W'MX \tag{3.2.3}$$

où M et W sont définis précédemment.

Preuve: Pour les notations de la preuve voir

(Section 2.1.1 du chap.2 (théorème de Gauss Markov)).

Soit

$$Y = Z\theta + V \quad \text{où } V = (V_t) \text{ et } V_t \hookrightarrow N(0, \sigma^2)$$

on a
$$\widehat{\theta}_{OLS} = (Z'Z)^{-1}Z'Y$$

or

$$\varepsilon = MX + W\varepsilon^* \quad \text{où } \varepsilon_t \hookrightarrow N(0, \sigma^2)$$

qui est équivalent à

$$-MX = W\varepsilon^* - \varepsilon$$

$$\widehat{\varepsilon}^* = (W'W)^{-1}W'(-MX)$$

$$= -(W'W)^{-1}W'MX$$

■

Corollaire 3.1

Soit $X_t = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$ où $\varepsilon_t \hookrightarrow N(0, \sigma^2)$ iid, et $\varepsilon_{1-q}, \dots, \varepsilon_0$ sont supposés des variables aléatoires.

Soit $K = \sigma^2 V$ la matrice de covariance de $X = (X_1, \dots, X_n)$,

où V une matrice $(n \times n)$.

Alors la fonction de log-vraisemblance (globale) est :

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |V| - \frac{1}{2\sigma^2} X'V^{-1}X \quad (3.1.8)$$

Preuve: On a

$$\varepsilon = MX + W\varepsilon^*$$

ce ci entraine

$$\varepsilon - W\varepsilon^* = MX$$

D'où

$$M^{-1}(\varepsilon - W\varepsilon^*) = X.$$

or $(\varepsilon - W\varepsilon^*)$ est gaussienne

Donc $X = (X_1, \dots, X_n)$ est un vecteur gaussien de densité:

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}(\det K)^{1/2}} e^{-\frac{1}{2}\langle X, K^{-1}X \rangle}$$

On a $|\sigma^2 V| = (\sigma^2)^n |V|$ et la log-vraisemblance est :

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |V| - \frac{1}{2\sigma^2} X'V^{-1}X$$

■

Box et Jenkins ont clarifiés le lien entre (3.1.5) et (3.1.8) dans le theoreme suivant:

Théorème 3.1 Soient W la matrice de dimension $(n+q) \times q$ définie dans la proposition 3.1 et $S(\theta) = \varepsilon'\varepsilon$. Alors:

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |W'W| - \frac{1}{2\sigma^2} S(\theta) \quad (3.2.4)$$

Preuve: On a

$$S(\theta) = \varepsilon'\varepsilon = (MX + W\varepsilon^*)'(MX + W\varepsilon^*)$$

et posons

$$\widehat{S}(\theta) = (MX + W \widehat{\varepsilon}^*)'(MX + W \widehat{\varepsilon}^*)$$

or

$$\begin{aligned} MX + W \varepsilon^* &= MX + W \widehat{\varepsilon}^* - W \widehat{\varepsilon}^* + W \varepsilon^* \\ &= MX + W \widehat{\varepsilon}^* + W(\varepsilon^* - \widehat{\varepsilon}^*) \end{aligned}$$

Donc

$$\begin{aligned} S(\theta) &= (MX + W \widehat{\varepsilon}^* + W(\varepsilon^* - \widehat{\varepsilon}^*))'(MX + W \widehat{\varepsilon}^* + W(\varepsilon^* - \widehat{\varepsilon}^*)) \\ &= [(MX + W \widehat{\varepsilon}^*)' + (\varepsilon^* - \widehat{\varepsilon}^*)'W'] [(MX + W \widehat{\varepsilon}^*) + W(\varepsilon^* - \widehat{\varepsilon}^*)] \\ &= (MX + W \widehat{\varepsilon}^*)'(MX + W \widehat{\varepsilon}^*) + \underbrace{(MX + W \widehat{\varepsilon}^*)'W(\varepsilon^* - \widehat{\varepsilon}^*)}_{=0} + \\ &\quad + \underbrace{(\varepsilon^* - \widehat{\varepsilon}^*)'W'(MX + W \widehat{\varepsilon}^*)}_{=0} + (\varepsilon^* - \widehat{\varepsilon}^*)'W'W(\varepsilon^* - \widehat{\varepsilon}^*) \\ S(\theta) &= \widehat{S}(\theta) + (\varepsilon^* - \widehat{\varepsilon}^*)'W'W(\varepsilon^* - \widehat{\varepsilon}^*) \quad (3.2.4.1) \end{aligned}$$

car $(MX + W \widehat{\varepsilon}^*)$ et $W(\varepsilon^* - \widehat{\varepsilon}^*)$ sont orthogonales

En effet

$$\begin{aligned} (MX + W \widehat{\varepsilon}^*)'W(\varepsilon^* - \widehat{\varepsilon}^*) &= (MX)'W \varepsilon^* - (MX)'W \widehat{\varepsilon}^* + \\ &\quad + (W \widehat{\varepsilon}^*)'W \varepsilon^* - (W \widehat{\varepsilon}^*)'W \widehat{\varepsilon}^* \end{aligned}$$

comme

$$W'W \widehat{\varepsilon}^* = -W'MX \quad \text{de (3.2.3)}$$

$$\Leftrightarrow (W'W \widehat{\varepsilon}^*)' = -(W'MX)'$$

$$\Leftrightarrow (W \widehat{\varepsilon}^*)'W = -(MX)'W$$

$$\Leftrightarrow \begin{cases} (W\widehat{\varepsilon}^*)'W\varepsilon^* = -(MX)'W\varepsilon^* \\ (W\widehat{\varepsilon}^*)'W\widehat{\varepsilon}^* = -(MX)'W\widehat{\varepsilon}^* \end{cases}$$

On obtient donc :

$$(MX + W\widehat{\varepsilon}^*)'W(\varepsilon^* - \widehat{\varepsilon}^*) = 0$$

On a:

$$\begin{cases} f(\varepsilon ; \theta, \sigma) = (2\pi\sigma^2)^{-\frac{(n+q)}{2}} e^{-\frac{\varepsilon'\varepsilon}{2\sigma^2}} \\ \text{et } \varepsilon \text{ est une fonction de } (\varepsilon^*, X) \\ \text{car } \varepsilon = MX + W\varepsilon^* \end{cases}$$

De plus

$$f(x, \varepsilon^*) = f(h^{-1}(x, \varepsilon^*)) |J_{h^{-1}}(x, \varepsilon^*)|$$

avec

$$\begin{cases} (X, \varepsilon^*) = h(\varepsilon) \Leftrightarrow \varepsilon = h^{-1}(X, \varepsilon^*) = MX + W\varepsilon^* \\ |J_{h^{-1}}(x, \varepsilon^*)| = 1 \end{cases}$$

Donc

$$f(x, \varepsilon^*; \theta, \sigma) = (2\pi\sigma^2)^{-\frac{(n+q)}{2}} \underbrace{\left| \frac{D(\varepsilon)}{D(x, \varepsilon^*)} \right|}_{\text{jacobien}=1} e^{-\frac{S(\theta)}{2\sigma^2}} \dots (3.2.4.2)$$

en remplaçant $S(\theta)$ par $\widehat{S}(\theta) + (\varepsilon^* - \widehat{\varepsilon}^*)'W'W(\varepsilon^* - \widehat{\varepsilon}^*)$ voir (3.2.4.1)

(3.2.4.2) \Leftrightarrow

$$f(x, \varepsilon^*; \theta, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} (2\pi\sigma^2)^{-\frac{q}{2}} e^{-\frac{\widehat{S}(\theta)}{2\sigma^2}} e^{-\frac{(\varepsilon^* - \widehat{\varepsilon}^*)'W'W(\varepsilon^* - \widehat{\varepsilon}^*)}{2\sigma^2}} \dots (3.2.4.3)$$

Cependant :

$$f(x, \varepsilon^*; \theta, \sigma) = f(x, ; \theta, \sigma) f(\varepsilon^* \setminus x; \theta, \sigma)$$

$$\text{avec } L(\theta, \sigma) = f(x, ; \theta, \sigma) \text{ de (3.1.8)}$$

$$\begin{aligned}
(3.2.4.3) \Leftrightarrow f(x, \varepsilon^*; \theta, \sigma) &= (2\pi\sigma^2)^{-\frac{n}{2}} (2\pi\sigma^2)^{-\frac{q}{2}} |W'W|^{1/2} |W'W|^{-1/2} \\
&\times \mathbf{e}^{-\frac{\widehat{S}(\theta)}{2\sigma^2}} e^{-\frac{(\varepsilon^* - \widehat{\varepsilon}^*)' W' W (\varepsilon^* - \widehat{\varepsilon}^*)}{2\sigma^2}} \\
&= \left[\underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} |W'W|^{-1/2} \mathbf{e}^{-\frac{\widehat{S}(\theta)}{2\sigma^2}}}_{f(x, ; \theta, \sigma)} \right] \left[\underbrace{(2\pi\sigma^2)^{-\frac{q}{2}} |W'W|^{1/2} e^{-\frac{(\varepsilon^* - \widehat{\varepsilon}^*)' W' W (\varepsilon^* - \widehat{\varepsilon}^*)}{2\sigma^2}}}_{f(\varepsilon^* \setminus x; \theta, \sigma)} \right]
\end{aligned}$$

$$L(\theta, \sigma) = f(x, ; \theta, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} |W'W|^{-1/2} \mathbf{e}^{-\frac{\widehat{S}(\theta)}{2\sigma^2}} \dots (3.2.4.4)$$

Comparons (3.2.4.4) et la fonction de vraisemblance (3.1.8)

$$\begin{cases} |V| = |W'W| \\ \widehat{S}(\theta) = X'V^{-1}X \end{cases}$$

D'où le résultat final :

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |W'W| - \frac{1}{2\sigma^2} S(\theta) \quad (3.2.4)$$

■

Propriétés :

- 1) $\widehat{\sigma^2} = \frac{S(\theta)}{n}$ l'estimateur de maximum de vraisemblance de σ^2 qui est donné par (3.1.6).
- 2) $\ell^c(\theta)$: la fonction de log-vraisemblance concentrée est donnée par

$$\ell^c(\theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \frac{S(\theta)}{n} - \frac{1}{2} \log |W'W| - \frac{n}{2} \quad (3.2.5)$$

Preuve:

1)

$$\begin{aligned}
\frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{S(\theta)}{2\sigma^4} \\
\frac{\partial \ell}{\partial \sigma^2} &= 0 \Rightarrow \widehat{\sigma^2} = \frac{S(\theta)}{n}
\end{aligned}$$

2) En remplaçant σ^2 par $\frac{S(\theta)}{n}$ dans (3.2.4) on obtient (3.2.5)

Remarque 3.2

La maximisation de la fonction de vraisemblance (3.2.5) est équivalente à la minimisation de

$$\ell^*(\theta) = n \log S(\theta) + \log |W'W| \tag{3.2.6}$$

Resumons les résultats obtenus dans la démonstration de la proposition 3.1:

Soient les matrices M et W données par (prop 3.1) :

$$(3.2.2.4) \Leftrightarrow m_{j,i} = 0 \quad j = 1 \dots q \text{ et } i = 1 \dots n \tag{3.2.7}$$

$$(3.2.2.6) \Leftrightarrow \begin{cases} m_{j,i} = 0 & i > j - q \quad j = q + 1, \dots, n + q \\ m_{j,i} = 1 & i = j - q \quad j = q + 1, \dots, n + q \end{cases}$$

$$(3.2.2.7) \Leftrightarrow \begin{cases} m_{j,i} = -\theta_1 m_{j-1,i} - \theta_2 m_{j-2,i} - \dots - \theta_q m_{j-q,i} \\ i < j - q \quad j = q + 1, \dots, q + n \end{cases}$$

Le (j, i) ème élément de W est donné par :

$$(3.2.2.5) \Leftrightarrow \begin{cases} w_{t,t} = 1 & t = 1 \dots q \\ w_{j,t} = 0 & t \neq j \quad t, j = 1 \dots q \end{cases} \tag{3.2.8}$$

$$(3.2.2.8) \Leftrightarrow \begin{cases} w_{j;t} = -\theta_1 w_{j-1,t} - \theta_2 w_{j-2,t} - \dots - \theta_q w_{j-q,t} \\ t = 1, \dots, q \quad j = q + 1, \dots, q + n \end{cases}$$

Utilisons ces résultats pour exprimer chaque colonne de W et de M . Ces matrices peuvent être écrites comme:

3.2.1 Evaluation de la fonction de vraisemblance $\ell^*(\theta)$

De la relation (3.2.6), $\ell^*(\theta)$ peut être évalué comme suit :

D'une part les résidus initiaux $\varepsilon^* = (\varepsilon_{1-q}, \dots, \varepsilon_0)$ sont estimés par la relation

$$\widehat{\varepsilon}^* = -(W'W)^{-1}W'MX$$

où W et M sont les matrices données par $\varepsilon = MX + W \varepsilon^*$ et la relation (3.2.11).

D'autre part les résidus $(\varepsilon_1, \dots, \varepsilon_n)$ sont calculables récursivement par la relation (3.2.1).

Finalement,

$$S(\theta) = \sum_{t=1-q}^n \varepsilon_t^2 \quad (3.2.12)$$

D'où l'évaluation de $\ell^*(\theta)$.

3.3 Difficultés de l'évaluation directe

Pratiquement, le calcul des estimateurs du maximum de vraisemblance sont déterminés par minimisation de (3.2.6) en utilisant un procédé d'optimisation numérique.

Quand les éléments de θ sont sans contraintes, le problème d'identification se pose dès que certaines racines, du polynôme associés au processus moyenne mobile se situent à l'intérieur et d'autres à l'extérieur du cercle unité.

Afin d'illustrer l'origine de ces problèmes, considérons une moyenne mobile de second ordre.

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} = \theta(L)\varepsilon_t$$

$$\text{où } \theta(L) = 1 + \theta_1 L + \theta_2 L^2$$

On utilise (3.1.4)

$$\begin{aligned}
\theta(L) &= \prod_{i=1}^2 (1 - \alpha_i L) \\
&= (1 - \alpha_1 L)(1 - \alpha_2 L) \\
&= 1 - (\alpha_1 + \alpha_2)L + \alpha_1 \alpha_2 L^2 \\
&\Leftrightarrow \begin{cases} \theta_1 = -(\alpha_1 + \alpha_2) \\ \theta_2 = \alpha_1 \alpha_2 \end{cases}
\end{aligned} \tag{3.3.1}$$

où α_1, α_2 sont réels.

La matrice W est donnée par :

$$W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\alpha_1 \alpha_2 & (\alpha_1 + \alpha_2) \\ -\alpha_1 \alpha_2 (\alpha_1 + \alpha_2) & \alpha_1 + \alpha_1 \alpha_2 + \alpha_2^2 \\ \cdot & \cdot \\ \cdot & \cdot \\ -\alpha_1 \alpha_2 (\alpha_1^{n-1} + \alpha_1^{n-2} \alpha_2 \dots & \alpha_1^n + \alpha_1^{n-1} \alpha_2 + \dots \\ \dots + \alpha_1 \alpha_2^{n-2} + \alpha_2^{n-1}) & \dots + \alpha_1 \alpha_2^{n-1} + \alpha_2^n \end{pmatrix} \tag{3.3.3}$$

en effet pour $n = 2$

$$w_{2+q,1} \stackrel{q=2}{=} w_{4,1} = -\theta_1 w_{3,1} - \theta_2 w_{2,1} = \theta_1 \theta_2 = -\alpha_1 \alpha_2 (\alpha_1 + \alpha_2)$$

Suppose qu'elle est vraie pour n . Montrons la relation pour $n + 1$.

$$\begin{aligned}
w_{n+q+1,1} \stackrel{q=2}{=} w_{n+3,1} &= -\theta_1 w_{n+2,1} - \theta_2 w_{n+2,1} \\
&= -\alpha_1 \alpha_2 (\alpha_1 + \alpha_2) (\alpha_1^{n-1} + \alpha_1^{n-2} \alpha_2 + \alpha_1 \alpha_2^{n-2} + \alpha_2^{n-1}) + \\
&\quad + \alpha_1 \alpha_2 (\alpha_1 \alpha_2 (\alpha_1^{n-2} + \alpha_1^{n-1} \alpha_2 + \alpha_1 \alpha_2^{n-3} + \alpha_2^{n-2})) \\
&= -\alpha_1 \alpha_2 (\alpha_1^n + \alpha_1^{n-1} \alpha_2 + \dots + \alpha_1^2 \alpha_2^{n-2} + \alpha_1 \alpha_2^{n-1} + \alpha_1^{n-1} \alpha_2 \\
&\quad + \alpha_1^{n-2} \alpha_2^2 + \dots + \alpha_1 \alpha_2^{n-1} + \alpha_2^n - \alpha_1^{n-1} \alpha_2 - \alpha_1^n \alpha_2^2 - \dots - \\
&\quad - \alpha_1^2 \alpha_2^{n-2} - \alpha_1 \alpha_2^{n-1}) \\
&= -\alpha_1 \alpha_2 (\alpha_1^n + \alpha_1^{n-1} \alpha_2 + \dots + \alpha_1 \alpha_2^{n-1} + \alpha_2^n)
\end{aligned}$$

D'où la formule pour $n + 1$.

.Si par exemple $|\alpha_1| \geq 1$ et $|\alpha_2| < 1$, les éléments de la matrice $(W'W)$ seront dominés par des puissances positifs de α_1 , par conséquent les éléments de $(W'W)^{-1}$ tendent vers l'infini avec n .

Afin d'assurer l'inversibilité du processus et d'éviter le problème d'identification, nous adoptons l'hypothèse que toutes les racines sont à l'intérieur du cercle unité.

3.4 Inversibilité

Proposition 3.3

Soit $(X_t)_t$ satisfaisant une représentation (3.1.1). le processus (X_t) est inversible si les racines du polynôme $Q(\lambda)$ notées $\alpha_i \in \mathbb{C}$, $\forall i \leq q$ sont en module strictement inférieur à 1 où

$$Q(\lambda) = \prod_{i=1}^q (\lambda - \alpha_i) \text{ et } |\alpha_i| < 1$$

Preuve:

On a vu que si les racines de $\theta(\lambda)$ sont en module supérieure à 1 (c'est à dire : $\left| \frac{1}{\alpha_i} \right| > 1$) alors (3.1.1) est inversible où

$$\theta(\lambda) = \prod_{i=1}^q (1 - \alpha_i \lambda)$$

Donc il suffit de remplacer λ par $\frac{1}{\lambda}$ dans le polynôme P ie :

$$\begin{aligned}\theta\left(\frac{1}{\lambda}\right) &= \prod_{i=1}^q \left(1 - \alpha_i \frac{1}{\lambda}\right) = \frac{1}{\lambda^q} \prod_{i=1}^q (\lambda - \alpha_i) \\ &= \frac{1}{\lambda^q} Q(\lambda)\end{aligned}$$

■

*Considérons de nouveau α_i , $i = 1, \dots, q$, les racines de $Q(\lambda)$: elles peuvent être regroupées comme:

$$\begin{aligned}Q(\lambda) &= [(\lambda - \alpha_1)(\lambda - \alpha_2)] [(\lambda - \alpha_3)(\lambda - \alpha_4)] \dots [(\lambda - \alpha_i)(\lambda - \alpha_{i+1})] \dots \\ &\quad \times [(\lambda - \alpha_{q-1})(\lambda - \alpha_q)]\end{aligned}$$

avec q un nombre pair et $i = 1, 3, 5, \dots, q - 1$.

Soit

$$\{(\alpha_1, \alpha_2); (\alpha_3, \alpha_4); \dots (\alpha_i, \alpha_{i+1}); \dots (\alpha_{q-1}, \alpha_q)\}$$

où chaque couple peut être considéré comme les racines de l'équation suivante :

$$(\lambda - \alpha_i)(\lambda - \alpha_{i+1}) = \lambda^2 + a_i \lambda + b_i \tag{3.4.1}$$

pour $i = 1, 3, \dots, q - 1$

$$\begin{cases} a_i = -(\alpha_i + \alpha_{i+1}) \\ b_i = (\alpha_i \alpha_{i+1}) \end{cases}$$

les α_i, α_{i+1} sont réels, ou bien complexes .

Lemme 3.2

Le processus $(X_t)_t$ est inversible si et seulement si a_i et b_i définis par (3.4.1) vérifient

$$|b_i| < 1 \quad \text{et} \quad |a_i| < 1 + b_i; \quad i = 1, 3, 5, \dots, q - 1 \tag{3.4.2}$$

Preuve:

Il suffit de montrer seulement pour (α_1, α_2) et a_1, b_1

$$\begin{aligned} b_1 &= \alpha_1 \alpha_2 \text{ et } |\alpha_1| < 1 \text{ et } |\alpha_2| < 1 \Leftrightarrow |b| < 1 \\ a_1 &= -(\alpha_1 + \alpha_2) \text{ et } \begin{cases} -1 < \alpha_1 < 1 \\ -1 < \alpha_2 < 1 \end{cases} \Leftrightarrow \\ &\Leftrightarrow \begin{cases} 1 + \alpha_1 > 0 \\ 1 + \alpha_2 > 0 \end{cases} \text{ et } \begin{cases} 1 - \alpha_1 > 0 \\ 1 - \alpha_2 > 0 \end{cases} \end{aligned}$$

donc

$$\begin{aligned} (1 + \alpha_1)(1 + \alpha_2) > 0 &\Leftrightarrow 1 + (\alpha_1 + \alpha_2) + \alpha_1 \alpha_2 > 0 \\ &\Leftrightarrow -(\alpha_1 + \alpha_2) < 1 + \alpha_1 \alpha_2 \\ &\Leftrightarrow a_1 < 1 + b_1 \dots \dots \dots (*) \end{aligned}$$

et

$$\begin{aligned} (1 - \alpha_1)(1 - \alpha_2) > 0 &\Leftrightarrow -(\alpha_1 + \alpha_2) + 1 + \alpha_1 \alpha_2 > 0 \\ &\Leftrightarrow -(\alpha_1 + \alpha_2) > -(1 + \alpha_1 \alpha_2) \\ &\Leftrightarrow a_1 > -(1 + b_1) \dots \dots \dots (**) \end{aligned}$$

de (*) et (**): $-(1 + b_1) < a_1 < 1 + b_1 \Leftrightarrow |a_1| < 1 + b_1$ ■

Remarque 3.3

1) Les conditions (3.4.2) sont satisfaites si on définit les a_i et b_i par :

$$b_i = \frac{\delta_i}{1 + |\delta_i|} \tag{3.4.3}$$

$$a_i = \frac{\gamma_i(1 + b_i)}{1 + |\gamma_i|} \tag{3.4.4}$$

où δ_i, γ_i sont des paramètres sans contraintes.

C'est à dire, avec a_i et b_i défini par (3.4.3) et (3.4.4) L'équation (3.4.1) peut être

résolue pour obtenir un couple de racine satisfaisant la condition d'inversibilité.

2) Si l'ordre de la moyenne mobile q est impair, la racine restante (après la prise des couples de racines), est réelle notée d , définie par

$$d = -\frac{\pi}{1 + |\pi|} \quad (3.4.5)$$

où π est un paramètre sans contrainte.

Un procédé par étape serait adopté pour l'estimation des paramètres d'une moyenne mobile $MA(q)$ pour $q > 1$.

3.5 Inférence

Rappelons les propriétés asymptotiques de l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ (voir la section 2.2.1 du chap. 2) :

$$n^{1/2}(\hat{\theta}_n - \theta) \rightarrow N(0, u^{-1})$$

où

$$\begin{aligned} u &= \mathfrak{I}(\theta) = \lim_{n \rightarrow \infty} \mathfrak{I}^n(\theta) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n I_t(\theta) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E\left(\frac{\partial \ell_t(\theta) \partial \ell_t(\theta)}{\partial \theta \partial \theta'}\right) \end{aligned}$$

et $I_t(\theta)$ l'information de Fischer.

Comme:

$$\begin{cases} \ell_t(\theta) \equiv \ell(\theta) \\ E\left(\frac{\partial \ell(\theta) \partial \ell(\theta)}{\partial \theta \partial \theta'}\right) = -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right) \end{cases}$$

Alors

$$u = \lim_{n \rightarrow \infty} -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right) \quad (3.5.1)$$

De plus on a vu aussi que :

$$V^\infty(n^{1/2}(\hat{\theta} - \theta)) = \mathfrak{S}^{-1}(\theta) = u^{-1} = - \left[\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right]^{-1} \quad (3.5.2)$$

D'après (3.2.5) et (3.2.6) **la maximisation** de $\ell(\theta)$ c'est équivalent a **la minimisation** de $\ell^*(\theta)$.On remarque aussi que $\max \ell(\theta) = \min(-\frac{1}{2}\ell^*(\theta))$ alors:

$$V(\hat{\theta}) = - \left[\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right]^{-1} = \left[\frac{1}{2} \frac{\partial^2 \ell^*(\theta)}{\partial \theta \partial \theta'} \right]^{-1}$$

$$V(\hat{\theta}) = 2 \left[\frac{\partial^2 \ell^*(\theta)}{\partial \theta \partial \theta'} \right]^{-1} \quad (3.5.3)$$

Puisque la fonction de vraisemblance $L(\theta)$ a été concentrée par rapport a ε^* , son utilisation dans (3.5.2) ou (3.5.3) est valable seulement quand les valeurs initiaux et les estimations de coefficients sont asymptotiquement indépendants. Par conséquent, si ces formules sont utilisées pour l'inférence, le processus estimé devrait être limité à la région d'inversibilité.

3.6 Exemple

Box et Jenkins[2] etudient l'estimation dans les modèles *ARMA* .

Les deux séries les plus courtes pour les quelles il ya identification à des modèles de moyenne mobile (*MA*) sont *A* et *C* .

Table1

Estimation des modèles pour les séries A et B

Séries	n	Méthode	Model Adapté	$\hat{\sigma}^2$
A	197	BJ	$\nabla X_t = \varepsilon_t - \frac{0.70}{(0.05)} \downarrow \varepsilon_{t-1}$	0.101
		ML	$\nabla X_t = \varepsilon_t - \frac{0.70}{(0.06)} \downarrow \varepsilon_{t-1}$	0.101
C	226	BJ	$\nabla^2 X_t = \varepsilon_t - \frac{0.13}{(0.07)} \downarrow \varepsilon_{t-1} - \frac{0.12}{(0.07)} \downarrow \varepsilon_{t-2}$	0.019
		ML	$\nabla^2 X_t = \varepsilon_t - \frac{0.13}{(0.07)} \downarrow \varepsilon_{t-1} - \frac{0.12}{(0.08)} \downarrow \varepsilon_{t-2}$	0.019

Le tableau1 compare les estimations données par Box-Jenkins[2] à ceux obtenus par le

maximum de vraisemblance.

Les erreurs standards (celles écrites entre parenthèse) sont estimées à partir de (3.5.3) pour des valeurs de maximum de vraisemblance, $\widehat{\sigma^2}$ est l'estimation *ML* de σ^2 , ∇ l'opérateur de différence.

Pour chaque série, dans les deux exemples nous avons obtenu les mêmes estimations. Ce n'est pas surprenant, car la série contenant environ 200 observations.

Les séries de Box-Jenkins sont utiles comme point de référence car ce sont des données réelles. Afin d'obtenir une série plus courte pour des comparaisons, nous avons subdivisé la série *C* (contenant 224 observations) en huit sous-séries de 28 observations, après différenciation d'ordre 2 (ie ∇^2).

L'estimation par le maximum de vraisemblance (*ML*) pour cette série, ne présente aucun problème et donne des racines à l'intérieur de la région d'inversibilité ($\widehat{\alpha}_1 = 0.41$, $\widehat{\alpha}_2 = -0.29$). Ces mêmes résultats sont en principe attendus pour les sous-séries.

Tableau2

Modèles $ARIMA(0, 2, 2)$ estimé pour les sub-series de C

Méthode	Evaluations de Fonction		Estimation Contraintes de Paramètre					
	Sans contr.	Contr.	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}^2$	racines	$ \hat{\alpha} $	
1	ML	42	47	-0.18 (0.34)	-0.16 (0.32)	0.016	0.5,-0.31	
	LS	31	36	-0.12 (0.27)	-0.14 (0.26)	0.016	0.44,-0.32	
2	ML	39	44	0.22 (0.20)	-0.37 (0.23)	0.011	0.51,-0.73	
	LS	73	165	0.28 (0.06)	-0.60 (0.06)	0.011	0.65,-0.92	
3	ML	42	44	0.067 (0.21)	0.56 (0.17)	0.033	-0.34, $\pm 0.67i$	0.75
	LS	32	35	0.064 (0.20)	0.55 (0.21)	0.039	-0.32, $\pm 0.67i$	0.74
4	ML	33	31	-0.59 (0.22)	0.09 (0.22)	0.0069	0.29, $\pm 0.04i$	0.30
	LS	33	33	-0.61 (0.23)	0.10 (0.24)	0.0069	0.31, $\pm 0.09i$	0.32
5	ML	32	38	-0.05 (0.21)	-0.27 (0.21)	0.0094	0.54,-0.50	
	LS	36	38	-0.04 (0.20)	-0.27 (0.20)	0.0094	0.54,-0.50	
6	ML	28	40	-0.39 (0.19)	0.26 (0.24)	0.017	0.20 $\pm 0.47i$	0.51
	LS	31	40	-0.38 (0.19)	0.27 (0.25)	0.017	0.19 $\pm 0.49i$	0.52
7	ML	30	105	-0.98 (0.16)	0.99 (0.25)	0.0031	0.49 $\pm 0.87i$	0.99
	LS	43	137	-0.92 (0.14)	0.80 (0.12)	0.0039	0.46 $\pm 0.77i$	0.89
8	ML	18	25	-0.04 (0.21)	-0.01 (0.29)	0.011	0.10,-0.06	
	LS	23	21	-0.04 (0.22)	-0.01 (0.29)	0.011	0.12,-0.08	

Les résultats des estimations du tableau 2 sont obtenus, par les moindres carrés (pour des conditions initiales nulles) et par le maximum de vraisemblance, et dans les deux cas sous la contrainte de l'appartenance des racines à la région d'inversibilité (Voir la section 3.4).

Pour la comparaison de l'exécution et le coût du calcul une estimation sans contrainte a été également effectuée dans chaque cas.

A l'exception des sous-série 2 et 7, les méthodes d'estimation (ML) et (LS) ont données des résultats comparables pour les six autres sous-series.

Nous remarquons pour la sous-série 2, que l'une des deux racines, obtenues par les moindres carrés est proche de la frontière de la région d'inversibilité, alors que les racines estimées par le maximum de vraisemblance, sont à l'intérieur. Cette situation est inversé dans le cas de sous-série 7 .

Il est possible que la contrainte d'une estimation à la région d'inversibilité a plus de conséquence sur l'évaluation de la fonction, sur tout quand les racines estimées sont plus près de la frontière de la région d'inversibilité.

Il est naturel que le coût du calcul de la fonction de vraisemblance totale soit plus chère que l'évaluation de la somme des carrés.

Le coût supplémentaire (qui dépend de taille de l'échantillon) est en grande partie engagé dans l'estimation des conditions initiales supposées aleatoires.

3.7 Conclusion

L'estimation de maximum de vraisemblance des processus MA n'est pas adapté pour des échantillons de petites tailles.

Une recherche approfondie sur l'exécution relative de la procédure maximum de vraisemblance et moindres carrés exigerait une étude de simulation à grande échelle, que nous n'avons pas essayé d'effectuer ici.

Le procédé de maximum de vraisemblance est efficace dans l'estimation numérique des paramètres proche à la frontière de la région d'inversibilité .

Chapitre 4

Estimateurs Exacte et Conditionnel du Maximum de Vraisemblance d'un MA(q) Vectoriel

4.1 Introduction

Dans ce chapitre, nous considérons un processus vectoriel moyenne mobile d'ordre q donné par :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (4.1.1)$$

Avec ε_t un vecteur $g \times 1$ tel que $\varepsilon_t \hookrightarrow N(0, \Sigma)$ *iid*, et θ_i ($i = 1, \dots, q$) matrices $g \times g$, Σ matrice de covariance sont les paramètres à estimer.

La fonction exacte de vraisemblance pour le processus de moyenne mobile est obtenue dans la section 2, et la procédure approximative est examinée dans la section suivante.

4.2 Fonction de Vraisemblance Exacte

La fonction de vraisemblance exacte pour le modèle vectoriel est obtenu en généralisant l'approche de Box et Jenkins (1970, A7.4) .

A partir de n observations du processus X_t .

Nous définissons le vecteur ε de dimension $g(n+q) \times 1$ comme :

$$\varepsilon = (\varepsilon'_{1-q}, \dots, \varepsilon'_0, \varepsilon'_1, \dots, \varepsilon'_n)' \quad (4.2.1)$$

où (' :est le transposer).

Propriétés .:

1) ε est un vecteur gaussien tq $\varepsilon \hookrightarrow N(0, \Lambda)$

où

$$\Lambda = I \otimes \Sigma \quad (4.2.2)$$

avec I matrice unitaire de dimension $(n+q) \times (n+q)$, et \otimes le produit de KRONIKER.

2) La densité du vecteur ε est donnée par:

$$f(\varepsilon) = (2\pi)^{-\frac{g(n+q)}{2}} |\Lambda|^{-1/2} e^{-\frac{\varepsilon' \Lambda^{-1} \varepsilon}{2}} \quad (4.2.3)$$

preuve:

1) les composantes du vecteur ε sont des v.a gaussienne iid normalement distribuées par hypothèse

($\varepsilon_t \hookrightarrow N(0, \Sigma)$) alors ε est un vecteur gaussien tq $\varepsilon \hookrightarrow N(0, \Lambda)$

or

$$\begin{aligned} E(\varepsilon) &= (E(\varepsilon'_{1-q}), \dots, E(\varepsilon'_0), E(\varepsilon'_1), \dots, E(\varepsilon'_n)) \\ &= (0, \dots, 0) \end{aligned}$$

donc

$$\Lambda = cov(\varepsilon)$$

Comme

$$\varepsilon_t \hookrightarrow N(0, \Sigma) \Rightarrow cov(\varepsilon_t) = \Sigma$$

d'où

$$\Lambda = cov(\varepsilon) = \begin{pmatrix} cov(\varepsilon_{1-q}) & & & & \\ & \cdot & & & 0 \\ & & cov(\varepsilon_t) & & \\ & & & 0 & \\ & & & & cov(\varepsilon_n) \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma & & & & \\ & \Sigma & & & 0 \\ & & \Sigma & & \\ & & & 0 & \cdot \\ & & & & \Sigma \end{pmatrix}$$

Ainsi

$$\Lambda = I \otimes \Sigma$$

2) Comme ε est vecteur gaussien alors la fonction de densité est :

$$f(\varepsilon) = (2\pi)^{-\frac{g(n+q)}{2}} |\Lambda|^{-1/2} e^{-\frac{\varepsilon' \Lambda^{-1} \varepsilon}{2}}$$

Considerons le systeme suivant obtenu à partir de (4.1.1) :

$$\left\{ \begin{array}{l} \varepsilon_{1-q} = \varepsilon_{1-q} \\ \cdot \\ \cdot \\ \varepsilon_0 = \varepsilon_0 \\ \varepsilon_1 = X_1 - \theta_1 \varepsilon_0 - \dots - \theta_q \varepsilon_{1-q} \\ \cdot \\ \varepsilon_n = X_n - \theta_1 \varepsilon_{n-1} - \dots - \theta_q \varepsilon_{n-q} \end{array} \right. \quad 4.2.4$$

Proposition 4.1

Posons

$$\varepsilon = (\varepsilon'_{1-q}, \dots, \varepsilon'_n), \varepsilon^* = (\varepsilon'_{1-q}, \dots, \varepsilon'_0)' \quad , \quad X = (X'_1, \dots, X'_n) \quad (4.2.5)$$

de taille $(n + q)g \times 1$; $qg \times 1$; $ng \times 1$ respectivement.

Alors le système (4.2.4) se réduit à :

$$\varepsilon = MX + K \varepsilon^* \quad (4.2.6)$$

où M et K sont deux matrices données, dépendant seulement de θ et de dimensions respectives $(n + q)g \times ng$; $(n + q) \times qg$. (voir la démonstration pour leurs expressions)

Preuve:

Il suffit de remplacer successivement dans le système (4.2.4), ε_t $t \geq 1$ en fonction de (X_1, \dots, X_{n-1}) et de $(\varepsilon_{1-q}, \dots, \varepsilon_0)$ pour obtenir (4.2.6).

On utilise le résultat (3.2.2) (comme le cas scalaire du chap. 3):

$$\varepsilon_i = M_i X_i + K_i \varepsilon_i^*$$

avec

$$\varepsilon_i = (\varepsilon_{1-q,i}, \dots, \varepsilon_{n,i}) \text{ et } \varepsilon_i^* = (\varepsilon_{1-q,i}, \dots, \varepsilon_{0,i}) \text{ et } X_i = (X_{1,i}, \dots, X_{n,i})$$

et $\dim \varepsilon_i = n + q$, $\dim X_i = n$, $\dim \varepsilon_i^* = q$, $\dim M_i, K_i$ est $(n + q) \times n$,

$(n + q) \times q$ resp

pour $i = 1, \dots, g$ on a le système suivant :

$$\left\{ \begin{array}{l} \varepsilon_1 = M_1 X_1 + K_1 \varepsilon_1^* \\ \vdots \\ \varepsilon_g = M_g X_g + K_g \varepsilon_g^* \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \begin{pmatrix} \varepsilon_{1-q,1} \\ \vdots \\ \varepsilon_{n,1} \end{pmatrix} = M_1 \begin{pmatrix} X_{1,1} \\ \vdots \\ X_{n,1} \end{pmatrix} + K_1 \begin{pmatrix} \varepsilon_{1-q,1} \\ \vdots \\ \varepsilon_{0,1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \varepsilon_{1-q,g} \\ \vdots \\ \varepsilon_{n,g} \end{pmatrix} = M_g \begin{pmatrix} X_{1,g} \\ \vdots \\ X_{n,g} \end{pmatrix} + K_g \begin{pmatrix} \varepsilon_{1-q,g} \\ \vdots \\ \varepsilon_{0,g} \end{pmatrix} \end{array} \right.$$

$$\Leftrightarrow \begin{bmatrix} \begin{pmatrix} \varepsilon_{1-q,1} \\ \cdot \\ \cdot \\ \varepsilon_{1-q,g} \end{pmatrix} \\ \cdot \\ \cdot \\ \begin{pmatrix} \varepsilon_{n,1} \\ \cdot \\ \cdot \\ \varepsilon_{n,g} \end{pmatrix} \end{bmatrix} = M \begin{bmatrix} \begin{pmatrix} X_{1,1} \\ \cdot \\ \cdot \\ X_{1,g} \end{pmatrix} \\ \cdot \\ \cdot \\ \begin{pmatrix} X_{n,1} \\ \cdot \\ \cdot \\ X_{n,g} \end{pmatrix} \end{bmatrix} + K \begin{bmatrix} \begin{pmatrix} \varepsilon_{1-q,1} \\ \cdot \\ \cdot \\ \varepsilon_{1-q,g} \end{pmatrix} \\ \cdot \\ \cdot \\ \begin{pmatrix} \varepsilon_{0,1} \\ \cdot \\ \cdot \\ \varepsilon_{0,g} \end{pmatrix} \end{bmatrix}$$

\Leftrightarrow

$$\varepsilon = MX + K \varepsilon^*$$

■

Lemme 4.1

Si $\varepsilon = MX + K \varepsilon^*$ et $f(\varepsilon) = (2\pi)^{-\frac{g(n+g)}{2}} |\Lambda|^{-1/2} e^{-\frac{\varepsilon' \Lambda^{-1} \varepsilon}{2}}$ Alors :

$$f(x, \varepsilon^*) = (2\pi)^{-\frac{g(n+g)}{2}} |\Lambda|^{-1/2} e^{-\frac{(MX+K \varepsilon^*)' \Lambda^{-1} (MX+K \varepsilon^*)}{2}} \quad (4.2.7)$$

Preuve:

Remplaçons ε par $MX + K \varepsilon^*$ dans (4.2.3) pour obtenir (4.2.7).

En effet

$$f(x, \varepsilon^*) = f(h^{-1}(x, \varepsilon^*)) |J_{h^{-1}}(x, \varepsilon^*)|$$

avec

$$\begin{cases} (X, \varepsilon^*) = h(\varepsilon) \Leftrightarrow \varepsilon = h^{-1}(X, \varepsilon^*) = MX + K \varepsilon^* \\ |J_{h^{-1}}(x, \varepsilon^*)| = 1 \end{cases}$$

donc

$$f(x, \varepsilon^*) = f(Mx + K \varepsilon^*) \times 1$$

d'où le resultat. ■

Proposition 4.2

Posons $\varepsilon^* = (\varepsilon_{1-q}, \dots, \varepsilon_0)$. L'estimateur de maximum de vraisemblance $\widehat{\varepsilon}^*$ de ε^* est obtenu à partir de (4.2.6) par les moindres carrés généralisés :

$$\widehat{\varepsilon}^* = -(K'\Lambda^{-1}K)^{-1}K'\Lambda^{-1}MX \quad (4.2.8)$$

où M et K sont définies précédemment.

Preuve:

Pour les notations (voir la section 2.1.2 du chap. 2)

Soit

$$Y = Z\theta + V \quad \text{où } V \hookrightarrow N(0, \Sigma)$$

on a

$$\widehat{\theta}_{GLS} = (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}Y$$

or

$$\varepsilon = MX + K \varepsilon^* \quad \text{où } \varepsilon \hookrightarrow N(0, \Lambda)$$

qui est équivalent à

$$-MX = K \varepsilon^* - \varepsilon$$

$$\widehat{\varepsilon}^* = (K'\Lambda^{-1}K)^{-1}K'\Lambda^{-1}(-MX)$$

$$\widehat{\varepsilon}^* = -(K'\Lambda^{-1}K)^{-1}K'\Lambda^{-1}MX$$

■

Lemme 4.2

Soit Λ matrice de covariance de ε , Alors $\varepsilon'\Lambda^{-1}\varepsilon$ peut se décomposer comme:

$$\varepsilon'\Lambda^{-1}\varepsilon = (MX + K \widehat{\varepsilon}^*)'\Lambda^{-1}(MX + K \widehat{\varepsilon}^*) + (\varepsilon^* - \widehat{\varepsilon}^*)'K'\Lambda^{-1}K(\varepsilon^* - \widehat{\varepsilon}^*) \quad (4.2.9)$$

Preuve:

De (4.2.6) nous avons

$$\varepsilon = MX + K \varepsilon^*$$

d'où

$$\varepsilon' \Lambda^{-1} \varepsilon = (MX + K \varepsilon^*)' \Lambda^{-1} (MX + K \varepsilon^*)$$

et par suite

$$MX + K \varepsilon^* = MX + K \widehat{\varepsilon}^* - K \widehat{\varepsilon}^* + K \varepsilon^* = MX + K \widehat{\varepsilon}^* + K(\varepsilon^* - \widehat{\varepsilon}^*)$$

donc

$$\begin{aligned} \varepsilon' \Lambda^{-1} \varepsilon &= [MX + K \widehat{\varepsilon}^* + K(\varepsilon^* - \widehat{\varepsilon}^*)]' \Lambda^{-1} [MX + K \widehat{\varepsilon}^* + K(\varepsilon^* - \widehat{\varepsilon}^*)] \\ &= (MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*) + (\varepsilon^* - \widehat{\varepsilon}^*)' K' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*) \\ &\quad + \underbrace{(MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*)}_{=0} + \underbrace{(\varepsilon^* - \widehat{\varepsilon}^*)' K' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*)}_{=0} \end{aligned}$$

car:

$$\begin{aligned} (MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*) &= (MX)' \Lambda^{-1} K \varepsilon^* - (MX)' \Lambda^{-1} K \widehat{\varepsilon}^* \\ &\quad + (K \widehat{\varepsilon}^*)' \Lambda^{-1} K \varepsilon^* - (K \widehat{\varepsilon}^*)' \Lambda^{-1} K \widehat{\varepsilon}^* \end{aligned}$$

de (4.2.8) on a :

$$\begin{aligned} (K' \Lambda^{-1} K) \widehat{\varepsilon}^* &= -K' \Lambda^{-1} MX \\ \Leftrightarrow (K \widehat{\varepsilon}^*)' (K' \Lambda^{-1})' &= -(MX)' (K' \Lambda^{-1})' \\ \Leftrightarrow (K \widehat{\varepsilon}^*)' \Lambda^{-1} K &= -(MX)' \Lambda^{-1} K \quad \text{car } (\Lambda^{-1} = \Lambda^{-1'}) \end{aligned}$$

Comme Λ^{-1} : diagonalisable

$$\Leftrightarrow \begin{cases} (K \widehat{\varepsilon}^*)' \Lambda^{-1} K \varepsilon^* = -(MX)' \Lambda^{-1} K \varepsilon^* \\ (K \widehat{\varepsilon}^*)' \Lambda^{-1} K \widehat{\varepsilon}^* = -(MX)' \Lambda^{-1} K \widehat{\varepsilon}^* \end{cases}$$

de ce système on obtient

$$(MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*) = 0$$

d'où le resultat finale. ■

Théorème 4.1

La densité de vecteur ε^ est donnée par:*

$$f(\varepsilon^* \setminus x) = (2\pi)^{-\frac{gq}{2}} |K' \Lambda^{-1} K|^{1/2} e^{-\frac{1}{2}(\varepsilon^* - \widehat{\varepsilon}^*)' K' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*)} \quad (4.2.10)$$

et la fonction de vraisemblance exacte est donnée par:

$$f(x) = (2\pi)^{-\frac{gn}{2}} |\Lambda|^{-1/2} |K' \Lambda^{-1} K|^{-1/2} e^{-\frac{1}{2}(MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*)} \quad (4.2.11)$$

Preuve:

La densité conditionnelle s'écrit comme suit :

$$f(x, \varepsilon^*) = f(x) f(\varepsilon^* \setminus x) \quad (4.2.12)$$

De la forme (4.2.7) et (4.2.9) on obtient:

$$\begin{aligned} f(x, \varepsilon^*) &= (2\pi)^{-\frac{gn}{2}} (2\pi)^{-\frac{gq}{2}} |\Lambda|^{-1/2} |K' \Lambda^{-1} K|^{1/2} |K' \Lambda^{-1} K|^{-1/2} \\ &\quad \times e^{-\frac{1}{2}(MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*)} \times e^{-\frac{1}{2}(\varepsilon^* - \widehat{\varepsilon}^*)' K' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*)} \\ &= \underbrace{(2\pi)^{-\frac{gn}{2}} |\Lambda|^{-1/2} |K' \Lambda^{-1} K|^{-1/2} e^{-\frac{1}{2}(MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*)}}_{\mathbf{f}(\mathbf{x})} \\ &\quad \times \underbrace{(2\pi)^{-\frac{gq}{2}} |K' \Lambda^{-1} K|^{1/2} e^{-\frac{1}{2}(\varepsilon^* - \widehat{\varepsilon}^*)' K' \Lambda^{-1} K (\varepsilon^* - \widehat{\varepsilon}^*)}}_{\mathbf{f}(\varepsilon^* \setminus \mathbf{x})} \end{aligned}$$

■

4.3 Approximation de la fonction de vraisemblance

Sous les conditions initiales $\varepsilon^* = (\varepsilon'_{1-q}, \dots, \varepsilon'_0) = 0$, La fonction de vraisemblance du vecteur $\varepsilon = (\varepsilon'_1, \dots, \varepsilon'_n)$ est définie par:

$$L(x, \varepsilon^*) = f(\varepsilon) = f(\varepsilon'_1, \dots, \varepsilon'_n)$$

d'où

$$L(x, \varepsilon^*) = f(x \setminus \varepsilon^* = 0) = \prod_{t=1}^n f(\varepsilon'_t)$$

avec ε_t un vecteur gaussien centré de dimension $g \times 1$ et de matrice de covariance Σ .

Ainsi

$$f(\varepsilon_t) = (2\pi)^{-\frac{g}{2}} |\Sigma|^{-1/2} e^{-\frac{\varepsilon'_t \Sigma^{-1} \varepsilon_t}{2}}$$

ce qui entraîne que

$$L(x, \varepsilon^*) = (2\pi)^{-\frac{ng}{2}} |\Sigma|^{-n/2} e^{-\frac{1}{2} \sum_{t=1}^n \varepsilon'_t \Sigma^{-1} \varepsilon_t} \quad (4.3.1)$$

Définition 4.1

$L(x, \varepsilon^*)$ est dite vraisemblance conditionnelle.

Remarque 4.1

Le facteur de l'exponentielle de la fonction de vraisemblance exacte (4.2.11) est égal à :

$$-\frac{1}{2} (MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*) = -\frac{1}{2} \widehat{\varepsilon}' \Lambda^{-1} \widehat{\varepsilon} = -\frac{1}{2} \sum_{t=1-q}^n \widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t \quad (4.3.2)$$

Le facteur équivalent de la fonction de vraisemblance conditionnelle est considéré comme son approximatif, ceci pour deux raisons: la première c'est d'avoir considéré les conditions initiales $\varepsilon^* = (\varepsilon'_{1-q}, \dots, \varepsilon'_0)$ nulles et la seconde l'utilisation de $|\Sigma|^{-n/2}$ au lieu de l'expression exacte

$$|\Lambda|^{-1/2} |K' \Lambda^{-1} K|^{-1/2} = |\Sigma|^{-\frac{1}{2}n+q} |K' (I \otimes \Sigma^{-1}) K|^{-1/2}$$

Ces approximations sont examinées par la suite composante par composante.

Remarque 4.2

1/ Concernant l'identification d'un processus moyenne mobile $MA(q)$, nous adoptons la terminologie de Box-Jenkins, q'un processus moyenne mobile est inversible si toute les racines de

$$\theta(z) = I + \theta_1 z + \dots + \theta_q z^q \tag{4.3.3}$$

sont situées en dehors de cercle unité.

2/Ces conditions sont nécessaires, pour justifier l'approximation de (4.2.11) par (4.3.1)

Nous retournons maintenant pour examiner les approximations implicites par l'utilisation de (4.3.1). L'effet de ε^* sur ε dans (4.2.6) est déterminé par la matrice K , d'où la propriété suivante:

Propriété

Chaque bloc de colonne de la matrice K (définie par la proposition 4.1) vérifié les relations de récurrence suivantes :

$$\begin{aligned} K_{i+q,j} &= -\theta_1 K_{i+q-1,j} - \theta_2 K_{i+q-2,j} - \dots - \theta_q K_{i,j} \\ i &= 1, \dots, n \quad j = 1, \dots, q \end{aligned} \tag{4.3.5}$$

$$\text{et } K_{i,j} = \delta_{ij} I \quad i, j = 1, \dots, q \tag{4.3.6}$$

Preuve:

On utilise le procédé du cas scalaire pour obtenir la relation de récurrence dans le cas vectoriel.

On a

$$\begin{aligned} w_{i+q,j} &= -\theta_1 w_{i+q-1,j} - \theta_2 w_{i+q-2,j} - \dots - \theta_q w_{i,j} \\ i &= 1, \dots, n \quad j = 1, \dots, q \end{aligned}$$

Dans ce cas les vecteurs ε_t sont de dimension $g \times 1$, donc les produits des éléments de K ($K_{i,j}$) par ε_t sont des matrices de dimension $g \times g$, alors il suffit de remplacer $w_{i,j}$ par $K_{i,j}$.

Ainsi

$$\begin{aligned}
w_{i,i} &= 1 \quad i = 1, \dots, q \\
w_{i,j} &= 0 \quad i, j = 1, \dots, q \quad i \neq j
\end{aligned}$$

equivalent à $K_{i,j} = \delta_{ij}I$, et par consequent

$$\begin{aligned}
K_{i+q,j} &= -\theta_1 K_{i+q-1,j} - \theta_2 K_{i+q-2,j} - \dots - \theta_q K_{i,j} \\
i &= 1, \dots, n \quad j = 1, \dots, q
\end{aligned}$$

Proposition 4.3

Considérons le terme $|K'(I \otimes \Sigma^{-1})K|$ de (4.2.12) alors le (i, j) éme bloc de $K'(I \otimes \Sigma^{-1})K$ est donné par:

$$\sum_{r=1}^{n+q} K'_{r,i} \Sigma^{-1} K_{r,j}$$

de plus

$$\sum_{r=1}^{n+q} K'_{r,i} \Sigma^{-1} K_{r,j} = \delta_{ij} \Sigma^{-1} + K'_{q+1,i} \Sigma^{-1} K_{q+1,j} + K'_{q+2,i} \Sigma^{-1} K_{q+2,j} + \dots + K'_{n+q,i} \Sigma^{-1} K_{n+q,j} \quad (4.3.7)$$

Preuve:

$$K = \begin{pmatrix} K_{11} & K_{12} & \dots & K_{1i} & \dots & K_{1j} & \dots & K_{1q} \\ \cdot & \cdot & \dots & \cdot & \dots & \dots & \dots & \cdot \\ K_{r1} & K_{r2} & \dots & K_{ri} & \dots & K_{rj} & \dots & K_{rq} \\ \cdot & \cdot & \dots & \dots & \dots & \dots & \dots & \cdot \\ K_{n+q,1} & K_{n+q,2} & \dots & K_{n+q,i} & \dots & K_{n+q,j} & \dots & K_{n+q,q} \end{pmatrix}$$

$$K' = \begin{pmatrix} K'_{11} & \cdot & \dots & K'_{r1} & \dots & \cdot & \dots & K'_{n+q,1} \\ \cdot & \cdot & \dots & \cdot & \dots & \dots & \dots & \cdot \\ K'_{1i} & \dots & K'_{ri} & \dots & \dots & \dots & \dots & K'_{n+q,i} \\ \cdot & \dots & \dots & \dots & \dots & \dots & \dots & \cdot \\ K'_{1j} & \dots & K'_{rj} & \dots & \dots & \dots & \dots & K'_{n+q,j} \\ \cdot & \dots & \cdot & \dots & \dots & \dots & \dots & \cdot \\ K'_{1q} & \dots & K'_{rq} & \dots & \dots & \dots & \dots & K'_{n+q,q} \end{pmatrix}$$

et

$$K'(I \otimes \Sigma^{-1}) = \begin{pmatrix} K'_{11}\Sigma^{-1} & \cdot & \dots & K'_{r1}\Sigma^{-1} & \dots & \cdot & \dots & K'_{n+q,1}\Sigma^{-1} \\ \cdot & \cdot & \dots & \cdot & \dots & \dots & \dots & \cdot \\ K'_{1i}\Sigma^{-1} & \dots & K'_{ri}\Sigma^{-1} & \dots & \dots & \dots & \dots & K'_{n+q,i}\Sigma^{-1} \\ \cdot & \dots & \dots & \dots & \dots & \dots & \dots & \cdot \\ K'_{1j}\Sigma^{-1} & \dots & K'_{rj}\Sigma^{-1} & \dots & \dots & \dots & \dots & K'_{n+q,j}\Sigma^{-1} \\ \cdot & \dots & \cdot & \dots & \dots & \dots & \dots & \cdot \\ K'_{1q}\Sigma^{-1} & \dots & K'_{rq}\Sigma^{-1} & \dots & \dots & \dots & \dots & K'_{n+q,q}\Sigma^{-1} \end{pmatrix}$$

Donc le (i, j) ème bloc de $K'(I \otimes \Sigma^{-1})K$ est :

$$\begin{aligned} & K'_{1i}\Sigma^{-1}K_{1j} + K'_{2i}\Sigma^{-1}K_{2j} + \dots + K'_{r,i}\Sigma^{-1}K_{r,j} + \dots + K'_{n+q,i}\Sigma^{-1}K_{n+q,j} \\ &= \sum_{r=1}^{n+q} K'_{r,i}\Sigma^{-1}K_{r,j} \end{aligned}$$

Puisque les q premiers blocs de K forment une matrice d'unit 

$(K_{i,j} = \delta_{ij}I \ i, j = 1, \dots, q)$ Alors: $\sum_{r=1}^{n+q} K'_{r,i} \Sigma^{-1} K_{r,j} = \delta_{ij} \Sigma^{-1} + K'_{q+1,i} \Sigma^{-1} K_{q+1,j} + K'_{q+2,i} \Sigma^{-1} K_{q+2,j} + \dots + K'_{n+q,i} \Sigma^{-1} K_{n+q,j}$ ■

Lemme 4.3

Sous les conditions d'inversibilité du processus MA(q), nous avons

$$K_{i+q,j} \rightarrow 0 \quad \text{qd } i \rightarrow \infty$$

et

$$K'_{t+q,j} \Sigma^{-1} K_{t+q,j} \rightarrow 0 \quad \text{si } t \rightarrow \infty \tag{4.3.8}$$

et les conditions initiales ont un effet asymptotiquement négligeable.

Preuve:

Si toutes les racines de (4.3.4) se situent en dehors du cercle unité alors les éléments de la matrice $K_{i+q,j}$, ont des puissances en i et en module inférieur à 1 donc $K_{i+q,j} \rightarrow 0$ quand $i \rightarrow \infty$.

Comme $\hat{\varepsilon}^* = -(K' \Lambda^{-1} K)^{-1} K' \Lambda^{-1} M X$ et $K_{i+q,j} \rightarrow 0$ alors $\varepsilon^* \rightarrow 0$. ■

Remarque 4.3

Les resultats précédents montrent que seul l'element $\delta_{ij} \Sigma^{-1}$ intervient asymptotiquement dans l'expression de l'element (i, j) de la matrice $K'(I \otimes \Sigma^{-1})K$, ce qui justifier notre approximation.

4.4 Evaluation de la Vraisemblance

La Maximisation de la fonction de vraisemblance exacte (4.2.11), pour estimer les paramètres θ_i ($i = 1, \dots, q$) et Σ , par la méthode du maximum de vraisemblance, est équivalente à la minimisation de la fonction L^* , et est donnée par la proposition suivante:

Proposition 4.4

$$L^* = (n + q) \log |\Sigma| + \log |K'(I \otimes \Sigma^{-1})K| + \sum_{t=1-q}^n \widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t \quad (4.4.1)$$

Preuve:

$$\begin{aligned} L &= \log f(x) = -\frac{gn}{2} \log(2\pi) - \frac{1}{2} \log |\Lambda| - \frac{1}{2} \log |K' \Lambda^{-1} K| \\ &\quad - \frac{1}{2} (MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*) \end{aligned}$$

comme

$$\left\{ \begin{array}{l} |\Lambda|^{-1/2} = |\Sigma|^{-\frac{n+q}{2}} \quad \text{et} \quad \Lambda = I \otimes \Sigma \\ (4.3.2) \Leftrightarrow (MX + K \widehat{\varepsilon}^*)' \Lambda^{-1} (MX + K \widehat{\varepsilon}^*) \\ \quad \quad \quad = \sum_{t=1-q}^n \widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t \end{array} \right.$$

donc

$$\begin{aligned} L &= -\frac{gn}{2} \log(2\pi) - \frac{1}{2} (\log |\Sigma|^{n+q} + \log |K'(I \otimes \Sigma^{-1})K| + \sum_{t=1-q}^n \widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t) \\ &= -\frac{gn}{2} \log(2\pi) - \frac{1}{2} ((n + q) \log |\Sigma| + \log |K'(I \otimes \Sigma^{-1})K| + \sum_{t=1-q}^n \widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t) \end{aligned}$$

■

Proposition 4.5

Sous les conditions $\varepsilon^ = 0$, alors le i éme bloc de $K' \Lambda^{-1} MX$ est donné par:*

$$\sum_{t=1}^n K'_{t+q,i} \Sigma^{-1} \widehat{\varepsilon}_{0t} \quad (4.4.2)$$

où $\widehat{\varepsilon}_{0t}$ le vecteur de résidues pour le temps t calculé avec $\varepsilon^* = 0$.

Preuve:

De (4.2.6): $\widehat{\varepsilon}_0 = MW$ quand $\varepsilon^* = 0$

de plus on a vu que le (i, j) eme bloc de $K'\Omega^{-1}K$ est:

$$\sum_{r=1}^{n+q} K'_{r,i} \Sigma^{-1} K_{r,j} = \sum_{t=1}^n K'_{t+q,i} \Sigma^{-1} K_{t+q,j} \quad \text{d'après (4.3.7)}$$

les q premier bloc sont nul car $\varepsilon^* = 0$ (c'est à dire les q premier de la somme de (4.3.7)) et en remplaçant $K_{t+q,i}$ par les elements de $MX = \widehat{\varepsilon}_0$ on obtient :

$$\sum_{t=1}^n K'_{t+q,i} \Sigma^{-1} \widehat{\varepsilon}_0$$

■

Remarque 4.4

L'évaluation de la fonction de vraisemblance L^ se fait en deux etapes .*

D'abord , on détermine les matrices $K'\Lambda^{-1}MX$ et $K'\Lambda^{-1}K$, ceci permet d'estimer les conditions initiales par la relation $\widehat{\varepsilon}^ = -(K'\Lambda^{-1}K)^{-1}K'\Lambda^{-1}MX$.*

Ensuite, on calcule $\widehat{\varepsilon}_t$ et la matrice $\widehat{\varepsilon}'_t \Sigma^{-1} \widehat{\varepsilon}_t$, $t = 1, \dots, n$.

Bien que les matrices M et K soient de grandes tailles, ni l'une ni l'autre n'a besoins d'être stoker entièrement:en effet, la matrice M n'est utilisée que dans le produit MX exigé dans (4.2.8) ou bien $MX = \varepsilon$ quand $\varepsilon^ = 0$.*

Quand à la matrice K , seul les q bloc lignes sont utilisés pour calculer $K_{i+q,j}$ ($j = 1, \dots, q$) de l'équation (4.3.5), ce traitement par étapes de K est commode pour accumuler les produits $K'\Lambda^{-1}MX$ et $K'\Lambda^{-1}K$ en raison de la forme diagonale de Λ^{-1} .

Par rapport à l'utilisation de la fonction approximative (4.3.1) ,l'évaluation de la vraisemblance exacte nécessite des opérations supplémentaires.

Dans les deux cas l'estimation, des parametres θ_i ($i = 1, \dots, q$) et Σ utilise une technique d'optimisation nonlinéaire appropriée.

4.5 Conclusion

Une etude comparative proposé par Tuniliffe Wilson, de l'estimateur exacte du maximum de vraisemblance des parametres d'un processus moyenne mobile, et de son approximation asymp-

toyique, a montré que la procedure exacte est preferable à la seconde que le cout des cacul la rend infaisable dans certains cas.

Annexe1

Le cas du premier ordre fournit une forme d'illustration naturel pour le comportement du l'estimateur de maximum de vraisemblance.

Un processus de moyenne mobile d'ordre un est défini par :

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1} \quad t = 1, \dots, n \quad \varepsilon_t \hookrightarrow N(0, \sigma^2) \quad (*)$$

Nous représentons les deux estimateurs de θ d'un modèle $MA(1)$.

En pratique les estimateurs du maximum de vraisemblance sont trouvés numériquement par les algorithmes itératifs, on cherche le maximum de la fonction de vraisemblance numériquement (par exemple : par la méthode de Newton-Raphson), démarrant d'une valeur initiale pour le vecteur du paramètre recherché. Cette valeur peut être fournie par les estimateurs des moindres carrés, ou bien supposons quel est donnée donc on obtient deux estimateurs

$\hat{\theta}_{C.M.L}$:C'est l'estimateur de max de vraisemblance conditionné à une valeur initiale donnée (dans notre cas la valeur initiale égal à zéro).

$\hat{\theta}_{U.M.L}$:l'estimateur de max de vraisemblance sans condition

Avec

$$\hat{\theta}_{C.M.L} = \min_{\theta} \sum_{t=1}^n \varepsilon_t^2(\theta)$$

telque $\varepsilon_t(\theta)$ sont calculés comme :

$$\begin{aligned} (*) \quad &\Leftrightarrow \quad \varepsilon_t(\theta) = X_t - \theta\varepsilon_{t-1}(\theta) \\ \varepsilon_0 &= 0, \varepsilon_1 = X_1, \dots, \varepsilon_n = . \end{aligned}$$

$$\hat{\theta}_{U.M.L} = \min_{\theta} \sum_{t=1-q}^n \varepsilon_t^2(\theta) \quad \text{pour } q = 1$$

ce resultat est obtenu par 3.2.12

Les statistiques descriptives pour les deux estimateurs du θ appraissent dans les trois tableaux obtenus à l'aide du logiciel *R*, les expériences ont été réaliser pour $\sigma^2 = 1$ et $\theta = 0$ et $\theta = \pm 0.2$ et $\theta = \pm 0.5$ et $\theta = \pm 0.9$, le nombre de répliques était 500 pour $n = 30$, et 200 pour $n = 100$ et 50 pour $n = 300$, et l'espace de paramètre à été confiner à $|\theta| < 1$ parce que les valeurs de paramètres en dehors de la région d'inversibilité sont superflue donc ignorée

Tableau1Statistiques descriptives pour les deux estimateurs du θ , $n=30$ (500réplique)

θ	Estimateur de θ	Moyenne	Variance	Biais	Er.quadratique
-0.9	C.M.L	-0.818	0.016	0.081	0.023
	U.M.L	-0.924	0.012	-0.024	0.013
-0.5	C.M.L	-0.517	0.038	-0.017	0.039
	U.M.L	-0.520	0.042	-0.020	0.042
-0.2	C.M.L	-0.208	0.047	-0.008	0.047
	U.M.L	-0.181	0.050	0.018	0.050
0.0	C.M.L	-0.005	0.048	-0.005	0.048
	U.M.L	0.017	0.050	0.017	0.050
+0.2	C.M.L	0.209	0.043	0.009	0.043
	U.M.L	0.197	0.053	-0.002	0.053
+0.5	C.M.L	0.511	0.040	0.011	0.040
	U.M.L	0.513	0.041	0.013	0.041
+0.9	C.M.L	0.818	0.017	-0.081	0.024
	U.M.L	0.923	0.012	0.023	0.013

Tableau2

Statistiques descriptives pour les deux estimateurs du θ , $n=100$ (200réplique)

θ	Estimateur de θ	Moyenne	Variance	Biais	Er.quadratique
-0.9	C.M.L	-0.870	0.003	0.029	0.003
	U.M.L	-0.904	0.002	-0.004	0.002
-0.5	C.M.L	-0.501	0.008	-0.001	0.008
	U.M.L	-0.514	0.008	-0.014	0.008
-0.2	C.M.L	-0.193	0.010	0.006	0.010
	U.M.L	-0.195	0.010	0.004	0.010
0.0	C.M.L	-0.003	0.008	-0.003	0.008
	U.M.L	-0.007	0.009	-0.007	0.009
+0.2	C.M.L	0.203	0.009	0.003	0.009
	U.M.L	0.209	0.010	0.009	0.010
+0.5	C.M.L	0.500	0.008	0.000	0.008
	U.M.L	0.514	0.009	0.014	0.009
+0.9	C.M.L	0.878	0.003	-0.021	0.003
	U.M.L	0.906	0.002	0.006	0.003

Tableau3

Statistiques descriptives pour les deux estimateurs du θ , $n = 300$ (50réplique)

θ	Estimateur de θ	Moyenne	Variance	Biais	Er.quadratique
-0.9	C.M.L	-0.893	0.0008	0.006	0.0008
	U.M.L	-0.904	0.0005	-0.004	0.0005
-0.5	C.M.L	-0.498	0.001	0.001	0.001
	U.M.L	-0.492	0.002	0.007	0.002
-0.2	C.M.L	-0.199	0.002	0.0001	0.002
	U.M.L	-0.201	0.002	-0.001	0.002
0.0	C.M.L	-0.001	0.003	-0.001	0.003
	U.M.L	0.005	0.002	0.005	0.002
+0.2	C.M.L	0.196	0.002	-0.003	0.002
	U.M.L	0.206	0.002	0.006	0.002
+0.5	C.M.L	0.499	0.002	-0.0003	0.002
	U.M.L	0.501	0.002	0.001	0.002
+0.9	C.M.L	0.890	0.0008	-0.009	0.0008
	U.M.L	0.896	0.0005	-0.003	0.0005

*Nous remarquons que les valeurs estimées du paramètres sont très proches aux valeurs exactes avec des variances et des erreurs très petites.

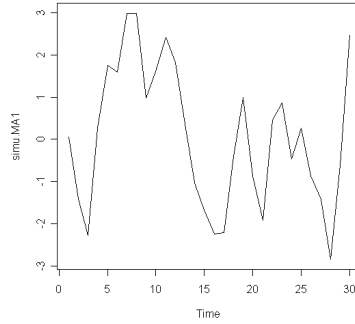
Ces dernières diminues et sont très proche de zéro, chaque fois que n (taille de l'échantillon) augmente .

Ceci confirme l'utilité des resultats théoriques et des propriétés asymptotiques de l'estimateur du maximum de vraisemblance.

Enfin nous presentons quelques trajectoires d'un processus moyenne mobile d'ordre un $MA(1)$ pour des échantillons de tailles différentes.

pour $n=30$ nous obtenons trajectoire suivante

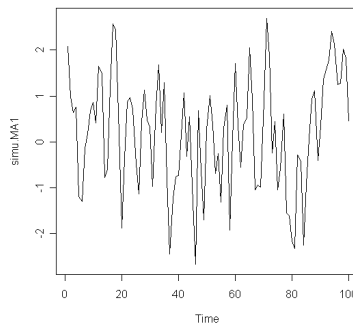
Grappe1



Simu.MA(0.5,1,30,0)

pour $n=100$ nous obtenons trajectoire suivante

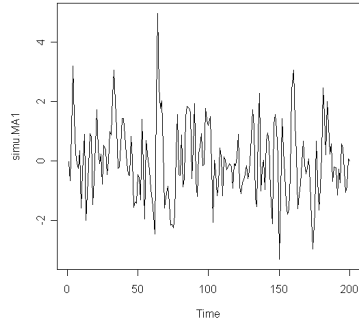
Graph2



Simu.MA(0.5,1,100,0)

pour $n=200$ nous obtenons trajectoire suivante

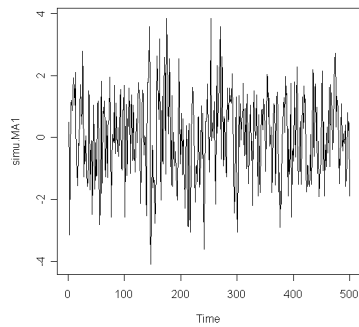
Graphe3



Simu.MA(0.5,1,200,0)

Pour $n=500$ nous obtenons la trajectoire suivante

Graphe4



Simu.MA(0.5,1,500,0)

Annexe2

```
PROGRAMME De  $\hat{\theta}_{C.M.L}$  :  
simu.B1<-function(teta,sigma2,n){  
  simu.ma1<-function(teta,sigma2,n)  
  {  
    epsilon<-rnorm(n,0,sigma2)  
    x<-numeric(n)  
    x[1]<-epsilon[1]+teta  
    for(i in 2:n)  
    {  
      x[i]<-epsilon[i]+teta*epsilon[i-1]  
    }  
    x  
  }  
  simu.ma1(0.2,1.000,30)  
  x<-simu.ma1(0.2,1.000,30)  
  x  
  f<-function(teta)  
  {  
    e<-numeric(30)  
    e[0]<-0  
    e[1]<-x[1]  
    for (i in 2:30)
```

```

{
e[i]<-x[i]-teta*e[i-1]
}
e
res<-sum(e^2)
res
}
nlm(f,0)
tetachapo<-nlm(f,0)
tetachapo1<-as.matrix(tetachapo)
tetachapo1[2,]
b1<-tetachapo1[2,]
b1
B1<-as.numeric(b1)
B1
}
B1<-numeric(500)
for(i in 1:500){
B1[i]<-simu.B1(0.2,1.000,30)
}
mean(B1)
var(B1)
mean(B1)-0.2 #biais
mean((B1-0.2)^2)#erreur quadratique
PROGRAMME De  $\hat{\theta}_{U.M.L}$  :
simu.B1<-function(teta,sigma2,n){
simu.ma1<-function(teta,sigma2,n)
{
epsilon<-rnorm(n,0,sigma2)
x<-numeric(n)

```

```

x[1]<-epsilon[1]+teta
for(i in 2:n)
{
x[i]<-epsilon[i]+teta*epsilon[i-1]
}
x
}
simu.ma1(0.2,1.000,30)
x<-simu.ma1(0.2,1.000,30)
x
f<-function(teta)
{
e<-numeric(30)
e[1]<-x[1]-teta*0.96
for (i in 2:30)
{
e[i]<-x[i]-teta*e[i-1]
}
e
res<-sum(e^2)
res
}
nlm(f,0)
tetachapo<-nlm(f,0)
tetachapo1<-as.matrix(tetachapo)
tetachapo1[2,]
b1<-tetachapo1[2,]
b1
B1<-as.numeric(b1)
B1

```

```
}  
B1<-numeric(500)  
for(i in 1:500){  
  B1[i]<-simu.B1(0.2,1.000,30)  
}  
mean(B1)  
var(B1)  
mean(B1)-0.2 #biais  
mean((B1-0.2)^2)#erreur quadratique
```


Conclusion

Dans ce mémoire, nous avons présenté une méthode d'estimation ML des parametre d'une moyenne mobile d'ordre q $MA(q)$, Elle est due aux auteurs Denise.R.Osborn et CharleR. Nélon, l'estimateur est obtenu en maximisant la fonction de vraisemblance ,En pratique cette maximisation obtenu numeriquement par des algorithmes itératifs.

Des resultats de simulations confirment les resultats theoriques.

Bibliographie

- [1] Ajai S.Gaur/Sanjaya S.Gaur. Statistical methode for practice and research. Response Books.
- [2] Box, G.E.P. and G.M. Jenkins. Time series analysis: forecasting and control (Holden Day, San Francsisco. 1970. revised edition. 1976).
- [3] Brockwell, P. J., & Davis, R. A. (1991). Time Series : Theory and Methods (Second ed.). New York : Springer.
- [4] Brockwell, P. J., & Davis, R. A. (1996). Introduction to Time Series and Forecasting. New York : Springer.
- [5] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity Journal of Econometrics, 31, 307-327.
- [6] Cryer, J. (1986). Time series analysis. Boston : Duxbury Press.
- [7] Chritophe Hurlin. Econométrie appliquée Serie temporelles.UFR Economie.
- [8] Charles R.Nelson(1973).The first-order moving average process.Graduats School of Busines University of Chicago,III,60637,USA.
- [9] Durbin, J. (1959). "Efficient Estimation of Parameters in Moving Average Models ". Biometrika, 46 (1959) p. 306-313.
- [10] DeniseR.Osborn(1976).Maximum likelihood estimation of moving average processes .Annals of Economic and Social Measurement 5/1,1976.

- [11] Denise R. Osborn (1976). Exact and approximate Maximum likelihood estimators for vector moving average processes. National Institute of Economic and Social Research, London.
- [12] Enders, W. (1995). Applied Econometric Time Series. New York : Wiley
- [13] Kendall, M., & Stuart, A. (1976). The Advanced Theory of Statistics (Vol. 3). London : Griffin.
- [14] Michel Sampson (2001). Time series analysis. Loglinear Publishing.
- [15] Monfort, A. (1997). Cours de statistique mathématique (3 ed.). Paris : Economica.
- [16] Pierre-André Cornillon et autres. Statistiques avec R. Pratique de La Statistique Presses Universitaires De Rennes 2008.
- [17] Robert H. Shumway-Cook, David Stoffer (1999). Time series analysis and its applications, Second Edition. Springer Science.