

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri, Tizi-Ouzou

Faculté des Sciences

Département: Mathématiques



Mémoire de Master

en
Mathématiques appliquées

Option: Processus aléatoires et statistique de la décision

Thème:

Prédiction dans les Séries Chronologiques

Présenté par

Hemsas Ouahiba

Devant le jury d'examen composé de:

| | | | |
|-------------------------------------|------------------------|-----------|--------------|
| <i>M^r</i> Mamou | Maître Assistant A | U.M.M.T.O | Président |
| <i>M^r</i> BERKOUN Youcef | Maître de conférence A | U.M.M.T.O | Rapporteur |
| <i>M^m</i> Boualam | Maître Assistant A | U.M.M.T.O | Examinatrice |

Année 2014

Remerciements

*Je tiens à exprimer toute ma reconnaissance à **Mr Berkoun Youcef** pour l'honneur qu'il m'a fait en assurant la direction et le suivi scientifique et technique du présent mémoire. Je le remercie pour sa grande contribution à l'aboutissement de ce travail, et pour sa disponibilité.*

*Je remercie également **Mr Mamou** pour l'honneur qu'il me fait en acceptant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adressent également à **Mm Boualam** pour avoir acceptée d'examiner ce travail.*

*En préambule de ce mémoire, je tiens à remercier mes soeurs **NANA** et sa belle famille, **ZAHRA** et sa belle famille, **DAOUIA** et sa belle famille, **CELIA**, **HAFIDA**, mes frères **GHANI**, **FODIL**, **OUAHAB** .*

*Je remercie tous mes amis **MOHAMMED** , **HAND**, **KHALED**, **WARDA**, **KAHINA**, **LAMIA** ainsi que tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.*

Table des matières

| | |
|---|----------|
| Introduction | 5 |
| 1 Généralités sur les séries chronologiques | 6 |
| 1.1 Séries chronologiques | 6 |
| 1.1.1 Périodicité | 6 |
| 1.2 Classification des séries chronologiques | 7 |
| 1.2.1 Domaines d'application | 7 |
| 1.2.2 Fonction d'autocovariance | 7 |
| 1.2.3 Fonction d'autocorrélation | 8 |
| 1.3 Stationnarité | 9 |
| 1.3.1 Fonction d'autocorrélation partielle | 9 |
| 1.3.2 Fonction d'autocovariance empirique | 10 |
| 1.3.3 Critères d'information | 10 |
| 1.4 Composantes d'une série chronologique | 11 |
| 1.4.1 La tendance (trend) | 11 |
| 1.4.2 Composante saisonnière (saisonnalité) | 11 |
| 1.4.3 bruit blanc | 11 |
| 1.4.4 Composante résiduelle (accidentelle ou irrégulière) | 11 |
| 1.4.5 Composante cyclique | 12 |
| 1.4.6 Perturbations | 12 |
| 1.5 Les modèles | 12 |
| 1.5.1 Modèle additif | 12 |
| 1.5.2 Modèle multiplicatif | 12 |
| 1.5.3 Modèle mixte | 13 |
| 1.6 Elimination de la tendance et saisonnalité | 13 |
| 1.6.1 Elimination de la tendance en l'absence de saisonnalité | 13 |
| 1.6.2 Elimination du trend et de la saisonnalité | 16 |
| 1.6.3 Méthode des différences | 16 |
| 1.6.4 calcul des variations saisonnières | 17 |
| 1.6.5 corrections des variations saisonnières | 17 |
| 1.7 Moyennes mobiles | 17 |
| 1.7.1 Moyenne mobile simple | 17 |
| 1.7.2 Moyenne mobile centrée | 18 |
| 1.8 Mise en œuvre sous R | 19 |
| 1.8.1 Décomposition de la série | 20 |

| | | |
|----------|---|-----------|
| 2 | Séries chronologiques linéaires | 23 |
| 2.1 | Processus autorégressifs AR | 23 |
| 2.2 | Processus moyenne mobile MA | 23 |
| 2.3 | Processus ARMA | 23 |
| 2.4 | Estimation des paramètres | 24 |
| 2.4.1 | Estimation des paramètres autorégressifs (les équations de Yule-Walker) | 24 |
| 2.4.2 | Estimation par maximum de vraisemblance | 26 |
| 2.4.3 | Vraisemblance à la Box-Jenkins | 30 |
| 2.4.4 | Quelques résultats asymptotiques | 31 |
| 2.5 | Test de stationnarité à une tendance déterministe près | 31 |
| 2.6 | Validation d'un modèle ARMA | 32 |
| 2.7 | Mise en œuvre sous R | 33 |
| 2.7.1 | Simulation des séries chronologiques | 33 |
| 2.7.2 | Estimation des paramètres | 35 |
| 2.7.3 | Estimation des paramètres d'un processus AR(p) | 35 |
| 2.7.4 | Estimation des paramètres d'un processus ARIMA | 36 |
| 3 | Séries chronologiques linéaires | 39 |
| 3.1 | Géométrie des variables aléatoires | 39 |
| 3.2 | Equations linéaires de prédiction | 41 |
| 3.2.1 | Algorithme de Durbin-Levinson | 42 |
| 3.2.2 | Prédiction d'un processus ARMA | 43 |
| 3.2.3 | La méthode de Box et Jenkins | 43 |
| 3.2.4 | L'erreur de prédiction | 44 |
| 3.3 | Prévision par lissage exponentiel | 45 |
| 3.3.1 | Lissage exponentiel simple | 46 |
| 3.3.2 | Lissage exponentiel double | 48 |
| 3.3.3 | Méthode de Holt-Winters | 50 |
| 3.4 | Mise en œuvre sous R | 52 |
| 3.4.1 | Lissage exponentiel simple | 52 |
| 3.4.2 | Lissage exponentiel double | 54 |
| | Conclusion | 56 |
| | Références | 56 |

Introduction

Depuis plus de trois décennies, l'étude des séries chronologiques a connu un essor considérable, surtout dans le domaine de la finance avec l'apparition de modèles à volatilité stochastique (ARCH, GARCH, IGARCH). L'analyse des séries chronologiques sont des défis scientifiques importants qui trouvent leurs applications dans des domaines aussi variés que la finance, la production électrique, la climatologie, etc ...

L'étude d'une série chronologique commence non seulement par la modélisation, mais aussi à évaluer l'importance d'une saisonnalité, à distinguer si une évolution constatée ne relève pas d'un mouvement structurel ou conjoncturel.

Une fois modélisée, on cherche le plus souvent à prévoir une valeur future à partir de l'existant. Réduire l'incertitude liée à la connaissance du futur, améliorer la qualité de l'information et des décisions qui en découlent sont les principaux objectifs de la prévision. Il existe de nos jours un ensemble de méthodes rigoureuses, basées sur des algorithmes, permettant de faire des prévisions. La qualité de la prévision dépend du choix de la méthode. Dans ce mémoire, on abordera certaines de ces méthodes, qu'on appliquera aux modèles ARMA(p,q). Le mémoire comprend trois chapitres.

Le premier chapitre contient des généralités sur les séries chronologiques. Le second est dédié à l'estimation des paramètres d'un ARMA(p,q). Le troisième est consacré à la prédiction. On termine avec une conclusion générale.

Chapitre 1

Généralités sur les séries chronologiques

1.1 Séries chronologiques

Définition 1. On appelle série chronologique (*série temporelle, ou chronique*) une suite d'observations numériques d'une grandeur effectuées à intervalles réguliers au cours du temps. A titre d'exemple: l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel.

Définition : soient (Ω, A) et (E, F) deux espaces probabilisables, Une série chronologique est une application définie par :

$$X : \Omega \times T \longrightarrow E$$

$$(w, t) \longrightarrow X(t, w) = X_t(w).$$

où T est un espace de temps et E est un espace de valeurs.

1.1.1 Périodicité

La périodicité d'une série chronologique est la durée qui sépare deux observations. La fréquence des observations peut être journalière, hebdomadaire, mensuelle, annuelle. En économie, un effet saisonnier est lié à une période connue. Une série chronologique journalière sera observée pendant plusieurs semaines avec une périodicité de 5, 6 ou 7 jours selon le cas. Une chronique mensuelle sera observée sur plusieurs années.

1.2 Classification des séries chronologiques

On peut classer les séries chronologiques selon des critères variés: domaines d'application, séries réelles, complexe, séries stationnaires ou non stationnaires (avec tendances, avec facteurs saisonniers).

1.2.1 Domaines d'application

Le domaine d'application pour des séries chronologiques est très vaste. En effet, nous pouvons classer tout phénomène évoluant dans le temps comme domaine d'application pour ces séries. Nous citons par exemple:

- *Finance et économétrie*: évolutions des indices boursiers, des prix, des données économiques des entreprises.
- *Médecine*: analyse d'électro-encéphalogrammes et d'électrocardiogrammes.
- *Traitement du signal*: signaux de communications, de radars, de sonars, analyse de la parole.

1.2.2 Fonction d'autocovariance

Définition 2. (Fonction d'autocovariance) L'autocovariance d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie par:

$$\begin{aligned} \rho &: \mathbb{Z} \longrightarrow \mathbb{R} \\ h &\longrightarrow \rho(h) = \text{Cov}(X_t, X_{t-h}) \end{aligned}$$

Proposition 1.

1. $\rho(-h) = \rho(h), \forall h$, (ρ est une fonction paire)
2. $\forall n \in \mathbb{N}, \forall a \in \mathbb{R}^N, \forall t \in \mathbb{Z}^N, \sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho(t_i - t_j) > 0$, (ρ est de type positif).
avec $a = (a_1, a_2, \dots, a_n)$ et $t = (t_1, t_2, \dots, t_n)$

Démonstration 1.

1. La parité:

$$\begin{aligned} \rho(h) &= \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t-h}, X_{(t-h)+h}) = \text{Cov}(X_{t-h}, X_t) \\ &= \text{Cov}(X_t, X_{t-h}) = \rho(-h). \end{aligned}$$

2. La Positivité:

$$\begin{aligned} \text{Var}\left(\sum a_i X_{t_i}\right) &= \text{Cov}\left(\sum_i a_i X_{t_i}, \sum_j a_j X_{t_j}\right) \\ &= \sum_{j,i} a_i a_j \text{Cov}(X_{t_i}, X_{t_j}) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho(t_i - t_j) > 0. \end{aligned}$$

1.2.3 Fonction d'autocorrélation

Définition 3. On appelle **fonction d'autocorrélation notée ACF**, la fonction ρ définie de \mathbb{Z} dans \mathbb{R} par:

$$\forall h \in \mathbb{Z}, \gamma(h) = \frac{\rho(h)}{\rho(0)} \tag{1.1}$$

Proposition 2. La fonction d'autocorrélation ρ vérifie:

1. $\forall h \in \mathbb{Z}, \gamma(-h) = \gamma(h)$ (γ elle est paire).
2. $\gamma(0) = 1$.
3. $|\gamma(h)| \leq 1, \forall h$.

ACF d'un AR(1)

$$\begin{aligned} X_t &= 0,9X_{t-1} + a_t \\ \text{ACF}(1) &= (\text{corr}(X_{t+1}), X_t) \\ &= \text{corr}(0,9X_t + a_{t+1}, X_t) \\ &= 0,9\text{corr}(X_t, X_t) + \text{corr}(a_{t+1}, X_t) \\ &= 0,9 + 0 = 0,9 \end{aligned}$$

car X_t et a_{t+1} sont indépendants

$$\begin{aligned} \text{ACF}(2) &= (\text{corr}(X_{t+2}), X_t) \\ &= \text{corr}(0,9X_{t-1} + a_t, X_t) \\ &= \text{corr}(0,9 \times 0,9X_{t-2} + a_{t-1} + a_t, X_t) \\ &= 0,81\text{corr}(X_{t-2} + a_{t-1} + a_t, X_t) \end{aligned}$$

$$\begin{aligned}
&= \text{corr}(X_{t-2}, X_t) + \text{corr}(a_{t-1}, X_t) + \text{corr}(a_t, X_t) \\
&= \text{corr}(X_{t-2}, X_t)
\end{aligned}$$

car X_{t-2} indépendant avec a_t et a_{t-1}

$$\begin{aligned}
ACF(p) &= (\text{corr}(X_t), X_{t-k}) = 0,9^k \\
ACF \text{ d'un } MA(q) &= 0 \quad \text{pour } k > q
\end{aligned}$$

1.3 Stationnarité

Définition 1.3.1. (Stationnarité stricte ou forte) $(X_t)_{t \in T}$ est un processus stationnaire au sens strict si: $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n), \forall h \in \mathbb{Z}$, la loi de $(X_{t_1}, \dots, X_{t_n})$ est identique à la loi de $(X_{t_1+h}, \dots, X_{t_n+h})$.

Théorème 1. (Théorème de Kolmogorov) $(X_t)_{t \in T}$ est un processus stationnaire au sens strict si et seulement si:

la loi de $(X_t)_{t \in T}$ est identique à la loi de $(Y_t)_{t \in T}$ où $Y_t = X_{t+h}$.

Définition 1.3.2. Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est stationnaire du second ordre (ou faiblement stationnaire) si et seulement si:

- (i) $\forall t \in \mathbb{Z}, E(X_t) = \mu$.
- (ii) $\forall t \in \mathbb{Z}, E(X_t) < \infty$.
- (iii) $\forall t, h \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h}) = \gamma(h)$ ($\gamma(h)$ ne dépend que de h).

Remarque 1. pour un processus, la stationnarité stricte implique la stationnarité faible, la réciproque est fausse.

1.3.1 Fonction d'autocorrélation partielle

Considérons une série stationnaire X_t et ses régressions linéaires sur son passé résumé à une, deux, trois... observations:

$$\begin{aligned}
X_t &= \phi_{1,1}X_{t-1} + U_1t \\
X_t &= \phi_{1,2}X_{t-1} + \phi_{2,2}X_{t-2} + U_2t \\
X_t &= \phi_{1,3}X_{t-1} + \phi_{2,3}X_{t-2} + \phi_{3,3}X_{t-3} + U_3t \\
&\vdots
\end{aligned}$$

Considérons les $\phi_{k,k}, k=1,2,\dots$. Ils ont la même interprétation que les coefficients d'une régression linéaire classique: $\phi_{k,k}$ représente l'apport d'explication de X_{t-k} à X_t , toutes choses égales par ailleurs, c'est-à-dire étant donné qu'on régresse également sur $X_{t-1}, \dots, X_{t-k+1}$. Ils forment la fonction d'autocorrélation partielle que nous abrègerons en PACF.

PACF d'un AR(1)

$$\begin{aligned}
 PACF(1) &= (corr(X_{t+1}), X_t) + 0,9 \\
 PACF(2) &= (corr(X_{t+2} - 0,9X_{t+1}), X_t - 0,9X_{t+1}) \\
 &= corr(a_{t+2}, X_t) - 0,9corr(a_{t+2}, X_{t+1}) \\
 PACF(2) &= 0 \\
 PACF(k) &= 0 \quad \text{pour } k > p \\
 PACF \text{ d'un } AR(p) &= 0 \quad \text{pour } k > p
 \end{aligned}$$

D'une façon générale on a:

- La PACF d'un AR(p) est nulle à partir de l'ordre p+1.
- La PACF d'un processus qui a une composante moyenne mobile à une décroissance exponentielle.
- Ainsi la PACF d'un ARMA(p,q), q > 0 présente une décroissance exponentielle.

1.3.2 Fonction d'autocovariance empirique

A partir des observations $\{x_1, \dots, x_n\}$ d'un processus stationnaire (X_t) , nous aurons souvent besoin d'estimer la fonction d'autocorrélation $\rho(h)$ du processus (X_t)

Définition 4. La fonction d'autocovariance empirique de $\{x_1, \dots, x_n\}$ est définie par:

$$\hat{\rho}_n(h) = n^{-1} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}) \quad 0 < h < n$$

et $\hat{\rho}_n(h) = \hat{\rho}_n(-h)$ lorsque $n < h < 0$

\bar{x} étant la moyenne empirique des x_i avec

$$\bar{x} = n^{-1} \sum_{j=1}^n x_j$$

Définition 5. La fonction d'autocorrélation empirique est définie par:

$$\hat{\gamma}(h) = \frac{\hat{\rho}_n(h)}{\hat{\rho}_n(0)} \quad h < n$$

1.3.3 Critères d'information

Le critère AIC (Akaike information criterion)

C'est un critère de choix de modèle et qui est défini par:

$$AIC(p, q) = \log \sigma^2 + 2 \frac{p+q}{T}$$

Le critère BIC (Bayesian information criterion)

C'est un critère de sélection de modèle, il consiste à mettre en balance la qualité de la représentation de la série par la tendance avec le nombre de paramètres de la tendance, et qui est défini par:

$$BIC(p,q) = \log \sigma^2 + [p + q] \frac{\log T}{T}$$

Les modèles ayant la plus petite valeur de critère devront être choisis.

Ces deux critères conduisent donc à sélectionner des modèles dont la vraisemblance est grande.

1.4 Composantes d'une série chronologique

1.4.1 La tendance (trend)

La tendance m_t , représente l'évolution à long terme de la série X_t . Elle traduit le comportement général de la série. L'exemple le plus fréquent est la tendance linéaire.

1.4.2 Composante saisonnière (saisonnalité)

La composante saisonnière (notée S_t) correspond à un phénomène qui se répète à intervalles de temps réguliers. Elle est donc périodique (de période p). En effet, il suffit seulement de connaître les p premières valeurs S_1, \dots, S_p , et par périodicité on a $S_T = S_{t+p}$ pour tout t .

1.4.3 bruit blanc

Une suite de variables aléatoires $(\epsilon_t)_t$, $t = 0, 1, \dots$ constitue un *bruit blanc faible* (respectivement *fort*) si les trois propriétés suivantes sont respectées.

- (i) $\forall t \in \mathbb{Z}, E(\epsilon_t) = 0$.
- (ii) $\forall t \in \mathbb{Z}, Var(\epsilon_t) = \sigma^2$.
- (iii) $\forall t, s \in \mathbb{Z}, Cov(\epsilon_t, \epsilon_s) = 0$ si $t \neq s$ (respectivement, les ϵ_t sont *i.i.d.* (*i.e.* indépendants et identiquement distribués)).

1.4.4 Composante résiduelle (accidentelle ou irrégulière)

La composante résiduelle (notée ϵ_t) regroupe tout ce qui n'a pas été pris en compte par la tendance et le facteur saisonnier. Elle est la résultante de fluctuations irrégulières et imprévisibles dues à des facteurs perturbateurs non permanents.

Remarque 2. Ces trois composantes ne sont pas toujours simultanément présentes dans

une série chronologique. En effet, certaines séries n'ont pas de tendance, d'autres n'ont aucune composante périodique et d'autre ne connaissent aucune variation accidentelle.

1.4.5 Composante cyclique

La composante cyclique est définie comme des fluctuations sur de grandes périodes indépendamment de l'effet saisonnier. On ignore souvent cette composante cyclique, soit parce qu'elle n'existe pas, soit parce que les données statistiques ne remontent pas suffisamment dans le temps.

1.4.6 Perturbations

Elles sont des fluctuations ponctuelles de forte amplitude, dues par exemple, à une grève, à des conditions météorologiques exceptionnelles, à un krach financier... Il convient de les éliminer avant tout traitement mené sur la série.

1.5 Les modèles

Un modèle est une image simplifiée de la réalité et qui peut résumer au mieux l'information. Cependant, un modèle peut être meilleur qu'un autre pour décrire la réalité. Ici, on donne une liste qui sert à résumer et classer ces modèles.

1.5.1 Modèle additif

On considère une série $(X_t)_{t \in \mathbb{Z}}$ admettant une décomposition additive:

$$X_t = m_t + S_t + \epsilon_t, \quad t = 1 \dots T.$$

Les observations X_t sont modélisées comme superposition additive d'une tendance déterministe m_t , d'une saisonnalité (déterministe) S_t et d'une perturbation aléatoire ϵ_t qui représente les composantes (erreurs).

Ce modèle classique connaît comme généralisation,

$$X_t = f(t, y_t) = g(t) + y_t.$$

1.5.2 Modèle multiplicatif

Soit une série $(X_t)_{t \in \mathbb{Z}}$ admettant une décomposition multiplicatif:

$$X_t = m_t(1 + S_t)(1 + \epsilon_t), \quad t = 1 \dots T$$

Remarque 3. *Le modèle multiplicatif est généralement utilisé pour des données de type économiques.*

1.5.3 Modèle mixte

Les deux modèles (additif et multiplicatif) peuvent en fait se présenter dans un seul modèle appelé *mixte ou hybride*. En effet, on peut supposer, par exemple, que la composante saisonnière agit de façon multiplicative, alors que les fluctuations irrégulières sont additives,

$$X_t = m_t(1 + S_t) + \epsilon_t, \quad t = 1 \dots T.$$

1.6 Elimination de la tendance et saisonnalité

1.6.1 Elimination de la tendance en l'absence de saisonnalité

Supposons, dans un premier temps, que la partie déterministe du modèle soit uniquement composée d'une tendance m_t :

$$X_t = m_t + Y_t \tag{1.2}$$

Dans cette section, on donne trois méthodes pour estimer m_t . Ensuite, on peut l'éliminer après l'avoir estimée.

Estimation paramétrique

Cette méthode consiste à estimer le trend par la méthode des *moindres carrés ordinaires*,

$$\hat{m}_t = \hat{a}t + \hat{b}. \tag{1.3}$$

où \hat{a} et \hat{b} sont des estimateurs des coefficients de la fonction linéaire estimant m_t . La quantité à minimiser est

$$\sum_{t=1}^T (X_t - a - bt)^2, \tag{1.4}$$

La solution de ce problème de minimisation est, en posant $\bar{X} = \sum_{t=1}^T X_t$

$$\hat{a} = \frac{4T + 2}{T - 1} \bar{X} - \frac{6}{T(T - 1)} \sum_{t=1}^T tX_t \tag{1.5}$$

$$\hat{b} = \frac{6}{T(T - 1)} \left(\frac{2}{T(T - 1)} \sum_{t=1}^T tX_t - \bar{X} \right) \tag{1.6}$$

Cependant, on peut avoir une tendance sous forme d'un polynôme de degré n ,

$$\hat{m}_t = \hat{a}_0 + \sum_{i=1}^n \hat{a}_i t^i \quad (1.7)$$

Mais un polynôme de degré élevé pose des problèmes dans l'estimation des paramètres.

Estimation non paramétrique

Dans certaines situations, il n'est pas facile de trouver le degré du polynôme d'ajustement pour m_t . Il n'est pas possible d'utiliser la méthode des moindres carrés car le polynôme utilisé au départ pour \hat{m}_t n'est ni linéaire, ni quadratique. Ainsi, on pourrait utiliser un polynôme avec un degré élevé, mais le nombre de paramètres à estimer serait important et rendrait les calculs fastidieux. Dans cette situation, on a recours à la théorie non paramétrique de l'estimation du trend, qui ne suppose pas que la tendance soit polynomiale *à priori*. Pour comprendre cette technique, supposons que m_t soit linéaire dans un intervalle $[t - q, t + q]$. Dans ce cas, un bon estimateur de la tendance est donné par

$$\hat{m}_t = \frac{1}{2q + 1} \sum_{i=-q}^q X_{t+i} \quad (1.8)$$

On peut à présent calculer \hat{m}_t pour chaque valeur de t en calculant cette moyenne sur les $2q$ observations autour de t . On dit alors qu'on effectue une estimation par *moyenne mobile* (voir page 15).

Méthode des différences itérées

On peut également éliminer le trend sans l'estimer. Définissons tout d'abord l'opérateur de retard comme étant la fonction linéaire B :

$$BX_t = X_{t-1}.$$

Par récurrence, on aura:

$$B^n X_t = X_{t-n}.$$

On introduit également l'opérateur de différence ∇ défini comme suit:

$$\nabla X_t = X_t - X_{t-1}$$

En vertu de la définition précédente, on a directement:

$$\nabla X_t - X_t - BX_t.$$

Plus générale,

$$\nabla^d X_t = X_t - X_{t-d} \quad \text{ou} \quad \nabla^d X_t = (1 - B^d)X_t.$$

Supposons à présent que la tendance soit de la forme d'un polynôme d'ordre k :

$$m_t = \sum_{i=0}^k a_i t^i.$$

Si on applique à m_t l'opérateur de retard ∇ (d'ordre 1), on obtient :

$$\begin{aligned} \nabla m_t &= m_t - m_{t-1} \\ &= \sum_{i=0}^k a_i t^i - \sum_{i=0}^k a_i (t-1)^i \end{aligned}$$

en développant $(t-1)^i$

On a $(t-1)^i = \sum_{k=0}^i \zeta_i^k t^k (-1)^{i-k}$

pour $i = 2$

$$(t-1)^2 = \sum_{k=0}^2 \zeta_2^k t^k (-1)^{2-k} = 1 - 2t + t^2$$

$$\sum_{k=0}^i a_i (t-1)^2 = \sum_{k=0}^i a_i (1 - 2t + t^2)$$

$$\sum_{k=0}^i a_i (t-1)^i = \sum_{k=0}^i a_i (1 - 2t^{i-1} + t^i)$$

$$\sum_{k=0}^i a_i (t-1)^i = \sum_{k=0}^i a_i - 2 \sum_{k=0}^i a_i (t)^{i-1} + \sum_{k=0}^i a_i (t)^i$$

$$\sum_{k=0}^i a_i (t-1)^i = \sum_{k=0}^i a_i (t)^i + \sum_{k=1}^i c_i (t)^{i-1}$$

∇m_t devient:

$$\nabla m_t = \sum_{k=0}^{i-1} c_i t^k. \quad (1.9)$$

On voit que l'opérateur ∇ à diminuer l'ordre du polynôme m_t (m_t est d'ordre i), tandis que ∇m_t est d'ordre $i-1$. Dès lors, si on applique i fois l'opérateur ∇ au polynôme m_t , on obtient :

$$\nabla^i m_t = i! a_i = \text{constante indépendante de } t.$$

Par conséquent,

$$\nabla^i X_t = \text{constante} + \nabla^i Y_t. \quad (1.10)$$

1.6.2 Elimination du trend et de la saisonnalité

Supposons à présent qu'en plus d'un trend m_t , la série comporte une composante saisonnière S_t . On considère donc le modèle

$$X_t = m_t + S_t + Y_t \quad (1.11)$$

où :

- S_t est une fonction périodique en t , $S_t = S_{t+d}$ pour un certain $d > 0$.
- S_t ne comporte pas de composante "trend"

$$\sum_{j=1}^d S_{j+d} = 0$$

Dans le cas où la saisonnalité est de période d , on réorganise les données de la manière suivante

$$\{X_t\}_{t=1\dots T} = \{X_{jk}\}_{j=1\dots n \text{ et } k=1\dots d} \text{ avec } n \times d = T$$

Ainsi, on estime la tendance de chaque période avec la moyenne empirique:

$$\hat{m}_j = \frac{1}{d} \sum_{k=1}^d X_{jk}, \quad j = 1 \dots n. \quad (1.12)$$

Remarquons que la composante saisonnière dépend, par définition, de j et non de k . D'où, un estimateur de la composante S_j est donné aussi par la moyenne empirique

$$\hat{S}_k = \frac{1}{n} \sum_{j=1}^n (X_{jk} - \hat{m}_{jk}), \quad k = 1 \dots d. \quad (1.13)$$

Par conséquent, on peut éliminer la tendance et la saisonnalité à la fois.

1.6.3 Méthode des différences

Dans la formule (1.11) on peut appliquer la différenciation afin d'éliminer la tendance et la saisonnalité.

$$\begin{aligned} \nabla^d X_t &= X_t - X_{t-d} \\ &= m_t + S_t + Y_t - m_{t-d} - S_{t-d} - Y_{t-d} \\ &= (m_t - m_{t-d}) + (Y_t - Y_{t-d}) + (S_t - S_{t-d}) \\ &= \nabla^d m_t + \nabla^d Y_t. \end{aligned}$$

$\nabla^d X_t$ est donc une série chronologique désaisonnalisée, mais contenant un trend $\nabla^d m_t$.

1.6.4 calcul des variations saisonnières

Modèle additif(sans tendance)

$$X_t - \epsilon_t = S_t$$

Modèle multiplicatif(sans tendance)

$\frac{X_t}{\epsilon_t} = S_t$ le coefficient saisonnier S_j est donné par:

$$S_j = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \epsilon_{ij}) \text{ Modèle additif}$$

pour un modèle multiplicatif:

$$S_j = \frac{1}{n} \sum_{i=1}^n \frac{X_{ij}}{\epsilon_{ij}}$$

1.6.5 corrections des variations saisonnières

$$\bar{S}_p = \frac{1}{p} \sum_{j=1}^p S_j$$

Pour un modèle additif, on corrige S_j par: $\acute{S}_j = S_j - \bar{S}_p$ Pour un modèle multiplicatif, on corrige S_j par: $\acute{S}_j = \frac{S_j}{\bar{S}_p}$

$$\sum_{j=1}^p \acute{S}_j = \sum_{j=1}^p S_j - p\bar{S}_p = p\bar{S}_p - p\bar{S}_p = 0$$

$$\sum_{j=1}^p \acute{S}_j = \frac{\sum_{j=1}^p S_j}{p\bar{S}_p} = p$$

1.7 Moyennes mobiles

Les moyennes mobiles (moving averages) sont utilisées pour lisser une série chronologique. C'est un outil simple qui permet de mettre en évidence la tendance en supprimant la composante saisonnière et en réduisant l'amplitude des fluctuations irrégulières. Elles permettent de lisser directement la série sans hypothèse *à priori* sur la forme du modèle. La méthode est valide d'ailleurs quel que soit le modèle de décomposition.

1.7.1 Moyenne mobile simple

Il est clair qu'afin d'estimer la tendance, c'est-à-dire: la moyenne d'un phénomène dans un intervalle du temps, il faut disposer d'une série statistique sur une grande période. Le principe de la méthode est de construire une nouvelle série obtenue en calculant les moyennes successives de longueur k fixe à partir des données originales. En effet, chaque

moyenne obtenue correspond au milieu de la période de la moyenne calculée. Dans ce qui suit, on se limitera au cas régulier des instants observés. Les moyennes mobiles d'ordre k constituent à calculer:

- Les instants mobiles

$$\bar{t}_i = \frac{t_1 + \dots + t_k}{k}$$

- Pour chaque instants \bar{t}_i on calcule la moyenne empirique des observations

$$\bar{X}_{t_i} = \frac{X_{t_i} + \dots + X_{t_i+k-1}}{k}$$

Pour simplifier, on note $\bar{X}_{t_i} = \bar{X}_i$ appelée la *moyenne mobile d'ordre k* , calculée au temps t_i . On la notera par $MM_k(t_i)$.

Si $k = 2m + 1$, alors,

$$\bar{X}_i = \frac{X_{i-m} + \dots + X_{i+m}}{k}$$

Si $k = 2m$,

$$\bar{X}_i = \frac{X_{i-m+1} + \dots + X_{i+m}}{k}$$

1.7.2 Moyenne mobile centrée

Afin de conserver les dates d'observations lorsque k est pair, on utilise les moyennes mobiles centrées d'ordre k , qui seront notées $MM_k C(t_i)$. on définit $MM_k C(t_i)$ de la façon suivante:

$$\begin{aligned} \bar{X}_{(t_i)} &= \frac{\bar{X}\left(\frac{t_{i-1}+t_i}{2}\right) + \bar{X}\left(\frac{t_i+t_{i+1}}{2}\right)}{2} \\ &= \left(\frac{1}{2}X_{i-m} + X_{i-m+1} + \dots + X_{i-1} + X_i + \dots + X_{i+m-1} + \frac{1}{2}X_{i+m}\right)/k \end{aligned}$$

Remarque 4. *La méthode des moyennes mobiles correspond à un filtrage linéaire qui consiste à remplacer les valeurs d'une chronique par une combinaison linéaire des autres valeurs.*

a) Effet d'une moyenne mobile sur une composante saisonnière

Soit une série purement périodique de période p , i.e. $S_{t+p} = S_t, \forall p$.

Calculons la moyenne mobile d'ordre p ($p = 2m + 1$) associée à S_i ,

$$\bar{S}_i = \frac{S_{i-m} + S_{i-m+1} + \dots + S_i + S_{i+1} + \dots + S_{i+m}}{p}$$

le numérateur de \bar{S}_i est la somme de p terme consécutif de $(S_i)_i$, ce qui signifie que, si on applique une moyenne mobile d'ordre p à une série de période p , alors la composante saisonnière va s'éliminer.

Remarque 5. *Le résultat reste valable même si $m = 2p$.*

b) Effet d'une moyenne mobile sur la composante résidentielle

Une moyenne mobile réduit l'amplitude de la fluctuation irrégulière d'une chronique. Soit $(\epsilon_t)_t$ une suite de variable aléatoire telle que:

$$E(\epsilon_t) = 0 \quad \text{et} \quad \text{Var}(\epsilon_t) = \sigma^2, \quad \forall t$$

La moyenne mobile d'ordre p ($p = 2m + 1$) appliquée à $(\epsilon_t)_t$ donne

$$\bar{\epsilon}_i = \frac{1}{p} \sum_{l=-m}^m \epsilon_{i+l}$$

donc

$$E(\bar{\epsilon}_i) = 0 \quad \text{et} \quad \text{var}(\bar{\epsilon}_i) = \frac{\sigma^2}{p}, \quad \forall t$$

qui signifie que l'amplitude est divisée par la période.

1.8 Mise en œuvre sous R

Le **R** est un langage de programmation conçu pour les mathématiques. Ces dernières années, ce langage est devenu de plus en plus utilisé, notamment en statistique. La cause revient à sa puissance et sa simplicité d'apprentissage .

Pour mieux voir les choses, nous choisissons de mettre en pratique sous **R** quelques éléments que nous avons abordé dans ce mémoire. Cependant, il existe beaucoup d'ouvrages qui traitent les séries chronologiques sous **R**, nous citons par exemples Shumway & Stoffer (2006); Cryer & Chan (2008); Aragon (2011).

Dans cette section nous présentons quelques fonctions de base pour étudier les séries temporelles.

Pour traiter des données enregistrées en format `.txt` et placées dans le répertoire de **R**, on utilise les instructions suivantes:

```
> data=read.table("données.txt",sep="\t",header=T)
> attach(data)
```

`read.table()`: conçue pour importer le fichier nommé "données".

`header=T`: est une valeur logique indiquant si le fichier contient le nom des variables sur la première ligne.

Il faut préciser, dans `sep=` , comment les lignes sont séparées ("`\t`" = tabulation; " " = séparés par un espace; ", " = séparés par une virgule ...).

`attach()`: cette instruction permet de lire le fichier `data`.

Exemple 1. Nous représentons l'évolution des prix du pétrole en dollar, du février 1990 au juillet 2014, en utilisant de code **R** suivant:

```
> data=read.table("pétrole.txt",header=T)
> attach(data)
> X=ts(data[,2],start=1990-01,frequency=12)
> plot(X,type="l",ylab="Prix du pétrole",xlab="Années")
```

`ts()`: faite pour réécrire les données sous forme d'une série chronologique. On indique la date de commencement avec `start` ainsi que la fréquence avec `frequency` (on utilise la valeur 12 pour préciser que les valeurs sont prises 12 fois par année).

La fonction `plot()` permet de tracer la figure.

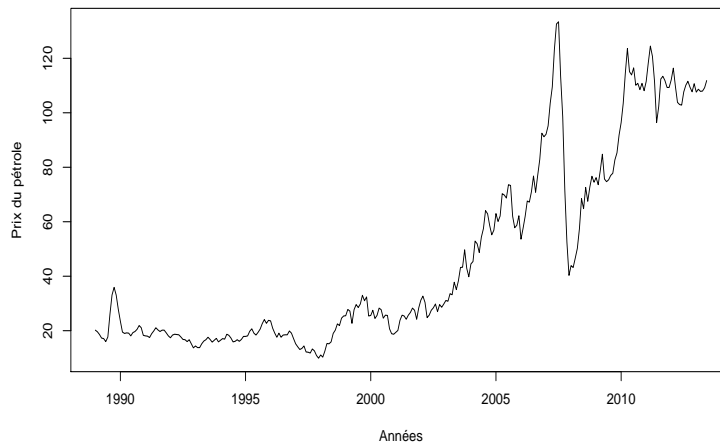


FIG. 1.1 – L'évolution des prix du pétrole, du février 1990 au juillet 2014.

Avec les instructions `acf()` et `pacf()` on représente la fonction d'autocorrélation empirique et la fonction d'autocorrélation empirique partielle respectivement.

```
> par(mfrow=c(1,2))
> acf(data[,2],lag.max=40,main="")
> pacf(data[,2],lag.max=40,main="")
```

1.8.1 Décomposition de la série

La sortie de la fonction `decompose` permet d'extraire la tendance et la saisonnalité ainsi que la partie aléatoire.

Exemple 2. Nous prenons l'exemple précédant. Le programme **R** est comme suit,

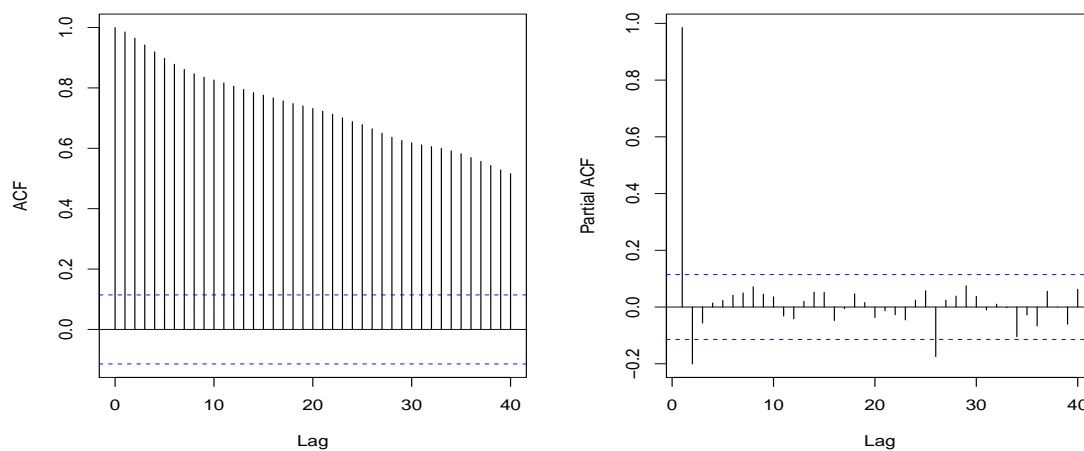


FIG. 1.2 – l'ACF (à gauche) et PACF (à droite) pour la série représentant les prix du pétrole.

```
> data=read.table("pétrole.txt",header=T)
> attach(data)
> X=ts(data[,2],start=1990-01,frequency=12)
> dec_X=decompose(X)
> plot(dec_X)
```

la figure résultante est comme suit,

Pour récupérer la composante saisonnière séparément, il suffit de taper `data$seasonal` (`data$trend` pour la tendance et `data$random` pour les résidus).

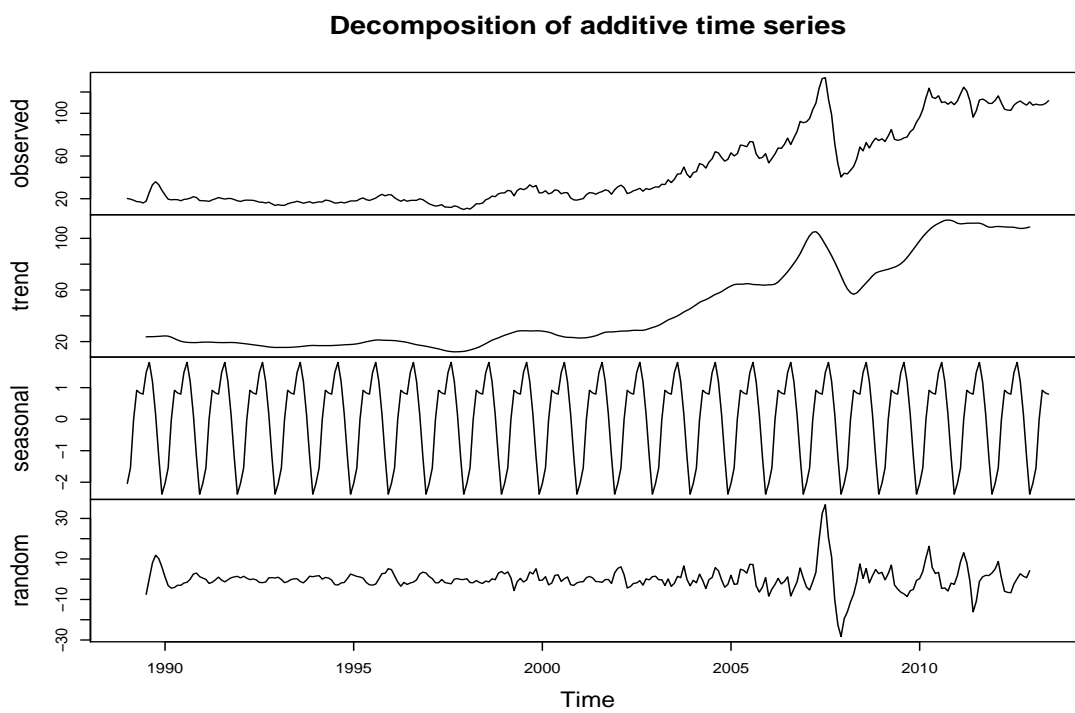


FIG. 1.3 – Décomposition additive pour la série consommation d'électricité.

Chapitre 2

Séries chronologiques linéaires

2.1 Processus autorégressifs AR

Un processus (X_t) est dit *autorégressif d'ordre p* noté $AR(p)$, si X_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période. Le processus est défini comme suit:

$$X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \epsilon_t \quad (2.1)$$

Les α_i sont des paramètres réels à estimer généralement et ϵ_t un bruit blanc centré de variance σ^2 .

2.2 Processus moyenne mobile MA

Dans un processus (X_t) de *moyenne mobile (Moving Average)* d'ordre q noté $MA(q)$, chaque observation x_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période dans le passé,

$$X_t = \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

2.3 Processus ARMA

Les modèles $ARMA(p, q)$ (*Autoregressive Moving Average*) sont représentatifs de processus générés par une combinaison des valeurs passées et des erreurs passées qui s'écrivent sous la forme

$$X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=0}^q \theta_j \epsilon_{t-j}$$

Il existe également d'autres processus comme: SARIMA, ARCH, GARCH,...

2.4 Estimation des paramètres

2.4.1 Estimation des paramètres autorégressifs (les équations de Yule-Walker)

Considérons tout d'abord un processus autorégressif d'ordre p causal et de moyenne nulle :

$$X_t = \sum_{k=1}^p \alpha_k X_{t-k} + \epsilon_t; \quad \text{avec } \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

En multipliant chaque membre de cette équation par: X_{t-h} ($h = 0, \dots, p$) et en prenant l'espérance de chaque côté, on obtient, pour le premier membre, la fonction d'autocovariance :

$$E(X_t X_{t-h}) = \rho_h \tag{2.2}$$

et, pour le second membre:

$$\sum_{k=1}^p \alpha_k E(X_{t-k} X_{t-h}) + E(\epsilon_t X_{t-h}) = \begin{cases} \sum_{k=1}^p \alpha_k \rho_{k-h}, & \text{si } h = 1, \dots, p \\ \sum_{k=1}^p \alpha_k \rho_k + \rho_\epsilon^2, & \text{si } h = 0 \end{cases} \tag{2.3}$$

car, par l'hypothèse de causalité du processus, on montre que $E(\epsilon_t X_{t-h})$ s'annule sauf pour $h = 0$. En égalant (2.2) et (2.3) on a, pour $h > 0$, la relation

$$\rho_h = \begin{cases} \sum_{k=1}^p \rho_{k-h} \alpha_k & \text{si } h = 1, \dots, p \\ \sum_{k=1}^p \rho_{k-h} \alpha_k + \rho_\epsilon^2 & \text{si } h = 0 \end{cases} \tag{2.4}$$

Ce système peut s'écrire sous forme matricielle. On note ρ_p le vecteur colonne de longueur p formé par $(\rho_1, \dots, \rho_p)'$, et α_p le vecteur colonne $(\alpha_1, \dots, \alpha_p)'$ et en formant la matrice $p \times p$ dont l'élément (i, j) vaut ρ_{i-j} c'est-à-dire:

$$\mathbf{R}_p = \begin{pmatrix} \rho_0 & \rho_1 & \dots & \rho_p \\ \rho_1 & \rho_0 & \dots & \rho_p \\ \vdots & \vdots & \ddots & \vdots \\ \rho_p & \rho_1 & \dots & \rho_0 \end{pmatrix} \tag{2.5}$$

alors le système (2.3) peut se réécrire matriciellement pour avoir:

$$\begin{cases} \rho_p = \mathbf{R}_p \alpha_p \\ \rho_0 = \hat{a}'_p \rho_p + \sigma_\epsilon \end{cases}$$

Ces dernières sont appelées *équations de Yule-Walker*. Elles relient les paramètres α_p du processus autorégressif à la fonction d'autocovariance théorique qui définit ρ_p et \mathbf{R}_p .

Si on veut estimer les paramètres $\alpha_1, \dots, \alpha_p$ du modèle, on peut se servir des équations de Yule-Walker en utilisant la fonction *d'autocovariance empirique* $\hat{\rho}_p$ comme un estimateur de ρ_h .

On construit alors le vecteur colonne $\tilde{\rho}_p = (\hat{\rho}_1, \dots, \hat{\rho}_p)'$ et la matrice $\hat{\mathbf{R}}_p$ construite en substituant ρ_h dans (2.5). Les relations trouvées ci-dessus entre α_p , ρ_p et \mathbf{R}_p permettent donc de trouver un estimateur $\hat{\alpha}_p$ de α_p et un estimateur $\hat{\sigma}_\epsilon$ de σ_ϵ à travers le système de Yule-Walker suivant:

$$\begin{cases} \hat{\rho}_p = \hat{\mathbf{R}}_p \hat{\alpha}_p \\ \hat{\rho}_0 = \hat{\alpha}'_p \hat{\rho}_p + \hat{\sigma}_\epsilon \end{cases} \quad (2.6)$$

Pour obtenir un estimateur de α_p il faut donc résoudre (2.5) en inversant la matrice \hat{R} :

$$\hat{\alpha}_p = (\hat{\mathbf{R}}_p)^{-1} \hat{\rho}_p$$

puis en calculant:

$$\hat{\sigma}_\epsilon = \hat{\rho}_0 - \hat{\alpha}'_p \hat{\rho}_p$$

Le cas d'un processus MA(q) est beaucoup plus difficile. Reprenons par exemple les relations reliant les autocorrélations et les paramètres d'un processus MA(2): $X_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$

$$\rho_X(1) = \frac{-\theta_1(1 - \theta_2)}{(1 + \theta_1^2 + \theta_2^2)} \quad \text{et} \quad \rho_X(2) = \frac{-\theta_2}{(1 + \theta_1^2 + \theta_2^2)}.$$

Ce système est non linéaire. Cela signifie en pratique la mise en oeuvre d'algorithme de résolution de système d'équations non linéaires. Les solutions obtenus n'ont pas la qualité de bons estimateurs. Excepté pour les cas des processus MA(1) et MA(2). La situation est identique pour les processus ARMA, même pour un processus ARMA(1, 1) le calcul est pour le moins peu "sympathique". En supposant utilisée la première relation $\hat{\rho}(2) = \alpha \hat{\rho}(1)$ et si $\alpha \neq \hat{\rho}(1)$ la deuxième équation est certes résoluble, dont les solutions sont (pour alléger l'écriture on pose $\rho = \hat{\rho}(1)$).

$$\frac{1}{\rho - \alpha} = \left(\rho\alpha - \frac{1}{2}\alpha^2 - \frac{1}{2}\sqrt{4\rho\alpha - 4\rho^2 - 2\alpha^2 + \alpha^4 - 4\rho\alpha^3 + 4\rho^2\alpha^2 + 1} - \frac{1}{2} \right)$$

$$\frac{1}{\rho - \alpha} = \left(\rho\alpha - \frac{1}{2}\alpha^2 + \frac{1}{2}\sqrt{4\rho\alpha - 4\rho^2 - 2\alpha^2 + \alpha^4 - 4\rho\alpha^3 + 4\rho^2\alpha^2 + 1} - \frac{1}{2} \right)$$

2.4.2 Estimation par maximum de vraisemblance

La méthode de Yule-Walker ne permet pas d'estimer de la façon efficace les paramètres d'un modèle contenant une partie MA parce qu'elle mène à un système d'équations non linéaires, ce que l'on peut facilement vérifier déjà pour le cas d'un MA(1). Néanmoins, il existe un algorithme itératif, l'algorithme des innovations qui permet de résoudre numériquement ce système d'équations non linéaires. Cet algorithme joue le même rôle que l'algorithme de Durbin-Levinson (voir page 38) pour le traitement itératif des équations de Yule-Walker.

Ici, on présente comme méthode alternative, la méthode générale du maximum de vraisemblance pour estimer les paramètres d'un modèle ARMA.

Considérons un modèle ARMA(p, q) de moyenne nulle

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \theta_0 \epsilon_t + \dots + \theta_q \epsilon_{t-q} \quad (2.7)$$

où $\theta_0 = 1$ et dans lequel on supposera que :

$$1 - \alpha_1 z - \dots - \alpha_p z^p \neq 0 \quad \text{pour } |z| \leq 1$$

et

$$1 - \theta_1 z - \dots - \theta_q z^q \neq 0 \quad \text{pour } |z| \leq 1$$

Notre objectif est à présent d'estimer les vecteurs colonnes $\alpha_p = (\alpha_1, \dots, \alpha_p)'$ et $\theta_q = (\theta_1, \dots, \theta_q)'$.

Dans un premier temps, supposons que l'on connaisse la distribution de X_t pour tout t . Nous supposons que le processus $\{X_t\}$ est gaussien, c'est-à-dire que toutes les fonctions de distributions de $\{X_t\}$ sont multivariées normales. Cette hypothèse permettra d'écrire simplement la fonction de vraisemblance du modèle qui est la probabilité d'observer $X_T = (X_1, \dots, X_T)'$ sous l'hypothèse gaussienne:

$$\mathcal{L}(\mathbf{R}_T) = (2\pi)^{-T/2} |\det \mathbf{R}_T|^{-1/2} e^{-\frac{1}{2} X_T' (\mathbf{R}_T)^{-1} X_T}$$

où \mathbf{R}_T est la matrice de covariance dont l'élément (i, j) est donné par $E(X_i X_j) = \rho_{|i-j|}$. Cette matrice est constituée par la fonction d'autocovariance ρ_h .

Or, la fonction d'autocovariance pouvait être exprimée en fonction des paramètres $\alpha_p, \theta_q, \sigma_\epsilon^2$, par la relation:

$$\rho_h = \sigma_\epsilon^2 \sum_{k=0}^{\infty} \alpha_k \alpha_{k+h}. \quad (2.8)$$

On peut donc également exprimer \mathbf{R}_T en fonction de ces paramètres:

$$\mathbf{R}_T = \mathbf{R}_T(\alpha_p, \theta_q, \sigma_\epsilon^2).$$

et la fonction de vraisemblance du processus gaussien s'écrit :

$$\mathcal{L}(\alpha_p, \theta_q, \sigma_\epsilon^2) = \mathcal{L}(\mathbf{R}_T(\alpha_p, \theta_q, \sigma_\epsilon^2)) \quad (2.9)$$

Si $\{X_t\}$ n'est pas un processus gaussien, on ne connaît pas en général les fonctions de distribution de $\{X_t\}$ et la fonction de vraisemblance n'est plus disponible. Toutefois, même dans ce cas, on utilisera (2.9) comme une mesure de la qualité de l'ajustement de la matrice de covariance aux données et on considérera cette fonction pour trouver des estimateurs de $\alpha_p, \theta_q, \sigma_\epsilon^2$.

Ces estimateurs sont donnés par la maximisation de $\mathcal{L}(\alpha_p, \theta_q, \sigma_\epsilon^2)$. Il est alors possible de montrer que, si le processus (2.7) est causal, inversible et $\epsilon_t \sim IID(0, \sigma_\epsilon^2)$, alors le processus

$$Y_t = \sum_{i=1}^p \hat{\alpha}_i X_{t-i} + \epsilon_t + \sum_{j=1}^q \hat{\theta}_j \epsilon_{t-j}, \quad IID(0, \hat{\sigma}_\epsilon^2),$$

est lui-même stationnaire causal et inversible. Il est également possible de démontrer un théorème de normalité asymptotique de ces estimateurs.

Proposition 3. *Si X_t est un processus causal et inversible de la forme (2.7) avec $\epsilon_t \sim IID(0, \sigma_\epsilon^2)$, et $\beta = (\alpha_p, \theta_q)$, l'estimateur de maximum de vraisemblance $\hat{\beta} = (\hat{\alpha}_p, \hat{\theta}_q)$ est asymptotiquement normal*

$$\sqrt{T}(\hat{\beta} - \beta) \longrightarrow \mathcal{N}(0, \Sigma_\beta)$$

où la matrice variance-covariance Σ_β est une fonction du paramètre β (pour les détails voir Brockwell and Davis 1991) qui coïncide, pour le cas d'un processus autorégressif pur d'ordre p , avec la matrice de la variance asymptotique des estimateurs Yule-Walker $\sigma_\epsilon^2 \mathbf{R}_p^{-1}$.

En pratique, les estimateurs du maximum de vraisemblance sont trouvés numériquement par des 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 initiale peut être fournie par les estimateurs de Yule-Walker ou les estimateurs des moindres carrés. Un modèle ARMA, la partie MA n'est pas observable, pour tous ces estimateurs il faut se servir du concept des résidus. Ceux-ci sont définis (supposant que $\mu_X = 0$) par une itération :

$$e_t := X_t - \sum_{k=1}^p \alpha_k X_{t-k} - \sum_{j=1}^q \theta_j e_{t-j} \quad (2.10)$$

avec $e_t := 0, t \leq 0$, et $X_t := 0, t \leq 0$.

Pour comprendre comment le concept des résidus aide à construire des estimateurs de $\alpha_p, \theta_q, \sigma_\epsilon^2$, étudions donc les estimateurs des moindres carrés définis par :

$$(\hat{\alpha}_p, \hat{\theta}_q)_{MC} = \arg \min \sum_{t=\max(p,q)+1}^T e_t^2$$

et

$$\sigma_{\epsilon,MC}^2 = T^{-1} \sum_{t=\max(p,q)+1}^T e_t^2.$$

On s'aperçoit que dans la somme des moindres carrés (à minimiser) tout ne dépend maintenant que des valeurs observables.

Exemple d'un processus AR(1)

Hypothèse essentielle: la distribution du bruit blanc est supposée connue (Dans notre cas supposée gaussienne). Soit $(X_t)_{n \in \mathbb{Z}}$ un processus AR(1):

$$X_t = C + \alpha X_{t-1} + \epsilon_t$$

Nous savons que l'équation de Yule-Walker d'ordre 1 donne un estimateur du paramètre autorégressif α . Le propos est de calculer la fonction de la Log-vraisemblance afin d'estimer le vecteur $\beta = (C, \alpha, \sigma^2)'$.

Proposition 4. *Soit $(X_t)_{n \in \mathbb{Z}}$ un processus AR(1): $X_t = C + \alpha_1 X_{t-1} + \epsilon_t$ où $(\epsilon_t)_{t \in \mathbb{Z}}$ est bruit blanc $\mathcal{N}(0, \sigma^2)$ et x_1, \dots, x_T une trajectoire observée. Alors*

$$\begin{aligned} LV(\beta) &= \log(f(x_1, \dots, x_T); \beta) \\ &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \left(\sigma^2 / (1 - \alpha^2) - \frac{1}{2} \frac{(x_1 - C / (1 - \alpha))^2}{\sigma^2 / (1 - \alpha^2)} \right) \\ &= -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \sum_{t=2}^T \left[\frac{(x_t - C - \alpha x_{t-1})^2}{2\sigma^2} \right]. \end{aligned}$$

La démonstration s'appuie, de façon équivalente, sur deux approches: la première utilise de façon itérative la vraisemblance conditionnelle; la seconde, repose sur la représentation de la vraisemblance d'un vecteur gaussien. La dernière proposition se généralise pour un processus ARMA(p, q), mais la fonction de vraisemblance a une écriture très complexe. Parmi les raisons de cette difficulté, le calcul de l'inverse de la matrice variance-covariance. Cependant, cette matrice possède une structure particulière, elle est symétrique et de sous et sur-diagonales identiques (appelée matrice de Toeplitz). Malheureusement l'inverse de cette matrice n'est pas une matrice de Toeplitz. Par contre, il est possible de mener un calcul récursif sur cet inverse. En effet, c'est le fondement et l'intérêt de l'algorithme de Durbin-Levinson. Il est préférable (et asymptotiquement, on démontre que c'est équivalent) d'utiliser des formes approchées comme l'estimateur des moindres carrés qui fait abstraction de la première observation. Dans le cas d'un processus AR(1), la fonction à

maximiser, conditionnellement à x_1 s'écrit,

$$\begin{aligned} LV(\beta|X_1 = x_1) &= -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \sum_{t=2}^T \left[\frac{(x_t - C - \alpha x_{t-1})^2}{2\sigma^2} \right] \\ &= -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \sum_{t=2}^T \left[\frac{e_t^2}{2\sigma^2} \right] \end{aligned}$$

La séparation des paramètres σ^2 , C et α ramène à la minimisation de la somme des carrés résiduels

$$\hat{\sigma}^2(C, \alpha) = \sum_{t=2}^T (x_t - C - \alpha x_{t-1})^2$$

Un tel estimateur possède, asymptotiquement, les propriétés de l'estimateur du maximum de vraisemblance.

Exemple d'un processus MA(1)

Soit un processus de moyenne mobile d'ordre 1

$$X_t = \mu + \epsilon_t - \theta \epsilon_{t-1}, \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

Le vecteur des paramètres à estimer est

$$\theta = (\mu, \theta, \sigma)^T \in \Theta \subset \mathbb{R}^3$$

a) Approche conditionnelle

On donne quelques étapes génériques des approches conditionnelles.

Supposons que $\epsilon_0 = 0$. La distribution conditionnelle de

$$X_1 | \epsilon_0 = \mu + \epsilon_1 \sim \mathcal{N}(\mu - \theta \epsilon_{t-1}, \sigma^2)$$

Plus généralement, si ϵ_{t-1} est connu, alors

$$X_t | \epsilon_{t-1} \sim \mathcal{N}(\mu - \theta \epsilon_{t-1}, \sigma^2)$$

et

$$f(x_t | \epsilon_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x_t - \mu + \theta \epsilon_{t-1})^2}{2\sigma^2} \right].$$

Le calcul de $t = 1$ à T est répété et par application successives de la formule de Bayes,

$$\mathcal{L}(x_1, \dots, x_T | \epsilon_0, \theta, \mu, \sigma^2) = \prod_{t=1}^T f(x_t | \epsilon_{t-1}; \theta, \mu, \sigma^2)$$

on obtient pour la log-vraisemblance, conditionnellement à $\epsilon_0 = 0$

$$\log(\mathcal{L}(x_1, \dots, x_T | \epsilon_0, \theta, \mu, \sigma^2)) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2} \sum_{t=1}^T \frac{\epsilon_t^2}{\sigma^2}$$

L'expression peut paraître simple, elle est cependant non linéaire en le paramètre θ ; par exemple $\epsilon_t = (x_t - \mu) + \theta(x_{t-1} - \mu) + \theta^2(x_{t-2} - \mu) + \dots + \theta^t \epsilon_0$.

b) Approche exacte

Utilisons la structure multivariée gaussienne du processus. Un processus MA(1) gaussien possède une matrice de variance-covariance,

$$\sigma^2 \Sigma = \sigma^2 \begin{bmatrix} (1 + \theta^2) & -\theta & 0 & \dots & 0 \\ -\theta & (1 + \theta^2) & -\theta & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (1 + \theta^2) \end{bmatrix}$$

Le log-vraisemblance s'écrit donc

$$-\frac{T}{2} (\log(2\pi\sigma^2) + \log(\det(\Sigma))) - \frac{1}{2\sigma^2} (x - \mu)' \Sigma_\epsilon^{-1} (x - \mu)$$

On sait qu'une matrice symétrique se factorise $\Sigma = UDU'$ avec U une matrice triangulaire inférieure et D une matrice diagonale de terme générique

$$d_{ii} = \sigma^2 \frac{1 + \theta^2 + \dots + \theta^{2i}}{1 + \theta^2 + \dots + \theta^{2(i-1)}}$$

En posant

$$\bar{x}_t = x_t - \mu - \theta \frac{1 + \theta^2 + \dots + \theta^{2(t-2)}}{1 + \theta^2 + \dots + \theta^{2(t-1)}} x_{t-1}$$

On obtient

$$\log(\mathcal{L}(x_1, \dots, x_T | \epsilon_0, \mu, \theta, \sigma^2)) = \frac{T}{2} \log(2\pi) - \frac{T}{2} \sum_{t=1}^T \log(d_{tt}) - \frac{1}{2} \sum_{t=1}^T \frac{\bar{x}_t^2}{d_{tt}^2}$$

La non linéarité des paramètres est évidente.

2.4.3 Vraisemblance à la Box-Jenkins

Soit un processus ARMA(p, q) gaussien suivant,

$$X_t = C + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

pour estimer les paramètres C, θ et σ_ϵ^2 , on suppose que les valeurs initiales x_0 et ϵ_0 sont données, donc la vraisemblance conditionnelle s'écrit comme suit,

$$LV(\beta|x_0, \epsilon_0) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^T \left[\frac{e_t^2}{2\sigma^2} \right]$$

Comme le cas précédent, cela revient à minimiser (Moindres carrés)

$$\sigma^2(\hat{C}, \hat{\phi}, \hat{\theta}) = \sum_{t=1}^T e_t^2.$$

2.4.4 Quelques résultats asymptotiques

Pour les cas les plus simples, les résultats suivants peuvent être utilisés:

– Processus AR(1)

$$\hat{\alpha}_1 \xrightarrow{\mathcal{L}} \mathcal{N} \left(\alpha_1, \frac{1}{T}(1 - \alpha_1^2) \right)$$

– Processus MA(1)

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \frac{1}{T} \begin{pmatrix} 1 - \alpha_1^2 & -\alpha_1(1 + \alpha_2) \\ -\alpha_1(1 + \alpha_2) & 1 - \alpha_2^2 \end{pmatrix} \right)$$

– Processus MA(1)

$$\hat{\theta}_1 \xrightarrow{\mathcal{L}} \mathcal{N} \left(\theta_1, \frac{1}{T}(1 - \theta_1^2) \right)$$

– Processus MA(2)

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \frac{1}{T} \begin{pmatrix} 1 - \theta_1^2 & -\theta_1(1 + \theta_2) \\ -\theta_1(1 + \theta_2) & 1 - \theta_2^2 \end{pmatrix} \right)$$

– Processus ARMA(1,1)

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\theta} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} \alpha \\ \theta \end{pmatrix}, \frac{1}{T} \frac{1 - \theta\alpha}{(\alpha - \theta)^2} \begin{pmatrix} (1 - \alpha^2)(1 - \theta\alpha) & (1 - \alpha^2)(1 - \theta^2) \\ (1 - \alpha^2)(1 - \theta^2) & (1 - \theta^2)(1 - \theta\alpha) \end{pmatrix} \right)$$

2.5 Test de stationnarité à une tendance déterministe près

Kwiatkowski et al. (1992) proposent un test (appelé KPSS) pour la stationnarité. Ce test suppose que la série est la somme d'une marche aléatoire, d'un trend déterministe et d'une erreur stationnaire:

$$X_t = R_t + \alpha_1 + \alpha_2 t + Z_t \tag{2.11}$$

avec R_t est la marche aléatoire: $R_t = R_{t-1} + \epsilon_t \mathcal{N}(0, \sigma_\epsilon^2)$, $\alpha_1 + \alpha_2 t$ une tendance déterministe et d'une erreur stationnaire Z_t . l'hypothèse nulle est $\sigma_\epsilon^2 = 0$ (qui veut dire que la marche aléatoire n'existe pas), contre $\sigma_\epsilon^2 > 0$. La statistique du test KPSS est comme suit,

$$\text{KPSS} = \frac{1}{(T\hat{\lambda})^2} \sum_{t=1}^T \hat{S}_t^2 \quad (2.12)$$

où $\hat{S} = \sum_{i=1}^t \hat{u}_i$, \hat{u}_i est le résidu de la régression de X_t sur la composante déterministe supposée et $\hat{\lambda}^2$ un estimateur convergent de la variance de long terme de u_t basé sur \hat{u}_t (voir Aragon 2011).

la loi de KPSS converge vers une loi non standard qui ne dépend pas des valeurs de α mais seulement de la forme de la tendance qui peu être un niveau ($\alpha_2 = 0$) ou une tendance linéaire $\alpha_1 + \alpha_2 t$. On rejette l'hypothèse nulle de stationarité pour de grandes valeurs de la statistique de test.

2.6 Validation d'un modèle ARMA

Dans une modélisation, la validation du modèle est une étape inévitable avant de faire les prévisions. Pour les modèles ARMA, on présente deux tests pour tester les résidus.

1- Premier test

Au cours de la prédiction d'un processus ARMA, le résidu est défini comme la série temporelle:

$$\hat{e}_t = \frac{X_t - \hat{X}_t(\hat{\alpha}, \hat{\theta})}{\sqrt{r_{t-1}(\hat{\alpha}, \hat{\theta})}} \quad (2.13)$$

où r_{t-1} étant la variance des erreurs de prédiction divisée par la variance de ϵ_t . Lorsque le processus est estimé correctement, la série des résidus suit asymptotiquement à peu près la même loi que le processus générateur ϵ_t (bruit blanc faible en général). On peut donc tester la qualité de l'estimation en observant la fonction de corrélation empirique $\hat{\rho}(h)$ de \hat{e}_t . Celle-ci vaut 1 pour $h = 0$, et pour $h \geq 1$, elle est comprise entre $\frac{-1.96}{\sqrt{n}}$ et $\frac{1.96}{\sqrt{n}}$ avec une probabilité de 95%. C'est le *test de "blancheur" de Bartlett*. Lorsque l'estimation du processus a été obtenue par maximisation de la vraisemblance, les intervalles de confiance peuvent être diminués pour les h petits. L'application de ce critère permet de dire si le modèle ARMA choisi (avec ses ordres et ses paramètres) modélise correctement le processus. Il est nécessaire pour le calculer de mettre en œuvre une prédiction du processus.

2- Deuxième test

Les *tests "portmanteau"* portent sur une statistique construite à partir de la corrélation empirique des résidus. On utilisera le test de Ljung-Box. Si on note $\hat{\rho}$ l'autocorrélation empirique des résidus de l'ajustement à un ARMA(p, q), alors, pour h assez grand, mais

petit devant n ,

$$Q = n(n-2) \sum_{i=1}^h \frac{\hat{\rho}(i)}{n-i} \xrightarrow{\mathcal{L}} \chi^2(h-df), \text{ pour } n \rightarrow \infty. \quad (2.14)$$

où df est le nombre de coefficients estimés (sans compter la moyenne), en général $p+q$. Au niveau α , on rejettera donc l'hypothèse ARMA(p, q) si Q est plus grand que le quantile d'ordre $1-\alpha$ de la loi $\chi^2(h-df)$.

2.7 Mise en œuvre sous R

2.7.1 Simulation des séries chronologiques

Pour simuler un modèle ARMA(p, q) défini comme suit,

$$X_t = C + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=0}^q \theta_j \epsilon_{t-j}$$

sous R, on utilise l'instruction suivante,

> `X=arima.sim(list(ar=c(0.4,0.5),ma=c(0.2,.5)),n)`, où:

- `n` représente le nombre de simulation.

- `ar=c(...)` et `ma=c(...)` représentent les vecteurs α et θ respectivement.

Voici quelques exemples simulés de séries temporelles,

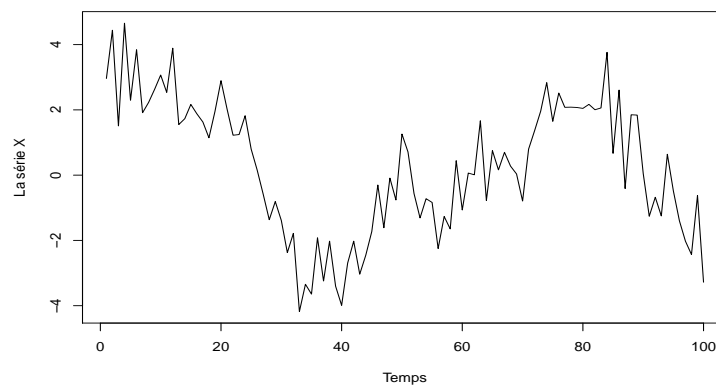


FIG. 2.1 – Série chronologique d'un processus AR(2) avec $\alpha = (0.4, 0.5)$.

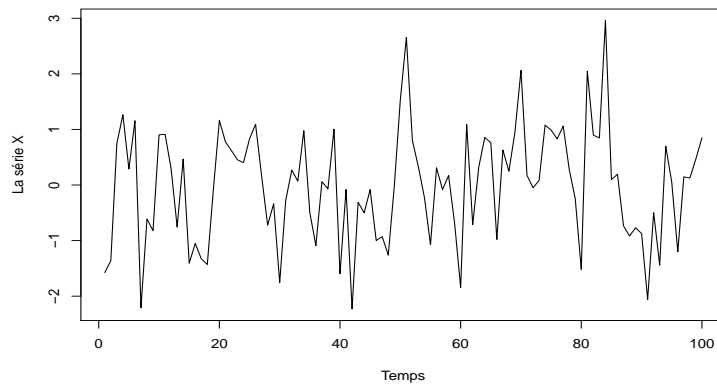


FIG. 2.2 – Série chronologique d'un processus MA(2) avec $\theta = (0.2, 0.5)$.

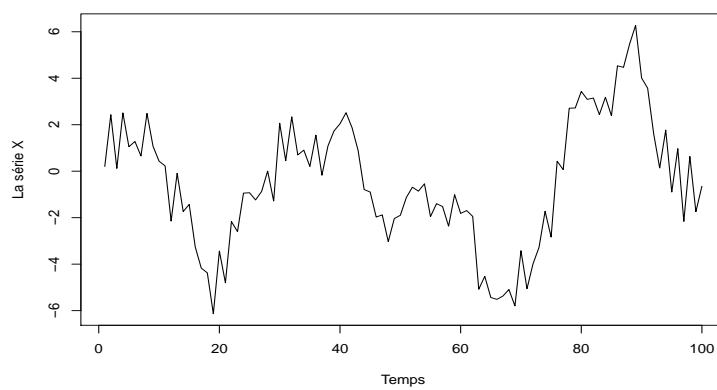


FIG. 2.3 – Série chronologique d'un processus ARMA(2,2) avec $\theta = (0.2, 0.5)$ et $\alpha = (0.4, 0.5)$.

2.7.2 Estimation des paramètres

2.7.3 Estimation des paramètres d'un processus $AR(p)$

Nous avons vu que l'estimation des paramètres d'un processus $AR(p)$ peut se faire avec l'équation de Yule-Walker. Sous **R** la fonction `ar.yw()` permet de fournir ces estimations.

Exemple 3. Nous simulons un processus $AR(3)$ avec $\alpha = (0.1, 0.5, 0.3)$ puis nous estimons α avec l'équation de Yule-Walker. Le code **R** est comme suit,

```
> set.seed(17)
> y=arima.sim(list(ar=c(0.1,0.5,0.3)),300)
> ar.yw(y,order=3)
Call:
ar.yw.default(x = y, order.max = 3)
Coefficients:
      1      2      3
0.0937 0.5541 0.2164
Order selected 3  sigma^2 estimated as  1.132
```

La commande `set.seed()` permet d'avoir les mêmes valeurs simulées pour chaque appel à cette fonction. C'est-à-dire que le lecteur peut reproduire les mêmes résultats en mettant le même code! (voir Robert et Casella 2011 pour plus d'explications). Si le paramètre p est inconnu, alors nous devons choisir une valeur arbitraire pour p . Cependant, cela peut créer des problèmes. Si on pose, pour le paramètre `order`, 2 ou 5 à la place de la vraie valeur ($p=3$), on aura,

```
> ar.yw(y,order=2)
Call:
ar.yw.default(x = y, order.max = 2)
Coefficients:
      1      2
0.2241 0.6026
> ar.yw(y,order=5)
Call:
ar.yw.default(x = y, order.max = 5)
Coefficients:
      1      2      3      4
0.1123 0.6017 0.2244 -0.0859
```

Nous remarquons que ces résultats ne sont pas proches au cas de la vraie valeur de p .

2.7.4 Estimation des paramètres d'un processus ARIMA

A l'aide de la fonction `auto.arima()` de package **forecast**, on peut chercher le meilleur modèle ARIMA (voir Hyndman 2007). Cette fonction contient plusieurs paramètres. Le paramètre d est l'ordre de différenciation simple. Sans précision, le test de KPSS en choisit une valeur automatiquement. Même chose pour l'ordre de différenciation saisonnière D . Il faut ensuite donner des ordres maximum d'autorégression et moyenne mobile. Par défaut, les ordres maximum sont `max.p=5`, `max.q=5` et pour les termes saisonniers `max.P=2`, `max.Q=2`. Notons qu'il existe d'autres paramètres. On peut les voir en tapant la commande `help(auto.arima)`.

Exemple 4. Nous prenons toujours l'exemple des prix du pétrole.

Premièrement, nous vérifions la stationnarité de cette série avec le test KPSS de package **tseries**.

```
> library(tseries)
> kpss.test(x)
```

Ce qui donne,

```
      KPSS Test for Level Stationarity
data:  x
KPSS Level = 6.046, Truncation lag parameter = 3, p-value = 0.01
Message d'avis :
In kpss.test(x) : p-value smaller than printed p-value
```

La *p-value* étant inférieure à 5%, alors on déduit que la série n'est pas stationnaire. Dans ce cas nous essayons de différencier la série une ou plusieurs fois pour avoir la stationnarité. Alors nous commençons par $d = 1$. L'instruction `diff()` se charge de cette fonction (la différenciation).

```
> kpss.test(diff(x,1))
      KPSS Test for Level Stationarity
data:  diff(x, 1)
KPSS Level = 0.0711, Truncation lag parameter = 3, p-value = 0.1
Message d'avis :
In kpss.test(diff(x, 1)) : p-value greater than printed p-value
```

On voit que la série différenciée est bien stationnaire, avec une *p-value* supérieure à 5%.

Deuxièmement, on passe à estimer les paramètres avec,

```
> auto.arima(x,max.p=5,max.q=5,d=1,stationary=T);
```

Le résultat de cette commande est donc,

```
Series: x
ARIMA(1,0,2) with non-zero mean
Coefficients:
          ar1      ma1      ma2  intercept
          0.9864  0.3806  0.1883    54.0275
s.e.      0.0094  0.0576  0.0578    22.7012
sigma^2 estimated as 16.52:  log likelihood=-831.77
AIC=1673.54  AICc=1673.75  BIC=1691.96
```

D'où, le modèle estimé est un ARIMA(1,1,2) avec $C = 54.0275$, $\alpha = 0.9864$ et $\theta = (0.3806, 0.1883)$.

Troisièmement, après avoir estimé les paramètres, le modèle doit être validé. Pour se faire on procède comme suit:

1. Tester la blancheur des résidus par le test de Ljung-Box, en utilisant la fonction `Box.test.2()`.
2. Tester la normalité des résidus avec le test de Jacs-Bera, en utilisant la fonction `jarque.bera.test()`.

```
> résidus=residuals(auto.arima(x,max.p=5,max.q=5,d=1,stationary=T))
> retard=c(1,3,6,12)
> Box.test.2(résidus,nlag=retard,type="Ljung-Box",decim=10)
      Retard  p-value
[1,]      1 0.9138615
[2,]      3 0.9811479
[3,]      6 0.4208851
[4,]     12 0.1341806
```

Les valeurs de *p-value* sont élevées pour les différents retards, donc on conclut à la blancheur du bruit.

Pour la normalité, les deux commandes: `hist()` et `qqnorm()` sont des outils graphiques permettant d'avoir une idée préalable sur la normalité.

```
> par(mfrow=c(1,2))
> hist(résidus, breaks=30,prob=TRUE)
> qqnorm(résidus)
```

Sur l'histogramme on peut remarquer que les résidus sont loin de la loi normale. Et sur le QQ-plot, on voit que les points sont très loin de la ligne bleue, ce qui suggère une

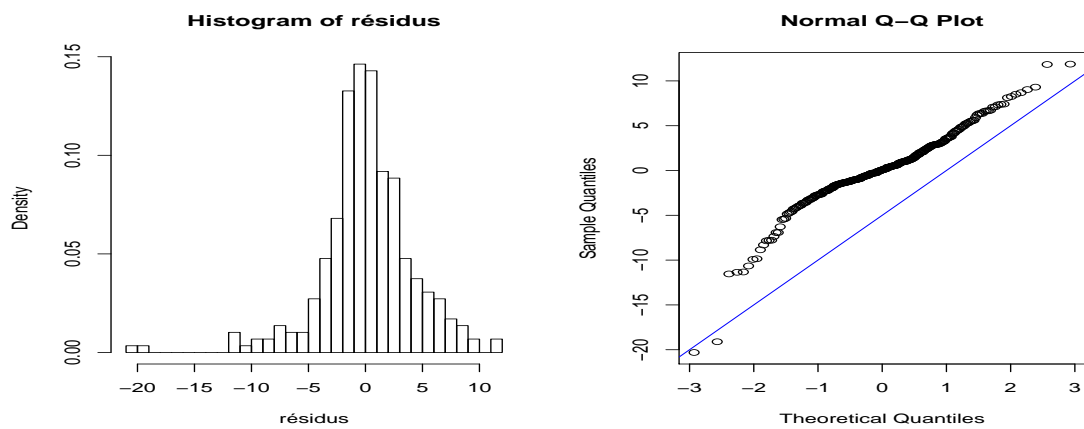


FIG. 2.4 – Histogramme et QQ-plot des résidus.

non-normalité. En effet, le test Jacs-Bera le confirme effectivement,

```
> jarque.bera.test(résidus);
```

Jarque Bera Test

data: résidus

X-squared = 217.6579, df = 2, p-value < 2.2e-16

Conclusion: Avec la non-normalité des résidus, le modèle ne peut être validé. Donc, il faut choisir un autre modèle.

Chapitre 3

Séries chronologiques linéaires

Pour traiter le problème de la prédiction des valeurs futures X_{t+1}, X_{T+2}, \dots d'une série chronologique observée jusqu'au temps T , nous supposons que les observations proviennent d'un processus stationnaire ARMA(p, q), causal, de moyenne nulle, et dont on connaît l'ordre p, q ainsi que les coefficients. Bien entendu, en pratique, lorsqu'on est face à T observations, il faudra ajuster un modèle adéquat (valide).

3.1 Géométrie des variables aléatoires

Soit $\mathbf{L}^2 = \{X/X \text{ v.a. réelle tel que } \mathbb{E}(X^2) < +\infty\}$ l'ensemble des variables à carré intégrable. On peut munir \mathbf{L}^2 d'un produit scalaire défini par :

$$\langle X, Y \rangle = \mathbb{E}(XY)$$

De même on définit $\|X\|^2 = \mathbb{E}(X^2) = \langle X, X \rangle = 0, =: \|X\| = [\mathbb{E}(X^2)]^{1/2}$

\mathbf{L}^2 muni de cette norme est dit espace de Hilbert (c'est un espace vectoriel normé complet de dimension infinie).

Remarques:

* $\forall X \in \mathbf{L}^2 / \mathbb{E}(X) = 0$ alors $\|X\|^2 = \mathbb{E}(X^2) = V(X) =: \|X\| = \sigma_X$.

* $\forall X \in \mathbf{L}^2 / \mathbb{E}(X) = \mathbb{E}(Y) = 0$ alors $cov(X, Y) = \mathbb{E}(X.Y) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X.Y) = \langle X, Y \rangle$.

* $\langle X, Y \rangle = \|X\| \|Y\| \cdot \cos(\vec{X}, \vec{Y}) = \|X\| \|Y\| \cdot \cos(\theta)$.

* $\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\|X\| \|Y\| \cdot \cos(\theta)}{\|X\| \|Y\|} = \cos(\theta)$.

* si $\rho = 0$ implique $\theta = \frac{\pi}{2}$. L'absence de corrélation se traduit par l'orthogonalité entre X et Y .

Théorème 2. Soit $Y \in \mathbf{L}^2$, alors la projection orthogonale de $Y \in \mathbf{L}^2$ sur D qu'on notera $P_D(Y) = \mathbb{E}(Y)$

Démonstration 2. On doit montrer que $\|Y - a\|^2$ est minimale pour $a = \mathbb{E}(Y)$

i.e: $\|Y - a\|^2 \geq \|y - \mathbb{E}(Y)\|^2, \forall a \in \mathbb{R}$

$$\|Y - a\|^2 = \mathbb{E}((Y - a)^2) = f(a)$$

$$\frac{\delta f}{\delta a} = 0 \Leftrightarrow 2\mathbb{E}(Y - a) = 0 \Leftrightarrow a = \mathbb{E}(Y)$$

$$\frac{\delta^2 f}{\delta a^2} < 0 \Leftrightarrow \mathbb{E}(Y) = P_D(Y)$$

Interprétation :

La meilleure approximation d'une v.a par une constante est $\mathbb{E}(Y)$

* Soit $\mathbf{L}_X^2 = \{Y \in \mathbf{L}^2 / Y = \varphi(X)\}$

Théorème 3. Soit $Y \in \mathbf{L}^2$ alors la projection orthogonale de y sur \mathbf{L}_X^2 est

$$P_{\mathbf{L}_X^2}(Y) = \mathbb{E}(Y|X)$$

Démonstration 3.

On doit montrer que $\|X - \varphi(X)\|^2 \geq \mathbb{E}\|(Y - \mathbb{E}(Y|X))\|^2 \forall \epsilon$

$$\|X - \varphi(X)\|^2 = \mathbb{E}[(Y - \varphi(X))^2] = \mathbb{E}[(Y - \mathbb{E}(Y|X) + \mathbb{E}((Y|X) - \varphi(X)))^2]$$

On sait que $\mathbb{E}\mathbb{E}(Y|X) = \mathbb{E}(y)$ et que:

$$\mathbb{E}(h(X).Y|X) = h(X).\mathbb{E}(Y|X)$$

$$\|Y - \varphi(X)\|^2 = \mathbb{E}(((Y - \mathbb{E}(Y|X)) + (\mathbb{E}(Y|X) - \varphi(X)))^2)$$

$$= \mathbb{E}(\mathbb{E}[(Y - \mathbb{E}(Y|X)) + (\mathbb{E}(Y|X) - \varphi(X))^2|X])$$

$$= \mathbb{E}(\mathbb{E}[(Y - \mathbb{E}(Y|X))^2 + \mathbb{E}(\mathbb{E}(Y|X) - \varphi(X))^2|X]) + 2\mathbb{E}[(\mathbb{E}(Y - \mathbb{E}(Y|X))\mathbb{E}(\mathbb{E}(Y|X) - \varphi(X))|X])$$

$\mathbb{E}(Y|X)$ est en fonction de X

$\varphi(X)$ est en fonction de X

$\mathbb{E}((Y - \varphi(X)))$ est en fonction de X

$$\mathbb{E}[\mathbb{E}((Y|X) - \varphi(X))^2|X] = \mathbb{E}(\mathbb{E}((Y|X) - \varphi(X))^2)$$

$$\mathbb{E}(\mathbb{E}(Y - \mathbb{E}(Y|X)))(\mathbb{E}((Y|X) - \varphi(X))|X) = \mathbb{E}(\mathbb{E}((Y|X) - \varphi(X))\mathbb{E}(\mathbb{E}(Y - \mathbb{E}(Y|X))|X))$$

$$= \mathbb{E}(Y - \mathbb{E}(Y|X)|X)$$

$$= \mathbb{E}(Y|X) - \mathbb{E}(Y|X) = 0$$

$$\|Y - \varphi(X)\|^2 = \|Y - \mathbb{E}(Y|X)\|^2 + b \text{ ou } b \geq 0$$

3.2 Equations linéaires de prédiction

On considère les prédictions linéaires de la série étudiée. Cela signifie que, si on observe les réalisations X_1, \dots, X_t de la série (X_t) , on cherche à prédire la valeur X_{T+h} (où h est un nombre positif fixé) à partir d'une combinaison linéaire des observations. Ainsi, en notant \hat{X}_{T+h} la valeur prévue au temps $t+h$ de (X_t) on étudie les prédictions qui s'écrivent sous la forme :

$$\hat{X}_{T+h} = k_0 + k_1 X_T + \dots + k_T X_1. \quad (3.1)$$

Les coefficients k_0, \dots, k_T dans cette expression sont choisis de telle sorte que la prédiction \hat{X}_{T+h} soit la plus proche possible de la véritable réalisation X_{T+h} du processus en $T+h$. Pour mesurer cette erreur, considérons l'erreur quadratique moyenne

$$S(k) = E(X_{T+h} - k_0 - k_1 X_T, \dots, k_T X_1)^2 \quad (3.2)$$

Les coefficients $k = (k_1, \dots, k_T)'$ dans (3.2) seront choisis de telle sorte qu'ils minimisent l'erreur $S(k)$. Ces coefficients forment donc la solution d'un système de T équations correspondantes à chaque dérivée partielle de $S(k)$. Ce système peut s'écrire de la forme suivante

$$E \left[X_{T+h} - k_0 - \sum_{i=1}^T k_i X_{T+1-i} \right] = 0$$

$$E \left[\left(X_{T+h} - k_0 - \sum_{i=1}^T k_i X_{T+1-i} \right) X_{T+1-j} \right] = 0 \quad j = 1, \dots, T.$$

La première équation donne $k_0 = 0$, qui correspond à l'hypothèse que le processus est à moyenne nulle. Quant aux $T-1$ équations suivantes, elles peuvent se réécrire (avec $k_0 = 0$):

$$\rho_{h-1+j} - \sum_{i=1}^T k_i \rho_{|i-j|} = 0 \quad j = 1, \dots, T,$$

ou encore, en notation matricielle:

$$\mathbf{R}_T k_T = \rho_T(h), \quad (3.3)$$

où $k_T = (k_1, \dots, k_T)'$, $\rho_T(h) = (\rho_h, \dots, \rho_{h+T-1})$ et est la matrice $T \times T$ de covariance dont l'élément (i, j)

$$(\mathbf{R}_T k_T)_{i,j} = \rho_{|i-j|} \quad i, j = 1, \dots, T.$$

L'équation (3.3) résout en principe le problème de la prédiction. Connaissant la fonction d'autocovariance du processus (qui, en pratique, devra être estimée), la prédiction \hat{X}_{T+h} est donnée par la relation (3.4) avec des coefficients k tels que

$$k_T = (\mathbf{R}_T)^{-1} \rho_T(h),$$

L'inverse de la matrice \mathbf{R}_T existe si $\rho_0 > 0$ et $\rho_h \rightarrow 0$ pour $|h| \rightarrow \infty$ (ce qui est vérifié pour un processus linéaire).

3.2.1 Algorithme de Durbin-Levinson

Même si, formellement, nous avons trouvé les équations de prédiction, il reste à inverser une matrice $T \times T$ pour trouver les coefficients de (3.2). Cette inversion est, pour un processus linéaire, théoriquement toujours possible mais parfois difficile et coûteuse en temps, surtout si le nombre T d'observations disponibles est élevé.

Dans cette perspective, il est utile de développer des algorithmes qui décrivent une méthode d'inversion plus rapide.

Le lecteur aura également remarqué la similitude formelle entre les équations des coefficients de prédiction (3.3) et les équations de Yule-Walker. Ces deux systèmes sont résolus par l'inversion de la matrice de covariance qui, dans le premier cas, est de rang T et dans le deuxième, de rang p . En conséquence, les méthodes algorithmiques que nous allons décrire maintenant pourront également être appliquées pour estimer les coefficients d'un modèle autorégressif par les équations de Yule-Walker.

L'idée de l'algorithme de Durbin-Levinson est que, pour prédire la valeur de la série au temps $T+h$, on pourrait utiliser une prédiction de la série au $T+h-1$. En remontant ainsi dans l'intervalle de temps de prédiction, cet algorithme propose le schéma de prédiction suivant: à partir des T observations, on évalue X_{t+1} . A partir des T observations auxquelles on ajoute \hat{X}_{T+1} , on évalue \hat{X}_{T+2} , et ainsi de suite jusqu'à \hat{X}_{T+h} pour l'horizon h . Pour la première étape de cette démarche, on trouve, en notant $\phi_{Ti} = k_i$ dans (3.2):

$$\hat{X}_{T+1} = \phi_{T1}X_T + \dots + \phi_{TT}X_1. \quad (3.4)$$

L'erreur quadratique moyenne de prédiction est égale à:

$$v_T = E(X_{T+1} - \hat{X}_{T+1})^2 \quad (3.5)$$

$$= \rho_0 - \phi'_T \rho_T \quad (3.6)$$

où $\phi_T = (\phi_{T1}, \dots, \phi_{TT})'$.

L'algorithme de Durbin-Levinson, énoncé ci-dessous, permet d'exprimer les coefficients ϕ_T en fonction des coefficients ϕ_{T-i} . Ainsi, pour prédire \hat{X}_{T+1} , on utilisera les coefficients ϕ_T trouvés pour cet algorithme. Récursivement, pour trouver \hat{X}_{T+2} , on aura besoin des coefficients ϕ_{T+1} que l'on calcule à partir de ϕ_T , et ainsi de suite.

Proposition 5. (Brockwell and Davis 1991)

Si (X_t) est un processus stationnaire de moyenne zéro et de fonction d'autocovariance ρ_h telle que $\rho_0 > 0$ et $\rho_h \rightarrow 0$ lorsque $h \rightarrow \infty$, alors les coefficients ϕ_{Tj} et les erreurs

quadratiques moyennes de prédiction h_T définis par (3.4) et (3.6) satisfont l'algorithme suivant:

- Étape 1: $\phi_{11} = \rho_1/\rho_0$, $v_0 = \rho_0$

- Étape T :

1. $\phi_{TT} = \left[\rho_T - \sum_{j=1}^{T-1} \phi_{T-1,j} \rho_{T-j} \right] v_{T-1}^{-1}$
2.
$$\begin{pmatrix} \phi_{T1} \\ \vdots \\ \phi_{T,T-1} \end{pmatrix} = \begin{pmatrix} \phi_{T-1,1} \\ \vdots \\ \phi_{T-1,T-1} \end{pmatrix} - \phi_{TT} \begin{pmatrix} \phi_{T-1,T-1} \\ \vdots \\ \phi_{T-1,1} \end{pmatrix}$$
3. $v_T = v_{T-1}(1 - \phi_{TT}^2)$.

Démonstration (voir Brockwell and Davis 1991).

Il est bien clair qu'en remplaçant les autocovariances ρ_h par les acf empiriques $\hat{\rho}_h$ on trouve la réalisation en pratique, l'algorithme de Durbin-Levinson empirique avec des coefficients empiriques $(\hat{\phi}_{T1}, \dots, \hat{\phi}_{TT})$ et un estimateur de l'erreur de prédiction \hat{v}_T .

3.2.2 Prédiction d'un processus ARMA

3.2.3 La méthode de Box et Jenkins

La méthode de Box et Jenkins est un outil systématique qui permet :

- *De déterminer le meilleur modèle de type ARMA décrivant le processus stochastique d'une série observée ou d'une transformation stationnaire de celle-ci ;
- *D'estimer ce modèle ;
- *De l'utiliser pour extrapoler les valeurs de la série.

La méthodologie de Box et Jenkins comporte essentiellement cinq étapes :

- Étape 1:Transformation des données afin de stabiliser la variance (*log, sqrt, ...*) et différenciation des données pour les stationariser.
- Étape 2:Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.
- Étape 3:Estimation des paramètres du(des) modèle(s) sélectionné(s).
- Étape 4:Diagnostique et tests adéquatation du modèle.
- Étape 5:Prévision: La dernière étape consiste la prévision des valeurs futures à travers le modèle retenu.

La prédiction de la partie MA(q) d'un processus ARMA(p, q) nécessite un peu de réflexion supplémentaire car, comme pour le problème de l'estimation des coefficients de la partie moyenne mobile, cette partie n'est pas observable. De nouveau, on s'en sort en

utilisant les résidus $e_t(a_p, \theta q)$; comme définis dans l'équation (2.10), pour rappel:

$$e_t(a_p, \theta q) = X_t - \sum_{k=1}^p \alpha_k X_{t-k} - \sum_{j=1}^q \theta_j e_{t-j}(a_p, \theta q) \quad (3.7)$$

avec $e_t(a_p, \theta q) := 0, t \leq 0$, et $X_t := 0, t \leq 0$.

D'après le paragraphe précédent, la partie $AR(p)$ est prédite par

$$\hat{X}_{T+1} = \sum_{k=1}^p \alpha_k X_{T+1-k} \quad (3.8)$$

Alors, pour un $ARMA(p, q)$ on utilise comme prédicteur

$$\hat{X}_{T+1} = \sum_{k=1}^p \alpha_k X_{T+1-k} + \sum_{j=1}^q \theta_j e_{T+1-j}(a_p, \theta q) \quad (3.9)$$

puisque les $\{\epsilon_t\}$ ne sont pas observables.

L'erreur de prédiction se comporte comme celle de la prédiction de la partie $AR(p)$: $E[(X_{T+1} - \hat{X}_{T+1})^2] \approx \sigma_\epsilon^2$, si T est grand.

En pratique, on estime le vecteur des coefficients $e_t(a_p, \theta q)$ par les estimateurs du maximum de vraisemblance $(\hat{\alpha}, \hat{\theta})$ et σ_ϵ^2 par: $T^{-1} \sum_t e_t^2(\hat{\alpha}_p, \hat{\theta}_q)$, alors que

$$\hat{X}_{T+1} = \sum_{k=1}^p \hat{\alpha}_k X_{T+1-k} + \sum_{j=1}^q \hat{\theta}_j e_{T+1-j}.$$

Pour une prédiction d'horizon $h > 1$ il faut remplacer les valeurs inconnues de $X_{T+h-1}, \dots, X_{T+1}$ par les prédictions précédentes $\hat{X}_{T+h-1}, \dots, \hat{X}_{T+1}$, avec $e_{T+h-1} = \dots = e_{T+1} = 0$.

3.2.4 L'erreur de prédiction

Notons \hat{X}_{T+h} la prévision de X_{n+h} sur les observations $\{X_1, X_2, \dots, X_T\}$.

Pour un $AR(p)$ causal,

$$X_t = \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + \epsilon_t, \quad (3.10)$$

Les résidus vérifient, $Cov(\epsilon_t, X_s) = 0$ si $s < t$. Supposons que $T > p$. Alors,

$$\begin{aligned} \hat{X}_{T+1} &= \theta_1 X_T + \dots + \theta_p X_{T+1-p} \\ X_{T+1} - \hat{X}_{T+1} &= \epsilon_t. \end{aligned}$$

d'où

$$E[(X_{T+1} - \hat{X}_{T+1})^2] = E[(\epsilon_t)^2] = Var(\epsilon_t) = \sigma_\epsilon^2.$$

On obtient de façon récursive les prévisions et erreurs de prévision a distance $h = 2, 3, \dots$
 Par exemple:

$$\begin{aligned}\hat{X}_{T+2} &= \theta_1 \hat{X}_{T+1} + \theta_2 X_T + \dots + \theta_p X_{T+2-p} \\ X_{T+2} - \hat{X}_{T+2} &= \theta_1 (X_{T+1} - \hat{X}_{T+1}) + \epsilon_{T+2} = \theta_1 \epsilon_{T+1} + \epsilon_{T+2}.\end{aligned}$$

donc

$$E[(X_{T+2} - \hat{X}_{T+2})^2] = Var(\theta_1 \epsilon_{T+1} + \epsilon_{T+2}) = (\theta_1 + 1) \sigma_\epsilon^2$$

Soit un processus ARMA(p, q) causal et inversible,

$$X_t = \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad (3.11)$$

avec $Cov(\epsilon_t, X_s) = 0$. D'après le théorème 5.5.1. de Brockwell and Davis (1991), on a:

$$X_t = \sum_{i=0}^{\infty} \pi_i \epsilon_{t-i} \quad (3.12)$$

Soit \hat{X}_{T+h} le meilleur prédicteur linéaire de X_{T+h} basé sur le passé infini (X_T, X_{T-1}, \dots) (voir Palma 2007 ou Brockwell and Davis 1991, Corollary 5.1.1.). Pour $h > 0$ on peut écrire

$$\hat{X}_{T+h} = E(X_{T+h} | X_T, \dots) = \sum_{i=1}^{\infty} \pi_i E(\epsilon_{T+h-i} | X_T, \dots) = \sum_{i=h}^{\infty} \pi_i \epsilon_{T+h-i}.$$

alors

$$X_{T+h} - \hat{X}_{T+h} = \sum_{i=0}^{h-1} \pi_i \epsilon_{T+h-i}.$$

Par conséquent, l'erreur de prévision s'écrit:

$$E[(X_{T+h} - \hat{X}_{T+h})^2] = Var\left(\sum_{i=0}^{h-1} \pi_i \epsilon_{T+h-i}\right) = \sigma_\epsilon^2 \sum_{i=0}^{h-1} \pi_i.$$

Donc on déduit que l'erreur de prévision augmente avec l'horizon h .

3.3 Prévision par lissage exponentiel

Introduite par Holt en 1958 puis par Winters en 1960. Les méthodes de lissage constituent l'ensemble des techniques empiriques de prévision qui accordent plus au moins d'importance aux valeurs du passé d'une série temporelle.

3.3.1 Lissage exponentiel simple

On dispose de n observations X_1, X_2, \dots, X_n d'une série chronologique $(X_t)_t$ et on voudrait prédire les valeurs X_{n+1}, \dots, X_{n+h} , avec h appelé *horizon*.

Le lissage exponentiel peut être considéré comme une régression pondérée.

Supposons qu'on s'intéresse à une série $X(t) = \psi(t) + \epsilon(t)$, $t = 1, \dots, n$, avec $\epsilon(t)$ est un bruit blanc et $\psi(t)$ est un polynôme de degré inférieur ou égal à 2, dont les coefficients varient lentement.

Objectif : Prédire la série aux temps $n + 1, \dots, n + h$.

Moyen : Estimer la tendance au voisinage de point n par un polynôme de degré inférieur à 2.

En lissage exponentiel, on mesure l'intensité du voisinage de n par un pas qui décroît exponentiellement en s'écartant de n . le poids au point n , sera pris de la forme:

$$W_{n-t} = \alpha(1 - \alpha)^{n-t}$$

avec $0 < \alpha < 1$. α étant un paramètre à estimer (appelé constante de lissage). Le lissage exponentiel simple correspond à $\psi(t) = C$ au voisinage de n .

au voisinage de n l'ajustement par régression pondérée d'une constante à X_n consiste à prendre comme estimation de C la constante qui minimise

$$\sum_{i=1}^n \alpha(1 - \alpha)^{n-i} (X_i - C)^2 = f(C)$$

f admet un extremum si:

$$\begin{aligned} f'(C) = 0 &\Leftrightarrow \left(\sum_{i=1}^n \alpha(1 - \alpha)^{n-i} (X_i - \hat{C})^2 \right)' = 0 \\ &\Leftrightarrow \sum_{i=1}^n (1 - \alpha)^{n-i} X_i - \hat{C} \sum_{i=1}^n (1 - \alpha)^{n-i} = 0 \end{aligned}$$

D'où

$$\hat{C} = \frac{\sum_{i=1}^n (1 - \alpha)^{n-i} X_i}{\sum_{i=1}^n (1 - \alpha)^{n-i}}$$

et on a:

$$\sum_{i=1}^n (1 - \alpha)^{n-i} = \frac{1 - (1 - \alpha)^n}{\alpha}$$

Par conséquent:

$$\hat{C} = \frac{\sum_{i=1}^n \alpha(1 - \alpha)^{n-i} X_i}{1 - (1 - \alpha)^n}$$

Supposons que l'on dispose d'un très grand nombre d'observations alors:

$$\hat{C} \simeq \sum_{i=1}^n \alpha(1 - \alpha)^{n-i} X_i = \hat{C}(n)$$

$\hat{C}(n)$ ne dépend pas de l'horizon h .

La formule de mise à jour:

On a:

$$(1 - \alpha)\hat{C}(n) = \sum_{i=1}^n \alpha(1 - \alpha)^{n-i} X_i$$

on pose $n - i = k$

$$\sum_{i=1}^n \alpha(1 - \alpha)^{n-i} (X_i - C)^2 = \sum_{k=0}^{n-1} \alpha(1 - \alpha)^k (X_{n-k} - C)^2$$

$$\hat{C}(n) = \alpha \sum_{k=1}^{n-1} (1 - \alpha)^k X_{n-k}$$

$$(1 - \alpha)\hat{C}(n - 1) = \alpha \sum_{k=0}^{n-1} (1 - \alpha)^{k+1} X_{n-k}$$

$$(1 - \alpha)\hat{C}(n - 1) = \alpha \sum_{k=0}^{n-2} (1 - \alpha)^{k+1} X_{n-k-h}$$

Posons $j = k + 1$

$$\begin{aligned} (1 - \alpha)\hat{C}(n - 1) &= \alpha \sum_{j=1}^{n-1} (1 - \alpha)^j X_{n-j} \\ &= \alpha \sum_{j=1}^{n-1} (1 - \alpha)^j X_{n-j} - \alpha X_n \\ &= \hat{C}(n) - \alpha X_n. \end{aligned}$$

Par conséquent:

$$\hat{C}(n) = (1 - \alpha)\hat{C}(n - 1) - \alpha X_n. \quad (3.13)$$

la prédiction \hat{X}_{n+1} est comme une moyenne pondérée entre la prévision au temps $n - 1$ est la dernière valeur. pour mettre à jour l'équation (3.13) il faut une valeur initiale qui est la valeur $\hat{C}(1) = X_1$.

Choix de la constante de lissage

pour un horizon h on cherche la valeur de α qui minimise

$$\sum_{k=1}^{n-h} \left(X_{n+h} - \hat{X}(h) \right)^2 = \sum_{k=1}^{n-h} \left(X_{n+h} - \sum_{j=0}^{n-h} (1 - \alpha)^j X_{n-j} \right)^2$$

3.3.2 Lissage exponentiel double

On suppose qu'au voisinage de n , la série peut être ajustée par une droite, c'est-à-dire:

$$X_n = a_1 + (t - n)a_2. \quad (3.14)$$

avec t appartient au voisinage de n . Cela va nous permettre d'écrire

$$\hat{X}_n = \hat{a}_1(n) + h\hat{a}_2 = \hat{X}_{n+h}. \quad (3.15)$$

Les coefficients \hat{a}_1 et \hat{a}_2 sont solution de :

$$\sum_{j=0}^{n-1} (1 - \alpha)^j (X_{n-j} - (a_1 + a_2 j))^2 = C(a_1, a_2)$$

On a :

$$\frac{\partial C(a_1, a_2)}{\partial a_1} = 0 \quad \text{et} \quad \frac{\partial C(a_1, a_2)}{\partial a_2} = 0$$

qui correspond à :

$$\begin{aligned} \sum_{j=0}^{n-1} (1 - \alpha)^j (X_{n-j} - (a_1 + a_2 j)) &= 0 \\ \sum_{j=0}^{n-1} (1 - \alpha)^j (X_{n-j} - (a_1 + a_2 j)) &= 0 \\ \sum_{j=0}^{n-1} (1 - \alpha)^j X_{n-j} - \hat{a}_1 \sum_{j=0}^{n-1} (1 - \alpha)^j + \hat{a}_2 \sum_{j=0}^{n-1} j(1 - \alpha)^j &= 0 \\ \sum_{j=0}^{n-1} (1 - \alpha)^j X_{n-j} - \hat{a}_1 \sum_{j=0}^{n-1} j(1 - \alpha)^j + \hat{a}_2 \sum_{j=0}^{n-1} j^2(1 - \alpha)^j &= 0 \end{aligned}$$

Supposons que n est relativement grand, alors :

$$\sum_{j=0}^{n-1} (1 - \alpha)^j \simeq \frac{1}{\alpha}$$

et

$$\sum_{j=0}^{n-1} j(1 - \alpha)^j \simeq \frac{1 - \alpha}{\alpha^2}$$

ce qui donne

$$\sum_{j=0}^{n-1} (1 - \alpha)^j X_{n-j} - \frac{\hat{a}_1}{\alpha} + \frac{\hat{a}_2(1 - \alpha)}{\alpha^2} = 0 \quad (3.16)$$

$$\sum_{j=0}^{n-1} (1 - \alpha)^j X_{n-j} - \frac{\hat{a}_1(1 - \alpha)}{\alpha^2} + \frac{\hat{a}_2(1 - \alpha)(2 - \alpha)}{\alpha^3} = 0 \quad (3.17)$$

$$\sum_{j=0}^{n-1} \alpha^3 (1-\alpha)^j X_{n-j} - \hat{a}_1 \alpha (1-\alpha) \hat{a}_2 (1-\alpha) (2-\alpha) = 0$$

posons

$$L_1(n) = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j X_{n-j}$$

$$\begin{aligned} L_2(n) &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j) \\ &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j \left(\sum_{i=0}^{n-1} \alpha (1-\alpha)^i X_{n-j-i} \right) \\ &= \alpha^2 \sum_{j=0}^{n-1} \sum_{i=0}^{n-1} (1-\alpha)^{i+j} X_{n-j-i} \\ &= \alpha^2 \sum_{k=0}^{n-1} k (1-\alpha)^k X_{n-k} + \alpha \left(\sum_{k=0}^{n-1} k \alpha (1-\alpha)^k X_{n-k} \right) \\ &= \alpha^2 \sum_{k=0}^{n-1} k (1-\alpha)^k X_{n-k} + \alpha L_1(n) \end{aligned}$$

Les équations (3.16) et (3.17) devient:

$$\begin{cases} L_1(n) = \hat{a}_1 + \hat{a}_2 \frac{(1-\alpha)}{\alpha} = 0 \\ L_1(n) = \alpha L_1(n) + \hat{a}_1 (1-\alpha) + \hat{a}_2 \frac{(1-\alpha)(2-\alpha)}{\alpha^3} = 0 \end{cases} \quad (3.18)$$

La résolution de ce système nous donne:

$$\begin{cases} \hat{a}_1(n) = 2L_1(n) - L_2(n) \\ \hat{a}_2(n) = \frac{\alpha}{1-\alpha} (L_1(n) + L_2(n)) \end{cases} \quad (3.19)$$

La formule de mise à jour

$$\begin{aligned} L_1(n) &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j X_{n-j} \\ (1-\alpha)L_1(n) &= \alpha \sum_{j=0}^{n-2} (1-\alpha)^{j+1} X_{n-j} \end{aligned}$$

$$\begin{aligned}
 (1 - \alpha)L_1(n - 1) &= \alpha(1 - \alpha) \sum_{j=0}^{n-2} (1 - \alpha)^j X_{n-j-1} \\
 &= \alpha \sum_{j=0}^{n-2} (1 - \alpha)^{j+1} X_{n-j-1} \\
 &= \alpha \sum_{k=1}^{n-1} (1 - \alpha)^k X_{n-k} \\
 &= \alpha \left(\sum_{k=1}^{n-1} (1 - \alpha)^k X_{n-k} - X_n \right) \\
 &= L_1(n) - \alpha X_n
 \end{aligned}$$

d'où

$$L_1(n) = (1 - \alpha)L_1(n - 1) + \alpha X_n \quad (3.20)$$

de même manière, on montre que:

$$\begin{aligned}
 L_2(n) &= \alpha L_1(n) + (1 - \alpha)L_2(n - 1) \\
 &= \alpha^2 X_n + \alpha(1 - \alpha)L_1(n - 1) + (1 - \alpha)L_2(n - 1)
 \end{aligned}$$

ce que nous donne:

$$\hat{a}_1(n) = \alpha_n^2 + (1 - \alpha)^2 (\hat{a}_1(n - 1) + \hat{a}_2(n - 1))$$

et les solution adaptatives sont:

$$\hat{a}_1(n) = \hat{a}_1(n - 1) + \hat{a}_2(n - 1) + (1 - (1 - \alpha)^2)(X_n - \hat{X}_{n-1}(1)). \quad (3.21)$$

et

$$\hat{a}_2(n) = \hat{a}_1(n - 1) + (1 - (1 - \alpha)^2)(X_n - \hat{X}_{n-1}(1)). \quad (3.22)$$

Pour les valeurs initiales, on prend

$$\hat{a}_1(1) = X_2 \quad \text{et} \quad \hat{a}_2(1) = X_2 - X_1.$$

3.3.3 Méthode de Holt-Winters

a - Méthode non saisonnière

Comme la méthode de lissage exponentiel double, celle de Holt-Winters non saisonnière revient à estimer au voisinage de l'instant n une droite

$$X_n = a_1 + (t - n)a_2. \quad (3.23)$$

La prévision prend la forme

$$X_{n+h} = \hat{a}_1 + \hat{a}_2 h. \quad (3.24)$$

La variante par rapport à la méthode de lissage exponentiel double est au niveau des formules de mise à jour dans l'estimation des paramètres a_1 et a_2 .

Soient deux constantes de lissages $0 < \alpha < 1$ et $0 < \beta < 1$.

Les formules de mise à jour sont :

$$\hat{a}_1(n) = \alpha X_n + (1 + \alpha)[\hat{a}_1(n-1) + \hat{a}_2(n-1)],$$

$$\hat{a}_2(n) = \beta[\hat{a}_1 - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1).$$

Remarque 6.

- l'introduction de deux constantes rend la méthode plus souple que le lissage exponentiel double : la constante α joue un rôle dans l'estimation de l'ordonnée à l'origine de la droite, a_1 , et la constante β dans celle de la pente de la droite, a_2 .
- si α et β sont petits le lissage est important car on tient compte du passé lointain.

b- Méthode saisonnière additive

On cherche maintenant à ajuster au voisinage de l'instant n une droite d'équation

$$\hat{X}_n = a_1 + (t - n)a_2 + S_n \quad (3.25)$$

où S_n est une composante périodique de période p .

Les formules récursives de mise à jour sont :

$$\hat{a}_1(n) = \alpha(X_n - \hat{S}_{n-p}) + (1 - \alpha) + [\hat{a}_1(n-1) + \hat{a}_2(n-1)]$$

$$\hat{a}_2(n) = \beta[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1)$$

$$\hat{S}(n) = \gamma[X_n - \hat{a}_1(n)] + (1 - \gamma)\hat{S}(n-p)$$

Les prévisions sont de la forme :

$$\hat{X}_{n+h} = \hat{a}_1 + \hat{a}_2 h + \hat{S}_{n+h-p} \quad 1 \leq h \leq p,$$

$$\hat{X}_{n+h} = \hat{a}_1 + \hat{a}_2 h + \hat{S}_{n+h-2p} \quad p+1 \leq h \leq 2p,$$

et ainsi de suite pour $h = 2p$.

Les trois constantes de lissages, α , β et γ ont le même effet que précédemment, plus elles sont petites et plus l'importance des données éloignées est significative. Elles agissent respectivement sur les paramètres a_1 , a_2 et S_n . Se référer à Gouriéroux et Monfort (1983) pour les valeurs d'initialisation.

c- Méthode saisonnière multiplicative

On ajuste au voisinage de l'instant n une droite d'équation

$$\hat{X}_n = (a_1 + (t - n)a_2)S_n \quad (3.26)$$

Les formules récursives de mise à jour sont :

$$\begin{aligned} \hat{a}_1(n) &= \alpha \frac{X_n}{\hat{S}_{n-p}} + (1 - \alpha)[\hat{a}_1(n-1) + \hat{a}_2(n-1)], \\ \hat{a}_2(n) &= \beta[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1), \\ \hat{S}(n) &= \gamma \frac{X_n}{\hat{a}_1(n)} + (1 - \gamma)\hat{S}_{n-p}. \end{aligned}$$

Les prévisions sont de la forme :

$$\begin{aligned} \hat{X}_{n+h} &= (\hat{a}_1 + \hat{a}_2 h)\hat{S}_{n+h-p} \quad 1 \leq h \leq p, \\ \hat{X}_{n+h} &= (\hat{a}_1 + \hat{a}_2 h)\hat{S}_{n+h-2p} \quad p+1 \leq h \leq 2p. \end{aligned}$$

3.4 Mise en œuvre sous R

3.4.1 Lissage exponentiel simple

La fonction `ets()` de `forecast` contient plusieurs méthodes de lissage exponentiel.

Exemple 5. Toujours avec l'exemple des prix du pétrole. Nous allons utiliser le lissage exponentiel simple pour prédire la série pour les 4 prochains mois.

```
> library(forecast)
> lissage=ets(y,model="ANN")
> summary(lissage);
ETS(A,N,N)
Call:
ets(y = y, model = "ANN")
Smoothing parameters:
  alpha = 0.9999
Initial states:
  l = 21.0363
sigma: 4.4191
      AIC      AICc      BIC
2548.706 2548.748 2556.074
Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.3088521 4.419127 2.837213 0.1880816 6.608954 0.09287177
```

- `alpha` est l'estimation de α .
- `l` représente l'estimation pour l'état initial.
- `sigma` représente l'écart type du bruit ϵ_t .
- Les dernières lignes représentent les critères d'information et les mesures d'erreur intra-échantillon (voir Aragon 2011 et Hyndman et al. 2008).

On remarque que `alpha` est très proche de 1, donc on déduit que la prédiction avec un lissage exponentiel ne pourra pas être une solution envisageable pour cette série.

Exemple 6. Soit la série `fmsales` (disponible dans le package `expsmooth`) qui représente les ventes d'un produit sur 62 semaines à partir du début de 2003.

Le code **R** pour le lissage simple est comme suit:

```
> require(expsmooth)
> y=ts(fmsales,start=2003,frequency=12)
> lissage=ets(y,model="ANN")
> summary(lissage)
ETS(A,N,N)
Call:
ets(y = y, model = "ANN")
Smoothing parameters:
  alpha = 0.7315
Initial states:
  l = 23.4661
sigma:  3.5496
      AIC      AICc      BIC
416.9693 417.1727 421.2236
Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.201251 3.549585 2.35029 0.09814127 6.949485 0.5799638
```

L'estimation des paramètres est donc $\hat{\alpha} = 0.7315$ et $l_0 = 23.4661$.

Pour voir le lissage sur un graphe, nous utilisons le code suivant:

```
x=numeric(1); x[1]=23.4661; n=length(y); a=0.7315;
for(i in 1:n){
x[i+1]=a*y[i]+(1-a)*x[i]}
x=ts(x,start=2003, frequency=12)
y=ts(fmsales,start=2003, frequency=12)
plot(y,type="l",ylab="Prix")
lines(x,type="l",col=2,lty=2,lwd=1)
```

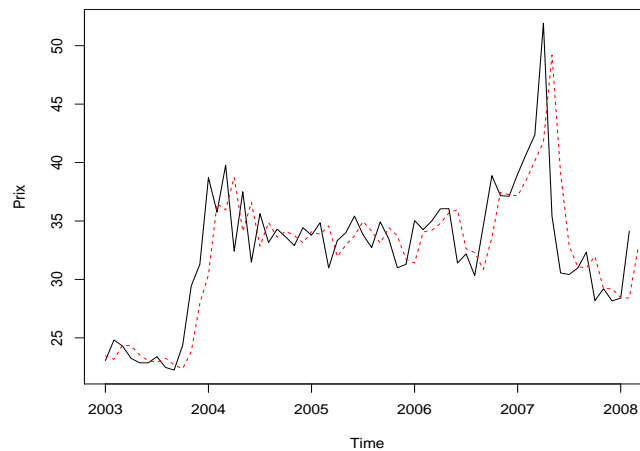


FIG. 3.1 – La série *fmsales* et son lissage exponentiel simple en rouge.

Pour la prédiction, on utilise simplement l'instruction `predict()` en indiquant l'horizon,

```
> predict(lissage,4)
      Point Forecast   Lo 80   Hi 80   Lo 95   Hi 95
Mar 2008      32.5938 28.04483 37.14278 25.63675 39.55086
Apr 2008      32.5938 26.95760 38.23001 23.97398 41.21363
May 2008      32.5938 26.04855 39.13906 22.58370 42.60391
Jun 2008      32.5938 25.25119 39.93642 21.36425 43.82336
```

- Les valeurs Lo 80 et Hi 80 représentent la borne inférieure et la borne supérieure de l'intervalle de confiance à 80% respectivement.

- Même chose pour Lo 95 et Hi 95 où la l'intervalle de confiance est à 95%.

Pour la prédiction, on voit qu'elle est bien constante, mais les bornes deviennent plus grandes pour chaque prochain mois.

3.4.2 Lissage exponentiel double

La même fonction est utilisée dans le lissage exponentiel double, en changeant seulement ANN par AAN.

Nous appliquons la fonction `ets()` pour la série `fmsales`,

```
> require(expsmooth)
> y=ts(fmsales,start=2003,frequency=12)
> lis_double=ets(y,model="AAN")
> summary(lis_double)
```

```
ETS(A,A,N)
```

```
Call:
```

```
ets(y = y, model = "AAN")
```

```
Smoothing parameters:
```

```
alpha = 0.7382
```

```
beta = 1e-04
```

```
Initial states:
```

```
l = 23.7075
```

```
b = 0.1667
```

```
sigma: 3.5447
```

```
      AIC      AICc      BIC
```

```
420.7979 421.4997 429.3065
```

```
Training set error measures:
```

| | ME | RMSE | MAE | MPE | MAPE | MASE |
|--------------|-----------|----------|----------|------------|----------|-----------|
| Training set | -0.030245 | 3.544682 | 2.367602 | -0.6372273 | 7.058176 | 0.5842358 |

La prédiction des 4 prochains mois se fait avec,

```
> predict(lis_double,4)
```

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|----------|----------------|----------|----------|----------|----------|
| Mar 2008 | 32.85680 | 28.31411 | 37.39950 | 25.90935 | 39.80425 |
| Apr 2008 | 33.02335 | 27.37640 | 38.67030 | 24.38708 | 41.65962 |
| May 2008 | 33.18989 | 26.62155 | 39.75824 | 23.14448 | 43.23531 |
| Jun 2008 | 33.35644 | 25.98072 | 40.73216 | 22.07625 | 44.63663 |

Conclusion

La modélisation et la prédiction sont deux outils indispensables pour de nombreuses disciplines. La science de nos jours ne repose plus sur l'intuition pour comprendre ou prédire le futur. La prédiction basée sur un appui solide (*i.e.* un bon modèle) peut éviter beaucoup de problèmes et crises, notamment dans le secteur économique. Ainsi, parfois elle nous permet de prendre les mesures convenables avant qu'il soit trop tard, c'est le cas, par exemple, des changements climatiques où des propagations des maladies.

Dans ce mémoire, nous avons vu de près quelques méthodes qui permettent de modéliser et prédire une série chronologique. Quelques éléments de base sont abordés sur les modèles ARMA tels que: les méthodes d'estimation des paramètres, validation du modèle, prédiction, etc. Nous avons discuté également le lissage exponentiel qui est un outil très simple permettant de faire des prévisions loin des complexités des modèles ARMA. Cependant, les formules de mise à jour est un avantage qui ne se trouve pas dans les modèles ARMA et qui leur donne une place considérable dans la pratique. En effet, l'arrivée d'une nouvelle valeur nous oblige de refaire les calculs pour les modèles ARMA, or que dans le lissage exponentiel cela n'est pas un problème grâce à ces formules. Mais cet avantage n'est pas un argument de préférence, car les modèles ARMA sont les plus élaborées et qui sont conçues généralement pour les situations compliquées. Ainsi, nous avons complété cette étude théorique par une mise en pratique sous **R**. A travers des exemples, nous avons présenté quelques fonctions de base importantes qui sert à modéliser et à prédire des séries temporelles.

On pourrait faire une extension de ce travail, aux prévisions dans des séries chronologiques à valeurs entières.

Bibliographie

- [1] Aragon, Y., (2011). *Séries temporelles avec R, méthodes et cas*. Springer-Verlag France.
- [2] Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics 31, 307-327.
- [3] Box, G.E.P. and G.M. Jenkins (1970). *Time Series Analysis, Forecasting, and Control*. Oakland, CA: Holden-Day.
- [4] Brockwell, P.J., and Davis, R. A. (1991). *Time Series : Theory and Methods* (Second ed.). New York: Springer.
- [5] Cryer, J.D., and Chan, K.S., (2008). *Time Series Analysis With Applications in R*. Springer.
- [6] Engle, R.F. (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*. Econometrica 50, 987-1008.
- [7] Durbin, J. and S.J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- [8] Fuller, W.A. (1976). *Introduction to Statistical Time Series*. New York: Wiley.
- [9] Fuller, W.A. (1995). *Introduction to Statistical Time Series*, 2nd ed. New York: Wiley.
- [10] Granger, C.W. and R. Joyeux (1980). *An introduction to long-memory time series models and fractional differencing*. J. Time Series Anal., 1, 15-29.
- [11] Gouriéroux, C., et Monfort, A., (1995). *Séries temporelles et modèles dynamiques*. Economica, 2 ed.
- [12] Hannan, E.J. (1970). *Multiple Time Series*. New York: Wiley.
- [13] Holt, S. J., and Withers, R. F. J. (1958). *Studies in enzyme cytochemistry. V. An appraisal of indigogenic reactions for esterase localization*. Proceedings of the Royal Society of London. Series B-Biological Sciences, 148(933), 520-532.
- [14] Hyndman, R.J., (2007). *forecast: Forecasting functions for time series*. URL: <http://CRAN.R-project.org>.
- [15] Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D., (2008). *Forecasting with exponential smoothing: the state space approach*. Springer.

-
- [16] Harvey, A.C. and P.H.J. Todd (1983). *Forecasting economic time series with structural and Box-Jenkins models: A case study*. J. Bus. Econ. Stat., 1, 299-307.
- [17] Harvey, A.C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- [18] Harvey, A.C. (1993). *Time Series Models*. Cambridge, MA: MIT Press.
- [19] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y., (1992). *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?* Journal of Econometrics, 54,159-178.
- [20] McQuarrie, A.D.R. and C.L. Tsai (1998). *Regression and Time Series Model Selection, Singapore*. World Scientific.
- [21] Mills, T.C. (1999). *The Econometric Modelling of Financial Time Series (Second ed.)*. Cambridge University Press.
- [22] Percival, D.B. and A.T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.
- [23] Priestley, M.B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. London: Academic Press.
- [24] Robert, C.P., et Casella G., (2011). *Méthodes de Monte Carlo avec R*. Springer-Verlag, France.
- [25] Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- [26] Shumway, R.H., and Stoffer, D.S, (2006). *Time series analysis and its applications with R examples*. Springer.
- [27] Taylor, S.J. (1986). *Modeling Financial Time Series. Chichester*. John Wiley & Sons.
- [28] Von Sachs, R., et Van Bellegem, S. (2005). STAT 2414 *Séries chronologiques*. STAT, 2414, 1.
- [29] Wei, W.W.S. (2005). *Time Series Analysis, 2nd ed*. Redwood City, CA: Addison-Wesley.
- [30] Winters, P. R. (1960). *Forecasting sales by exponentially weighted moving averages*. Management Science, 6(3), 324-342.