

Introduction aux séries temporelles

Cathy Maugis-Rabusseau

Last Updated on December, 2023

Table of contents

Préface	4
1 Introduction	5
1.1 Quelques exemples de séries temporelles	5
1.2 Quelles questions autour de l'étude d'une série temporelle ?	8
2 Tendances et saisonnalités	11
2.1 Décomposition d'une série temporelle	11
2.2 Estimation / élimination d'une tendance en l'absence de saisonnalité	12
2.2.1 Estimation de la tendance par moindres carrés	13
2.2.2 Estimation de la tendance par filtrage de moyenne mobile	15
2.2.3 Opérateurs retard et avance	15
2.2.4 Estimation de la tendance par lissage exponentiel	17
2.2.5 Elimination de la tendance par différenciation	20
2.3 Estimation / élimination de la tendance et de la saisonnalité	23
2.3.1 Estimation par moindres carrés	26
2.3.2 Estimation par Moyenne Mobile	27
2.3.3 Elimination de la saisonnalité par différenciation	34
3 Modélisation aléatoire des séries temporelles	36
3.1 Processus stochastiques	36
3.1.1 Définition	36
3.1.2 Premiers exemples de processus	37
3.2 Rappels sur l'espace L^2 et les processus du second ordre	40
3.2.1 Espace L^2	40
3.2.2 Convergence dans L^2	41
3.2.3 Projection orthogonale	42
3.3 Processus stationnaires	43
3.4 Fonctions d'autocovariance et d'autocorrélation	44
3.4.1 Fonction d'autocovariance (ACVF)	44
3.4.2 Fonction d'autocorrélation (ACF)	45
3.4.3 Exemples	45
3.4.4 CNS pour une fonction d'autocovariance	53
3.4.5 Matrice d'autocorrélation	54
3.4.6 Premiers pas vers les processus ARMA	55

3.5	Densité spectrale	57
4	Statistique des processus stationnaires du second ordre	63
4.1	Quelques estimateurs	63
4.1.1	Estimateur de la moyenne du processus stationnaire	63
4.1.2	Estimateur de la fonction d'auto-covariance / -corrélation	65
4.1.3	Estimateur de la matrice d'auto-covariance / -corrélation	66
4.1.4	Estimateur de la densité spectrale	67
4.2	Prévision linéaire optimale	72
4.2.1	Espaces linéaires engendrés par un processus du second ordre	72
4.2.2	Régression linéaire	74
4.2.3	Prévision linéaire optimale	75
4.2.4	Prévision linéaire optimale dans le cas d'un passé fini	76
4.2.5	Evolution des prévisions linéaires optimales en fonction de la taille de la mémoire	78
4.3	Autocorrélations partielles	81
4.4	Tests de blancheur d'un processus	84
5	Les modèles ARMA	88
5.1	Polynômes et séries en B	88
5.1.1	Définitions	88
5.1.2	Inversion de $I - \lambda B$	90
5.1.3	Inverse d'un polynôme en B	93
5.2	Processus AR	94
5.2.1	Définition	94
5.2.2	Un exemple instructif : le processus AR(1)	95
5.2.3	Propriétés	98
5.2.4	Liaisons temporelles	101
5.3	Processus MA	106
5.3.1	Définition	106
5.3.2	Propriétés	107
5.3.3	Liaisons temporelles	109
5.4	Processus ARMA	115
5.4.1	Définition	115
5.4.2	Ecritures MA(∞) et AR(∞)	116
5.4.3	Liaisons temporelles	117
5.5	Vers les processus ARIMA/ SARIMA	118
5.5.1	ARIMA	118
5.5.2	Sélection de modèle	120
5.5.3	SARIMA	123
	References	134

Préface

Ce polycopié a été rédigé dans le cadre de l'UF de Processus Stochastiques de 4ème année, spécialité Mathématiques Appliquées de l'INSA Toulouse.

Je tiens tout particulièrement à remercier Jean-Yves DAUXOIS qui a accepté de partager ses sources de cours pour mes premiers pas dans l'enseignement des séries temporelles !

1 Introduction

L'objectif de ce cours est de donner les bases pour modéliser l'évolution dans le temps, ici supposé discret, d'un phénomène aléatoire. On parle de **série temporelle**.

Nous allons ici donner quelques exemples de séries temporelles pour fixer les idées et introduire les questions qui peuvent être adressées autour de l'étude des séries temporelles.

1.1 Quelques exemples de séries temporelles

Les séries temporelles sont présentes dans de nombreux domaines d'applications. On donne ici quelques exemples illustratifs.

Exemple 1.1. On peut par exemple s'intéresser à l'évolution de la taille d'une population. La Figure 1.1 représente la population (en million d'habitants) en France entre 1846 à 1951 (à gauche) et aux USA entre 1790 et 1990 (à droite). Ces deux séries temporelles n'ont pas la même tendance générale.

Exemple 1.2. On donne ici un exemple bien connu en écologie. La Figure 1.2 représente le nombre de fourrures de lièvres (Hare) et de lynx échangées à la Compagnie de la Baie d'Hudson de 1845 à 1935. On constate la présence de cycles des populations de lynx (le prédateur) et du lièvre (sa proie). Ce jeu de données est disponible dans la librairie `fpp3` sous le nom de `pelt`.

Exemple 1.3. On peut aussi donner un exemple en climatologie. La Figure 1.3 représente l'évolution de la surface de la glace dans l'Arctique de 1972 à 2018. On peut observer sur cette série temporelle une périodicité et une tendance à décroître.

Exemple 1.4. Comme dernier exemple, on s'intéresse à la célèbre série temporelle *AirPassengers* donnant le nombre mensuel (en milliers) de passagers des lignes aériennes entre les années 1949 et 1960 (à gauche de la Figure 1.4). On peut observer une périodicité dans la série temporelle et une croissance exponentielle avec une augmentation de la variabilité. On peut transformer cette dernière série en considérant le logarithme népérien du nombre de passagers aériens (à droite de la Figure 1.4).

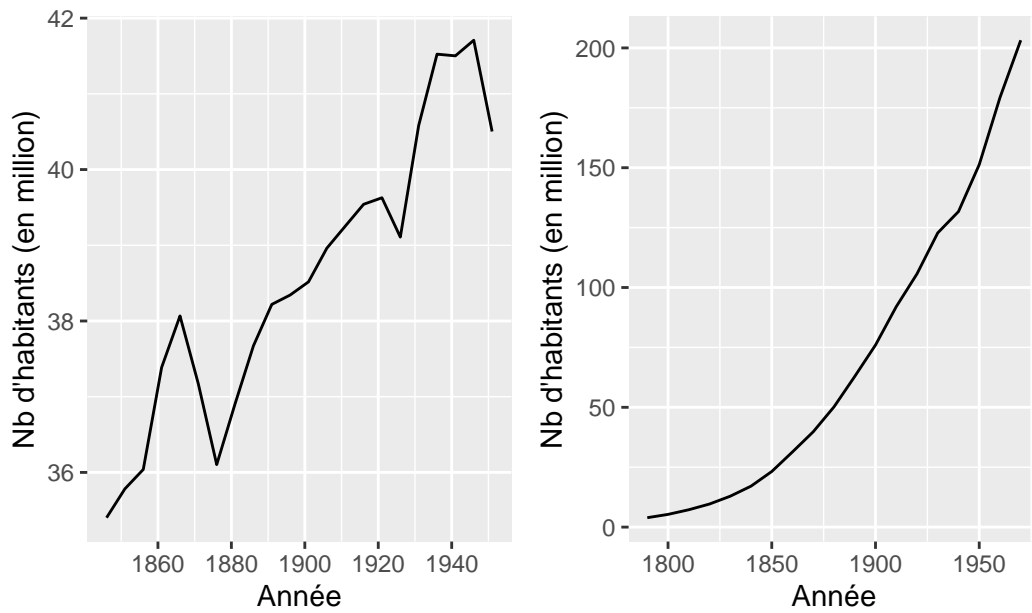


Figure 1.1: Taille de la population en France (à gauche) et aux USA (à droite).

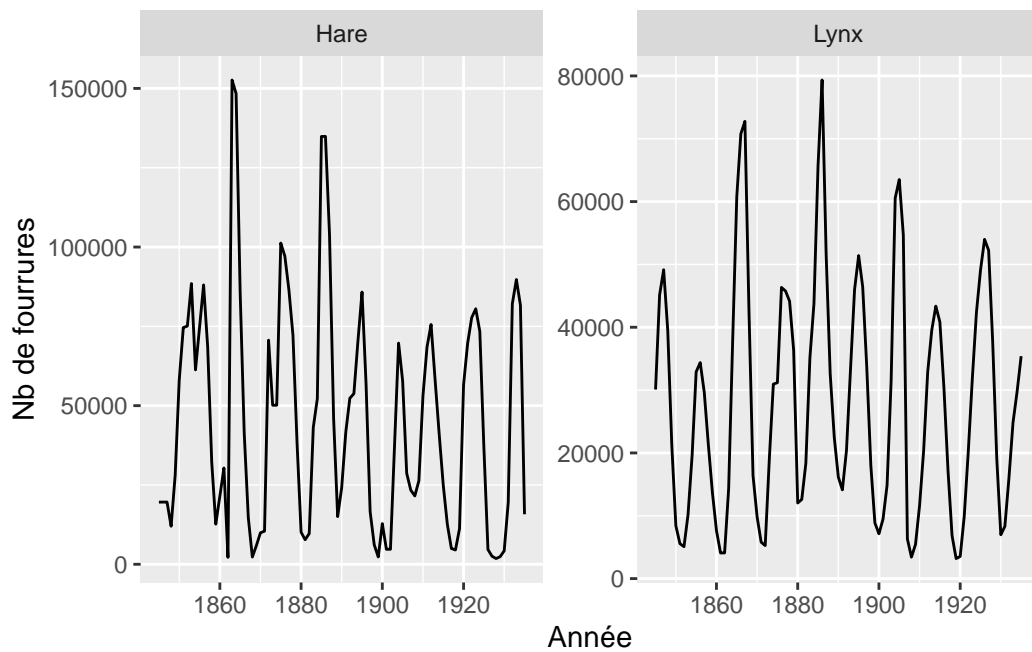


Figure 1.2: Nombre de fourrures échangées de lièvres et de lynx à la Compagnie de la Baie d'Hudson de 1845 à 1935.

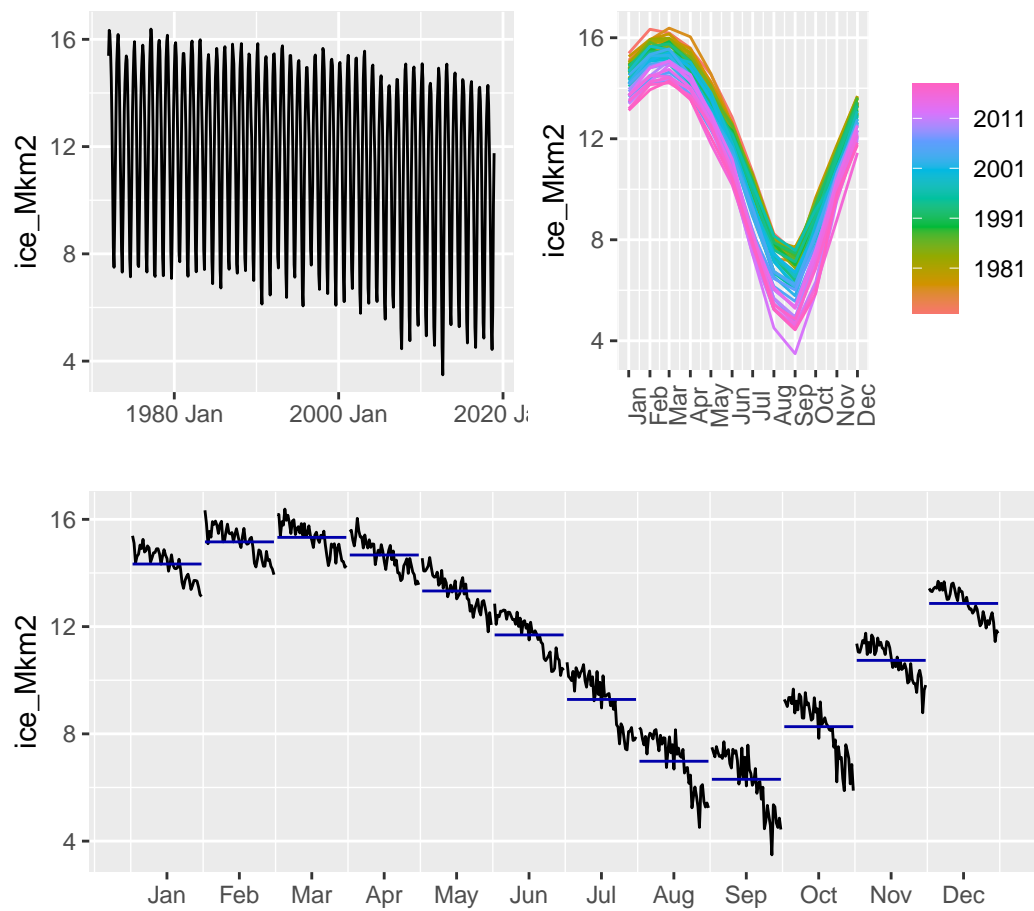


Figure 1.3: Evolution de la surface de la glace dans l'Arctique. En haut à gauche, la série temporelle. En haut à droite, chaque courbe représente l'évolution d'une année selon les différents mois. En bas, regroupement des mesures par mois.

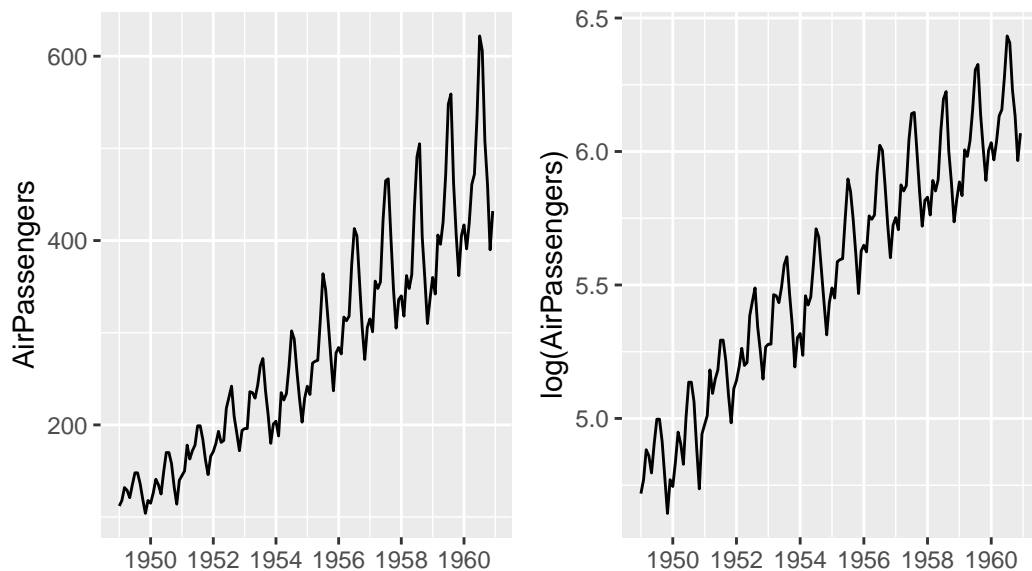


Figure 1.4: Nombre de passagers (en milliers) ayant emprunté les lignes aériennes de 1949 à 1960 à gauche. La série est log-transformée à droite.

Les exemples de séries temporelles ne manquent pas. On pourrait encore citer l'évolution de la température sur une période donnée, la concentration en polluants au cours du temps, le cours d'une action en finance, la consommation en électricité, l'évolution des recherches d'un mot sur internet, l'électrocardiogramme d'une personne en médecine,

Dans la suite de ce cours, on suppose que la série observée est une réalisation d'une suite de variables aléatoires. Il faut bien noter que **l'ordre est important** puisque l'on étudie un phénomène au cours du temps.

On note Y_t la valeur du phénomène au temps $t \in T$ où T est l'espace de **temps discret** (souvent $T = \mathbb{N}$ voire \mathbb{Z}). Le processus $(Y_t)_{t \in T}$ est alors appelé **série temporelle**.

1.2 Quelles questions autour de l'étude d'une série temporelle ?

Nous sommes tout d'abord confrontés au problème de la **modélisation** d'une série temporelle. On va chercher à ajuster un modèle qui décrit "au mieux" le comportement de la série temporelle. Ceci va nécessiter d'estimer les paramètres pour ajuster le modèle, tester son adéquation à la série temporelle étudiée, voire faire de la sélection de modèle entre plusieurs modèles en compétition.

Il est important de noter que l'on ne recherche pas un ajustement exact aux données. L'objectif est d'extraire la structure générale du signal et d'éliminer le bruit. On va en particulier chercher à déceler la présence d'une tendance, d'une saisonnalité dans les données.

Si l'on reprend la série temporelle du nombre de passagers aériens (voir Exemple 1.4), on peut estimer une saisonnalité et en déduire la série corrigée des variations saisonnières (voir Figure 1.5). On reviendra plus tard sur cette notion.

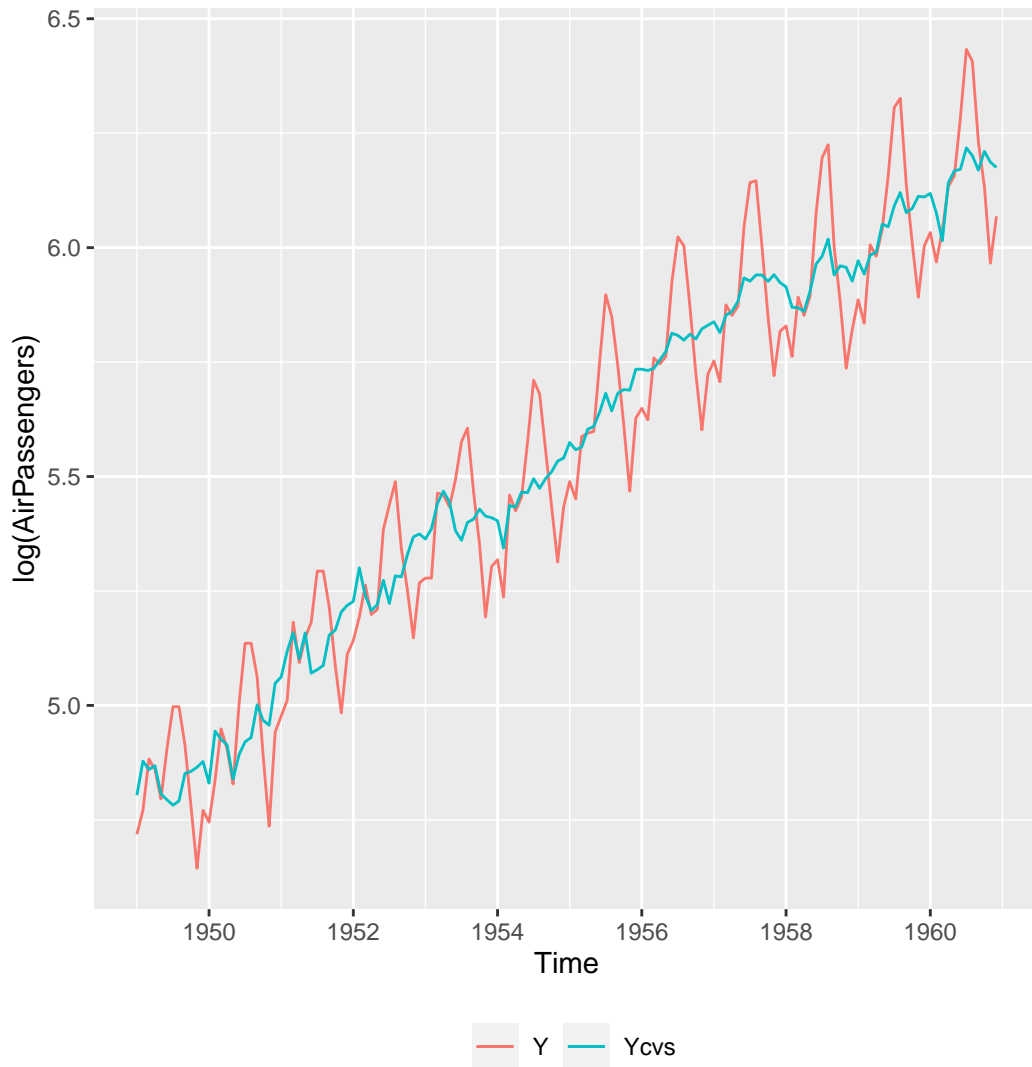


Figure 1.5: Série corrigée des variations saisonnières de la série temporelle AirPassengers (log-transformée).

Après l'analyse et la modélisation d'une série temporelle, on est souvent intéressé par la **prévision** des futures valeurs de la série. Par exemple, on peut chercher à prédire la température

dans les jours à venir, la concentration en ozone, ... Plus formellement, à partir des n premiers instants Y_1, \dots, Y_n de la série, on souhaite prévoir la valeur suivante Y_{n+1} ou plus éloignée dans le temps Y_{n+h} . Un exemple de prévision sur l'année suivante de la série temporelle AirPassengers est données en Figure 1.6. On devra s'intéresser alors à l'erreur de prédiction, à la taille des intervalles de prédiction, ...

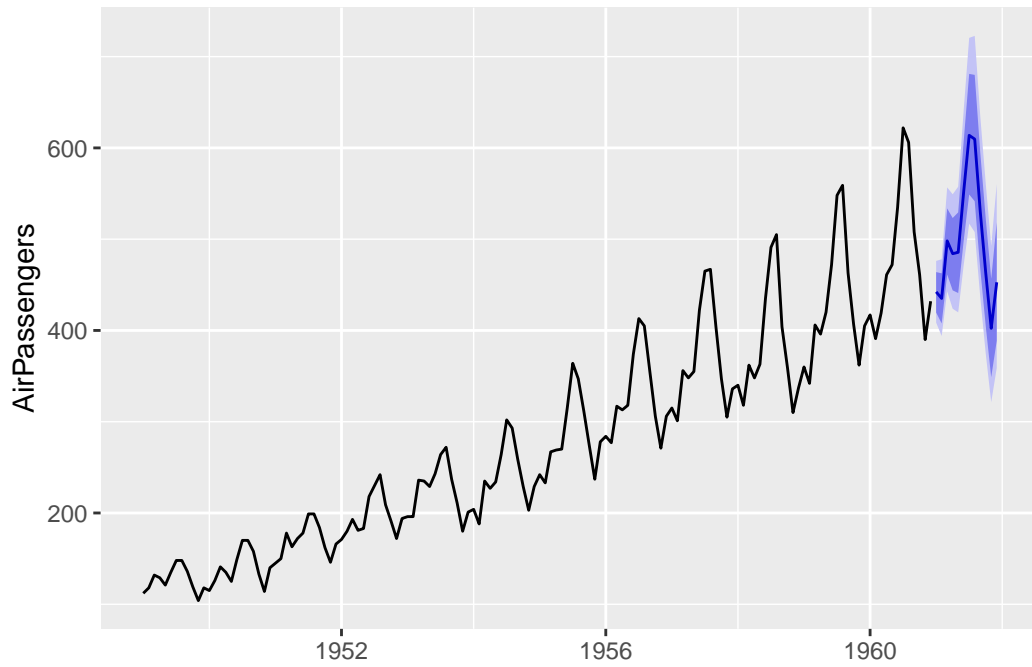


Figure 1.6: Prévision sur l'année suivante pour la série AirPassengers avec intervalle de prédiction.

2 Tendances et saisonnalités

2.1 Décomposition d'une série temporelle

Comme on a pu le constater dans les exemples de l' [introduction](#), quitte à faire une transformation des données au préalable, on peut décomposer une série temporelle en un **modèle additif** composé de trois termes de la façon suivante.

Definition 2.1 (Décomposition en modèle additif). La série temporelle $(Y_t)_{t \in T}$ se décompose en

$$Y_t = m_t + s_t + X_t, \quad \forall t \in T \quad (2.1)$$

où

- m_t est la **tendance** : une **fonction déterministe** à variation lente qui capte les variations de niveau et que l'on espère assez lisse
- s_t est la **saisonnalité** : une **fonction déterministe périodique** de période r ($s_{t+r} = s_t, \forall t$) telle que

$$\sum_{h=1}^r s_{t+h} = 0, \quad \forall t \in T$$

- X_t est un **bruit aléatoire stationnaire** appelé parfois résidu. Ce terme sera à définir dans la suite.

Remarque

L'hypothèse de somme nulle de la saisonnalité sur la période r n'est pas une hypothèse contraignante. En effet, on peut s'y ramener facilement en modifiant la tendance : si $\sum_{h=1}^r s_{t+h} = a$ alors on définit $\tilde{s}_t = s_t - \frac{a}{r}$ et $\tilde{m}_t = m_t + \frac{a}{r}$.

Si la saisonnalité et les variations semblent croître, on peut parfois atténuer ce phénomène en tentant une transformation des données. C'est en particulier ce que l'on peut constater sur les données *AirPassengers* quand on prend leur logarithme (voir Figure 1.4). Donc quitte à faire une transformation des données, on peut supposer le modèle additif (Equation 2.1).

Objectif du cours

- Apprendre à modéliser et estimer les composantes tendance $(m_t)_{t \in T}$ et saisonnalité $(s_t)_{t \in T}$
- Apprendre à modéliser le bruit résiduel $(X_t)_{t \in T}$
- Faire des prévisions sur les valeurs futures de la série temporelle initiale $(Y_t)_{t \in T}$.

La méthode générale pour étudier une série temporelle est la suivante :

Méthode générale d'étude

- **Etape 1** : on trace la série des données observées (Y_1, \dots, Y_n) et on essaie de déceler ses principales caractéristiques : une tendance, une composante saisonnière, une ou des ruptures dans le comportement de la série, une ou des observations aberrantes.
- **Etape 2** : On estime / supprime la tendance $(m_t)_{t \in T}$ et la composante saisonnière $(s_t)_{t \in T}$ pour obtenir une série $(X_t)_{t \in T}$ de résidus stationnaires. Pour cela, on peut utiliser plusieurs techniques: transformer les données, estimer les tendances et composantes saisonnières puis les supprimer des données, différencier la série.
- **Etape 3** : Choisir un modèle de processus stationnaire pour la série des résidus
- **Etape 4** : Prévoir les valeurs futures de la série en prévoyant d'abord celles des résidus puis remonter jusqu'à la série initiale en utilisant les transformations inverses.

2.2 Estimation / élimination d'une tendance en l'absence de saisonnalité

Dans cette partie, on suppose que la série $(Y_t)_{t \in T}$ n'a pas de saisonnalité. Elle suit donc le modèle additif suivant

$$Y_t = m_t + X_t, \quad \forall t \in T.$$

Sans perte de généralité, on suppose que $(X_t)_{t \in T}$ est un processus centré ($\mathbb{E}[X_t] = 0, \forall t \in T$). En effet, si $\mathbb{E}[X_t] \neq 0$, on remplace m_t et X_t par $m_t + \mathbb{E}[X_t]$ et $X_t - \mathbb{E}[X_t]$ respectivement.

On suppose également que l'on observe le processus sur les instants de temps $t = 1, \dots, n$: (Y_1, \dots, Y_n) .

Sans exhaustivité, la fonction tendance peut prendre l'une de ces formes

- Tendance linéaire : $m_t = \alpha_0 + \alpha_1 t$
- Tendance quadratique : $m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$
- Tendance polynomiale : $m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_k t^k$
- Tendance exponentielle : $m_t = c_0 + c_1 \alpha^t$
- Tendance de Gompertz : $m_t = \exp(c_0 + c_1 \alpha^t)$
- Tendance Logistique : $m_t = 1/(c_0 + c_1 \alpha^t)$
- ou bien des mélanges de ces types de fonctions.

2.2.1 Estimation de la tendance par moindres carrés

Dans cette section, on suppose que la tendance est une combinaison linéaire de fonctions temporelles, **connues et déterministes** :

$$m_t = \sum_{j=1}^p \alpha_j m_t^{(j)}.$$

Pour déterminer m_t , on cherche donc à estimer les coefficients inconnus α_j , pour $j = 1, \dots, p$. Pour cela, nous pouvons utiliser l'estimation par moindres carrés.

$$\begin{aligned} (\hat{\alpha}_1, \dots, \hat{\alpha}_p) &= \underset{(\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{t=1}^n (Y_t - m_t)^2 \\ &= \underset{(\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{t=1}^n \left(Y_t - \alpha_1 m_t^{(1)} - \dots - \alpha_p m_t^{(p)} \right)^2. \end{aligned}$$

On peut constater que l'on se ramène à un problème de régression linéaire de la forme

$$\underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbb{Y}} = \underbrace{\begin{pmatrix} m_1^{(1)} & \dots & m_1^{(p)} \\ \vdots & \ddots & \vdots \\ m_n^{(1)} & \dots & m_n^{(p)} \end{pmatrix}}_{\mathbb{X}} \underbrace{\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

D'après les résultats de la régression linéaire, si $\mathbb{X}'\mathbb{X}$ est inversible,

$$\hat{\boldsymbol{\theta}} = (\mathbb{X}'\mathbb{X})^{-1} \mathbb{X}'\mathbb{Y}.$$

On obtient alors les **données corrigées de la tendance** via l'expression suivante

$$\hat{Y}_t^{\text{CT}} = Y_t - \hat{m}_t = Y_t - \sum_{j=1}^p \hat{\alpha}_j m_t^{(j)}$$

Exemple 2.1. On considère la série temporelle $(Y_t)_{t \in \mathbb{N}}$ définie par la relation

$$Y_t = (1 + 0.01 t^2) + X_t \text{ avec } X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 4^2).$$

On observe les $n = 100$ premières valeurs de cette série (voir Figure 2.1, courbe noire).

On considère les trois fonctions tendances suivantes

- $m_{1,t} = \alpha_0 + \alpha_1 t$
- $m_{2,t} = \alpha_0 + \alpha_1 t + \alpha_2 t^2$
- $m_{3,t} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$

Les coefficients sont estimés par moindres carrés pour les trois tendances :

- pour une tendance linéaire :

```
Call:
lm(formula = y ~ x1, data = dataaux)

Coefficients:
(Intercept)          x1
   -18.547         1.045
```

- pour une tendance quadratique :

```
Call:
lm(formula = y ~ x1 + x2, data = dataaux)

Coefficients:
(Intercept)          x1          x2
   -0.07851    -0.04172     0.01076
```

- pour une tendance cubique :

```
Call:
lm(formula = y ~ ., data = dataaux)

Coefficients:
(Intercept)          x1          x2          x3
   3.993e-01   -9.711e-02   1.212e-02  -9.005e-06
```

Les estimations associées à $\hat{m}_{1,t}$, $\hat{m}_{2,t}$ et $\hat{m}_{3,t}$ sont représentées sur la Figure 2.1 en rouge, bleu et rose respectivement. Les courbes bleue et rose se superposent car la tendance recherchée $m_t = 1 + 0.01t^2$ est quadratique.

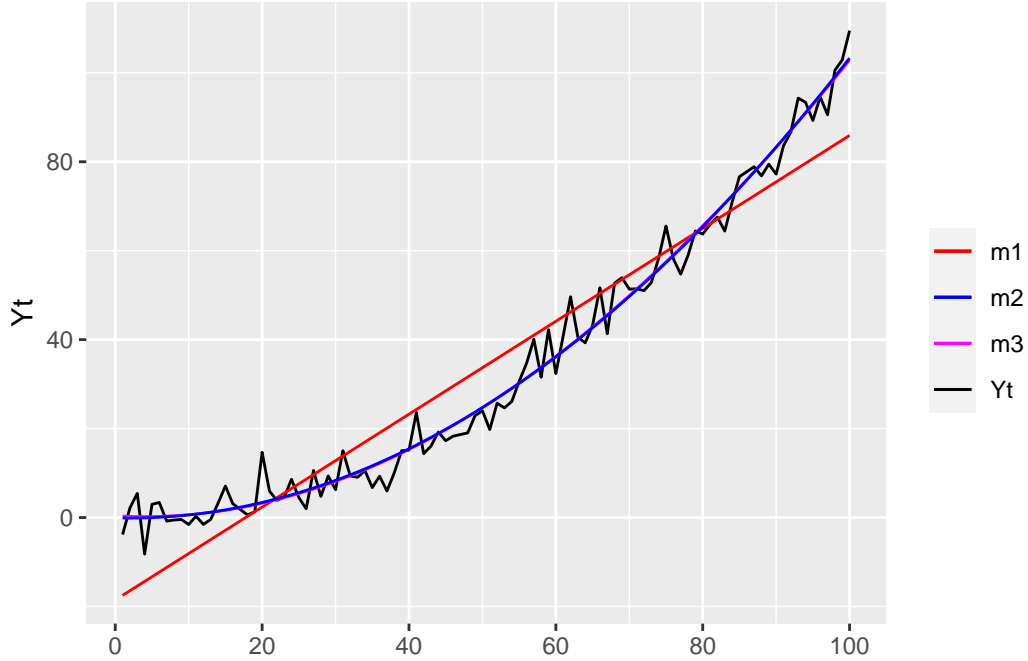


Figure 2.1: Exemple pour l'estimation de la tendance par moindres carrés

2.2.2 Estimation de la tendance par filtrage de moyenne mobile

2.2.3 Opérateurs retard et avance

Pour pouvoir parler de moyenne mobile, nous devons commencer par définir deux opérateurs importants pour ce cours de séries temporelles: les opérateurs **retard** B et **avance** F .

Definition 2.2. L'opérateur **retard** B sur une série temporelle $(Y_t)_{t \in T}$ est défini par :

$$B Y_t = Y_{t-1}, \quad \forall t \in T.$$

On note de manière naturelle : $B^h Y_t = Y_{t-h}$, $\forall t \in T$ et $\forall h \in \mathbb{N}^*$.

L'opérateur **avance** F sur une série temporelle $(Y_t)_{t \in T}$ est défini par :

$$F Y_t = Y_{t+1}, \quad \forall t \in T.$$

On note aussi $F^h Y_t = Y_{t+h}$, $\forall t \in T$ et $\forall h \in \mathbb{N}^*$ et $B^{-h} = F^h$, pour tout h .

A partir de ces deux opérateurs, on peut définir la notion de moyenne mobile.

Definition 2.3. Une **moyenne mobile** est un opérateur linéaire de la forme

$$M = \sum_{h=-m_1}^{m_2} \theta_h B^{-h},$$

où $(m_1, m_2) \in \mathbb{N} \times \mathbb{N}$ et $\theta_h \in \mathbb{R}$ pour tout h .

L'**ordre** de la moyenne mobile est l'entier $m_1 + m_2 + 1$.

La moyenne mobile est dite

- **normalisée** si $\sum_{h=-m_1}^{m_2} \theta_h = 1$. Il s'agit alors d'une moyenne au sens où on l'entend habituellement.
- **centrée** si $m_1 = m_2$. On prend autant d'instantanés du passé que du futur.
- **symétrique** si $m_1 = m_2 = m$ et $\theta_h = \theta_{-h}$, pour $h = 1, \dots, m$. On donne des poids identiques aux instantanés passés et futurs de même ordre.

Ainsi la moyenne mobile M appliquée à la série temporelle $(Y_t)_{t \in T}$ donne

$$\begin{aligned} M Y_t &= \sum_{h=-m_1}^{m_2} \theta_h B^{-h} Y_t \\ &= \theta_{-m_1} Y_{t-m_1} + \dots + \theta_{-1} Y_{t-1} \\ &\quad + \theta_0 Y_t \\ &\quad + \theta_1 Y_{t+1} + \dots + \theta_{m_2} Y_{t+m_2} \end{aligned}$$

Appliquer l'opérateur M revient donc à faire une “moyenne locale pondérée” des termes $Y_{t-m_1}, \dots, Y_0, \dots, Y_{t+m_2}$.

Example 2.2 (Exemple important de la moyenne mobile M_{2q+1}). Soit $q \in \mathbb{N}^*$. On considère la moyenne mobile M_{2q+1} définie par

$$M_{2q+1} Y_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}.$$

M_{2q+1} est une moyenne mobile avec $m_1 = m_2 = q$ et $\theta_h = \frac{1}{1+2q} \mathbb{1}_{|h| \leq q}$. C'est donc une moyenne mobile finie, symétrique et normalisée d'ordre $2q+1$. Au vu de la définition des coefficients, cette moyenne mobile peut être vue comme un filtre “passe-bas”.

La moyenne mobile M_{2q+1} laisse invariante les tendances linéaires : si $m_t = a + bt$ alors

$$M_{2q+1}m_t = \frac{1}{2q+1} \sum_{j=-q}^q m_{t-j} = \frac{1}{2q+1} \sum_{j=-q}^q (a + bt - bj) = a + bt = m_t.$$

Soit une série temporelle $(Y_t)_{t \in T}$ de la forme $Y_t = m_t + X_t$, où m_t est la tendance et X_t un processus centré. Si la tendance est pratiquement linéaire et que la moyenne empirique des X_t est proche de 0 (ce qui est en tout cas vrai pour q grand), on a

$$M_{2q+1}Y_t = \frac{1}{2q+1} \sum_{j=-q}^q m_{t-j} + \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j} \approx m_t.$$

\Rightarrow la moyenne mobile nous donne une estimation de la tendance quand elle est pratiquement linéaire.

Pour illustrer ce point, on observe les 50 premières réalisations d'une série temporelle $(Y_t)_{t \in \mathbb{N}}$ définie par $Y_t = (5 + 0.1t) + X_t$ avec $X_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

La Figure 2.2 montre l'estimation de la tendance par la moyenne mobile M_{2q+1}

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}, \quad \forall q+1 \leq t \leq n-q$$

pour $2q+1 = 3$ et $2q+1 = 7$.

2.2.4 Estimation de la tendance par lissage exponentiel

Definition 2.4. Le **lissage exponentiel simple** consiste à estimer la tendance via la formule récursive suivante : pour $\alpha \in [0, 1]$,

$$\begin{cases} \hat{m}_t = \alpha Y_t + (1 - \alpha) \hat{m}_{t-1} & \text{pour } t = 2, \dots, n \\ \hat{m}_1 = Y_1 \end{cases}$$

Proposition 2.1. *Le lissage exponentiel simple est une moyenne mobile normalisée.*

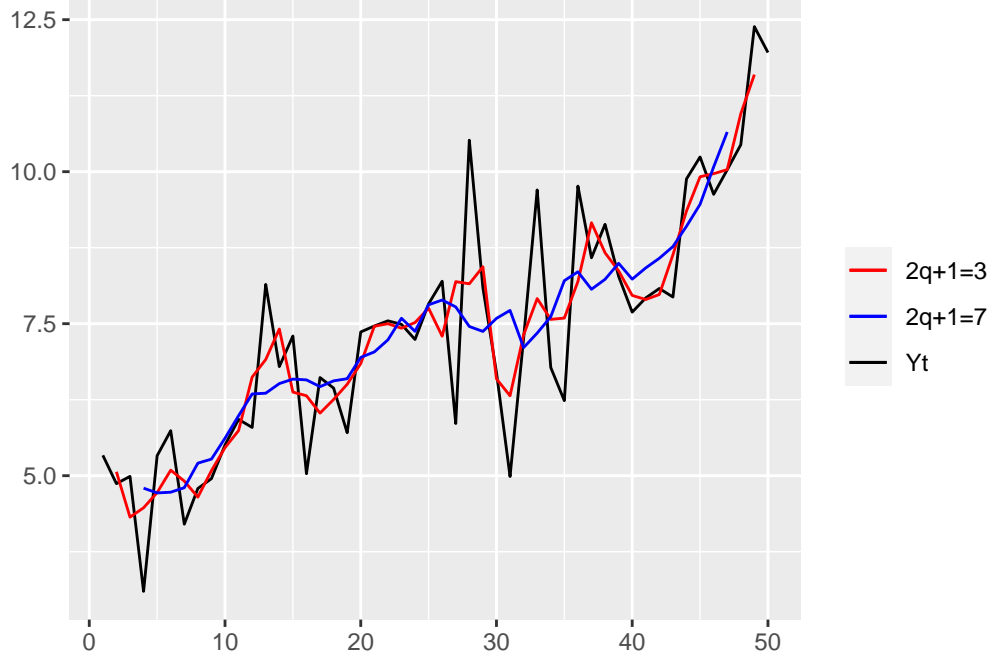


Figure 2.2: Exemples d'estimation de la tendance d'une série temporelle à l'aide d'une moyenne mobile M_{2q+1} .

Preuve

En résolvant l'équation de récurrence on a :

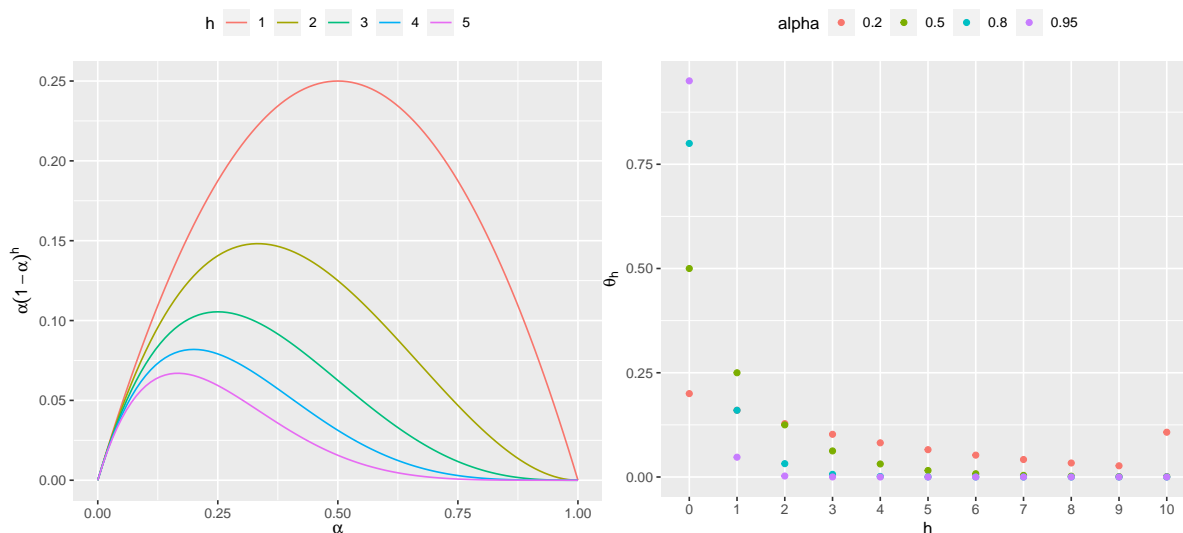
$$\begin{aligned}
 \hat{m}_t &= \alpha Y_t + (1 - \alpha) \hat{m}_{t-1} = \alpha Y_t + \alpha(1 - \alpha) Y_{t-1} + (1 - \alpha)^2 \hat{m}_{t-2} \\
 &= \alpha Y_t + \alpha(1 - \alpha) Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + (1 - \alpha)^3 \hat{m}_{t-3} \\
 &= \sum_{k=0}^{t-2} \alpha(1 - \alpha)^k Y_{t-k} + (1 - \alpha)^{t-1} Y_1 \\
 &= \sum_{h=0}^{t-1} \theta_h Y_{t-h}
 \end{aligned}$$

avec les coefficients

$$\theta_h = \begin{cases} \alpha(1 - \alpha)^h & \text{si } 0 \leq h \leq t-2 \\ (1 - \alpha)^{t-1} & \text{si } h = t-1 \\ 0 & \text{sinon.} \end{cases}$$

$$\text{Donc } \sum_{h=0}^{t-1} \theta_h = \sum_{h=0}^{t-2} \alpha(1 - \alpha)^h + (1 - \alpha)^{t-1} = \alpha \frac{1 - (1 - \alpha)^{t-1}}{1 - (1 - \alpha)} + (1 - \alpha)^{t-1} = 1.$$

On parle de **lissage exponentiel** car c'est une moyenne pondérée des valeurs précédentes avec une décroissance exponentielle des poids (voir Figure 2.3b). Ainsi les observations les plus récentes ont le plus de poids. Le choix de α est fondamental (voir Figure 2.3a): Plus α est proche de 1, plus on donne de poids à la dernière observation. L'estimation est alors moins lisse et on tend vers du sur-ajustement. Plus α est proche de 0, plus le lissage exponentiel s'appuie sur une mémoire longue de la série temporelle.



(a) Les poids $\alpha(1 - \alpha)^h$ en fonction de α pour plusieurs valeurs de h (b) Les poids θ_h en fonction de h pour plusieurs valeurs de α

Figure 2.3: Comportement des poids du lissage exponentiel

Le lissage exponentiel simple est parfois utilisé pour la prévision. Si l'on observe la série temporelle sur les instants $\{1, \dots, n\}$ alors une prévision à l'horizon h de la série est donnée par :

$$\hat{Y}_{n,h} = \hat{Y}_{n+h} = \hat{Y}_{n+1} = \sum_{k=0}^{n-2} \alpha(1-\alpha)^k Y_{n-k} + (1-\alpha)^{n-1} Y_1 = \hat{m}_n.$$

La formule de mise jour du lissage exponentiel permet de voir qu'une observation supplémentaire de la série ne nécessite pas de recalculer entièrement la prévision. En effet, si on observe en plus la valeur au temps $n+1$ alors

$$\hat{Y}_{n+1,1} = \hat{m}_{n+1} = \alpha Y_{n+1} + (1-\alpha) \hat{m}_n.$$

Comme évoqué précédemment, le choix du paramètre α est important. On peut chercher à minimiser l'erreur de prévision :

$$\hat{\alpha} = \underset{\alpha_1, \dots, \alpha_p}{\operatorname{argmin}} \sum_{t=1}^{n-h} \left(Y_{t+h} - \hat{Y}_{t+h}^{(\alpha_i)} \right)^2$$

Exemple 2.3. On applique le lissage exponentiel simple pour deux valeurs de α sur les trois séries temporelles suivantes, observées pour $t \in \{1, \dots, 50\}$:

- $Y_t^{[1]} = 1 + X_t$ avec $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.01)$
- $Y_t^{[2]} = 1 + 0.05t + X_t$ avec $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.1)$
- $Y_t^{[3]} = \mathbb{1}_{1 \leq t \leq 25} + 2\mathbb{1}_{t > 25} + X_t$ avec $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.1)$

Dans le cadre du lissage exponentiel simple, on vient de voir que la prévision est constante. Une extension est le lissage exponentiel double où l'on souhaite une prédiction linéaire. Cette prédiction est donnée par

$$\hat{Y}_{n+h} = \hat{a}_n h + \hat{b}_n$$

avec

$$\begin{cases} \hat{a}_n = \hat{a}_{n-1} + (1 - \alpha)^2 (Y_n - \hat{Y}_{n-1,1}) \\ \hat{b}_n = \hat{b}_{n-1} + \hat{a}_{n-1} + (1 - \alpha^2) (Y_n - \hat{Y}_{n-1,1}) \\ \hat{a}_2 = Y_2 - Y_1, \hat{b}_2 = Y_1 \end{cases}$$

Dans la même famille de méthode, on peut citer la méthode de Holt-Winters qui est un lissage exponentiel double avec une formule de mise à jour différente (voir Figure 2.5).

Exemple 2.4. On reprend les trois séries de l'exemple Exemple 2.3 et on applique un lissage exponentiel double avec la méthode de Holt-Winters.

2.2.5 Elimination de la tendance par différenciation

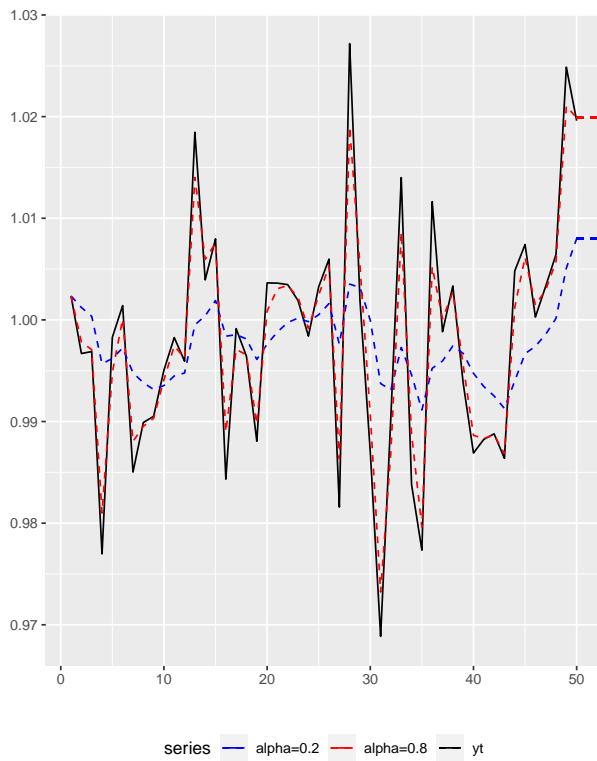
Définition 2.5. Soit $(Y_t)_{t \in T}$ une série temporelle.

L'opérateur **différenciation à l'ordre 1** ∇ d'une série temporelle est défini par

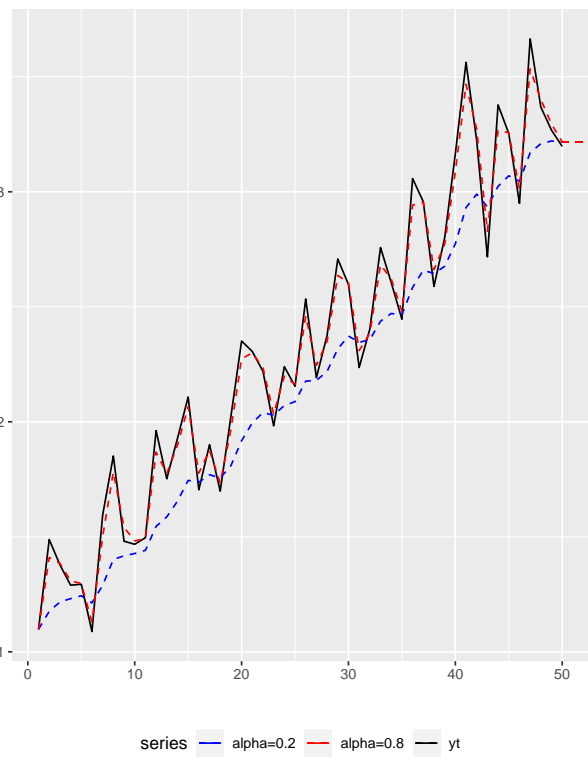
$$\nabla Y_t = Y_t - Y_{t-1} = Y_t - B Y_t = (I - B) Y_t, \quad \forall t \in T.$$

La **différenciation à l'ordre k** de la série $(Y_t)_{t \in T}$ est définie par

$$\nabla^k Y_t = (I - B)^k Y_t, \quad \forall t \in T.$$



(a) La série temporelle $Y_t^{[1]}$

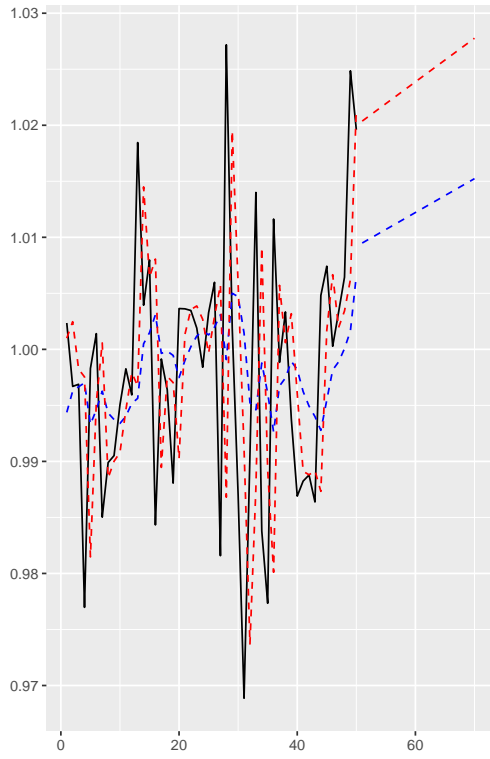


(b) La série temporelle $Y_t^{[2]}$

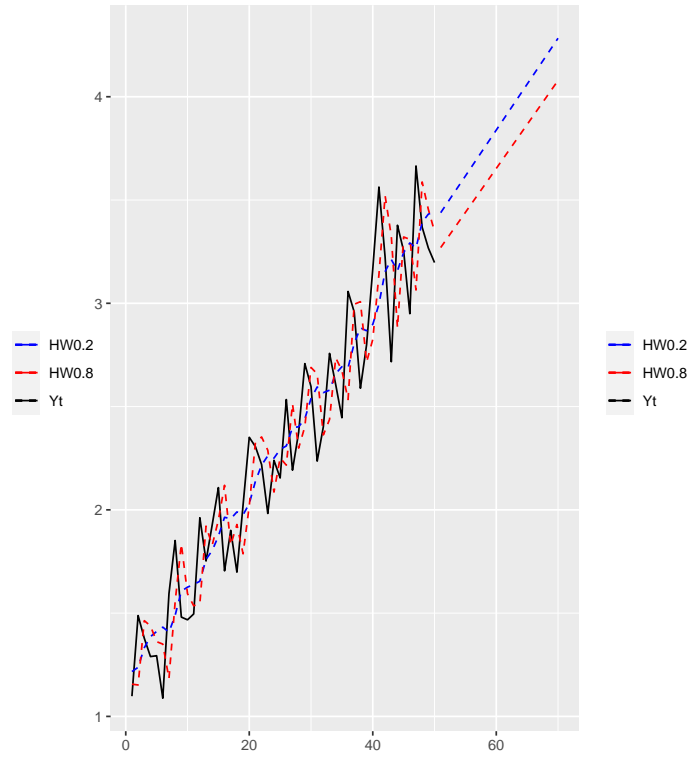


(c) La série temporelle $Y_t^{[3]}$

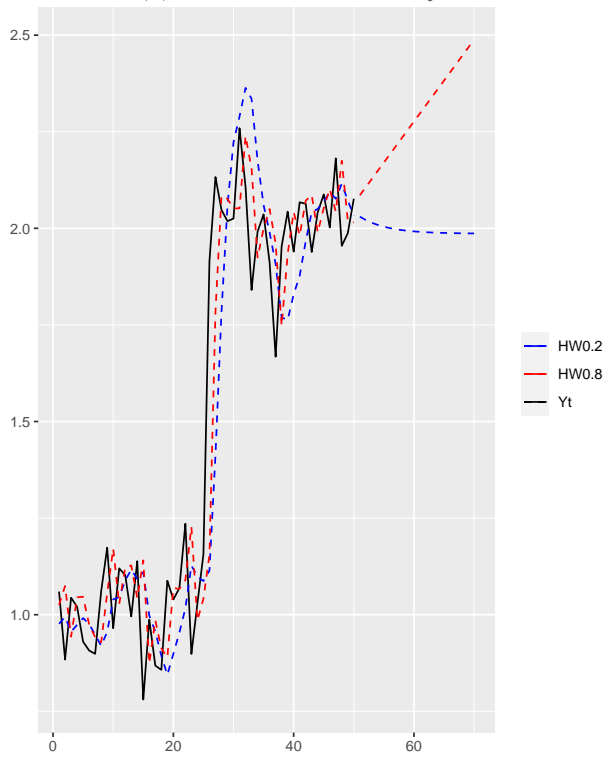
Figure 2.4: Exemples de lissage exponentiel simple



(a) La série temporelle $Y_t^{[1]}$



(b) La série temporelle $Y_t^{[2]}$



(c) La série temporelle $Y_t^{[3]}$

Figure 2.5: Reprise des exemples précédents avec ici la méthode de Holt-Winters.

⚠ Attention à la manipulation de cet opérateur.

- $\nabla^2 Y_t = (I - B)(I - B)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$
- $\nabla^3 Y_t = (I - B)\nabla^2 Y_t = Y_t - 3Y_{t-1} + 3Y_{t-2} - Y_{t-3}$

Proposition 2.2. Une tendance polynomiale m_t de degré k est réduite à une constante par une différenciation à l'ordre k .

Preuve

On raisonne par récurrence.

- Pour $k = 1$: pour une tendance linéaire $m_t = a + bt$, on a $\nabla m_t = a + bt - a - b(t-1) = b$.
- On suppose la proposition vraie au rang k
- Au rang $k + 1$: soit la tendance $m_t = \sum_{j=0}^{k+1} a_j t^j = a_{k+1} t^{k+1} + P_k(t)$. Alors

$$\begin{aligned}
 \nabla^{k+1} m_t &= \nabla^k (I - B)(a_{k+1} t^{k+1} + P_k(t)) \\
 &= \nabla^k [a_{k+1} (t^{k+1} - (t-1)^{k+1}) + P_k(t) - P_k(t-1)] \\
 &= \nabla^k [a_{k+1} (t - t + 1) Q_k(t) + P_k(t) - P_k(t-1)] \\
 &= \text{constante}
 \end{aligned}$$

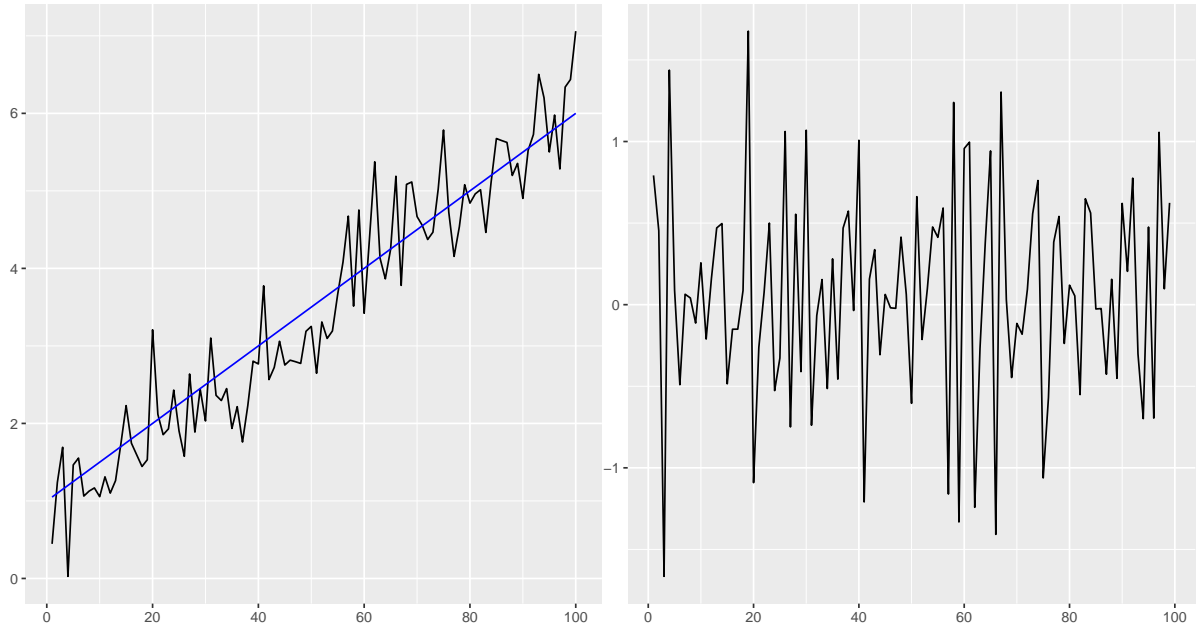
La différenciation permet d'éliminer les tendances polynomiales et donc pratiquement toutes les tendances car elles peuvent très souvent être approchées par des polynômes. Attention, il faut bien noter que cette technique permet d'éliminer la tendance mais ne l'estime pas.

Exemple 2.5. En Figure 2.6, on a simulé une série temporelle définie par $Y_t = 1 + \frac{1}{20}t + X_t$ avec $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.25)$. On voit que la différenciation de la série ∇Y_t élimine la tendance linéaire.

Exemple 2.6. Dans cet exemple, on a simulé une série temporelle définie par $Y_t = 1 - 5t + 0.25t^2 + X_t$ avec $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 100)$.

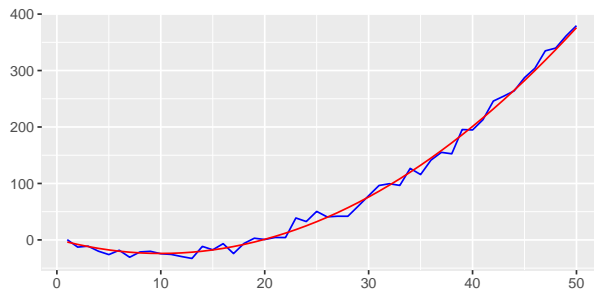
2.3 Estimation / élimination de la tendance et de la saisonnalité

On considère maintenant le cas général d'une série temporelle avec une tendance et une saisonnalité présentes dans la décomposition :

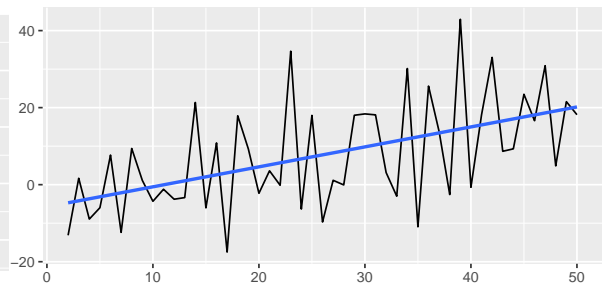


(a) Représentation de la série $(Y_t)_{t \in T}$ et la tendance linéaire $t \mapsto 1 + \frac{1}{20}t$ en bleu
 (b) Représentation de la différenciation de la série ∇Y_t

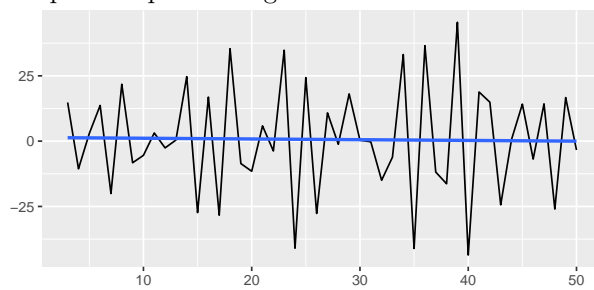
Figure 2.6: Illustration de l'élimination d'une tendance linéaire par la différenciation à l'ordre 1.



(a) Représentation de la série $(Y_t)_t$ et de la tendance quadratique en rouge



(b) Différenciation à l'ordre 1



(c) Différenciation à l'ordre 2

Figure 2.7: Illustration de l'élimination d'une tendance quadratique par une différenciation d'ordre 2 mais pas d'ordre 1

$$Y_t = m_t + s_t + X_t, \forall t \in T$$

avec $\mathbb{E}[X_t] = 0$, $s_{t+r} = s_t$ et $\sum_{k=1}^r s_{t+k} = 0$, $\forall t \in T$.

2.3.1 Estimation par moindres carrés

On peut reprendre la méthode des moindres carrés en supposant cette fois-ci que la tendance mais aussi la composante saisonnière sont des combinaisons linéaires de fonctions connues. On suppose donc qu'il existe des fonctions $m_t^{(j)}$ pour $j = 1, \dots, p$, et $s_t^{(\ell)}$ pour $\ell = 1, \dots, q$ telles que :

$$Y_t = \sum_{j=1}^p \alpha_j m_t^{(j)} + \sum_{\ell=1}^q \beta_\ell s_t^{(\ell)} + X_t, \forall t \in T.$$

Pour les fonctions de saisonnalité $s_t^{(\ell)}$, on peut par exemple considérer

- des indicatrices : par exemple dans le cas d'une saisonnalité trimestrielle

$$\forall \ell \in \{1, \dots, 4\}, s_t^{(\ell)} = \begin{cases} 1 & \text{si le trimestre à l'instant } t \text{ est } \ell \\ 0 & \text{sinon} \end{cases}.$$

- une combinaison de fonctions sinusoïdales.

On peut noter que les saisonnalités $s_t^{(\ell)}$ peuvent avoir des périodes différentes.

On cherche donc à estimer le vecteur des coefficients inconnus $\theta = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$ par la méthode des moindres carrés. On obtient alors les estimateurs $\hat{\theta} = (\hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{\beta}_1, \dots, \hat{\beta}_q)'$ et on récupère les données ajustées

$$\hat{Y}_t = \hat{m}_t + \hat{s}_t = \sum_{j=1}^p \hat{\alpha}_j m_t^{(j)} + \sum_{\ell=1}^q \hat{\beta}_\ell s_t^{(\ell)}, \forall t \in T.$$

La **série corrigée des variations saisonnières** (CVS) est alors définie par

$$\hat{Y}_t^{\text{CVS}} = Y_t - \hat{s}_t.$$

Exemple 2.7. Dans cet exemple, on s'intéresse à la série temporelle $Y_t = 0.5t + 3\cos\left(\frac{\pi t}{6}\right) + X_t$, $X_t \underset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$. On considère les fonctions de tendance polynomiales $m_t^{(j)} = t^j$ pour $j \in \{1, 2, 3\}$ et les fonctions de saisonnalité définies par $s_t^{(\ell)} = \cos(\theta_\ell t)$ avec $(\theta_1, \theta_2, \theta_3) = (\frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{3})$. A l'aide de la fonction `lm()`, on ajuste un modèle de régression linéaire

```

Call:
lm(formula = Yt ~ ., data = dfaux)

Residuals:
    Min       1Q   Median       3Q      Max
-2.47779 -0.63375 -0.02225  0.67613  2.20892

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.541e-02  7.165e-01   0.063  0.949761
m1           4.701e-01  1.207e-01   3.895  0.000338 ***
m2           2.578e-03  5.484e-03   0.470  0.640591
m3          -4.399e-05  7.094e-05  -0.620  0.538477
s1           3.117e+00  2.385e-01  13.068 < 2e-16 ***
s2          -4.606e-01  2.379e-01  -1.936  0.059497 .
s3          -3.389e-01  2.364e-01  -1.434  0.158912
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.168 on 43 degrees of freedom
Multiple R-squared:  0.9808,    Adjusted R-squared:  0.9781
F-statistic: 365.5 on 6 and 43 DF,  p-value: < 2.2e-16

```

On retrouve des estimations des coefficients en cohérence avec les données simulées. La Figure 2.8 représente les $n = 50$ premiers temps simulés de la série Y_t , les données ajustées \hat{Y}_t et les données corrigées des variations saisonnières \hat{Y}_t^{CVS} .

2.3.2 Estimation par Moyenne Mobile

On suppose toujours que l'on a la décomposition

$$Y_t = m_t + s_t + X_t, \quad \forall t \in T$$

L'idée générale est de trouver une moyenne mobile M qui vérifie les propriétés suivantes :

- la moyenne mobile laisse invariante la tendance : $M m_t = m_t$
- la moyenne mobile absorbe la saisonnalité : $M s_t = 0$
- la moyenne mobile réduit la variance du processus observé : $M Y_t$ a une variance plus faible que Y_t .

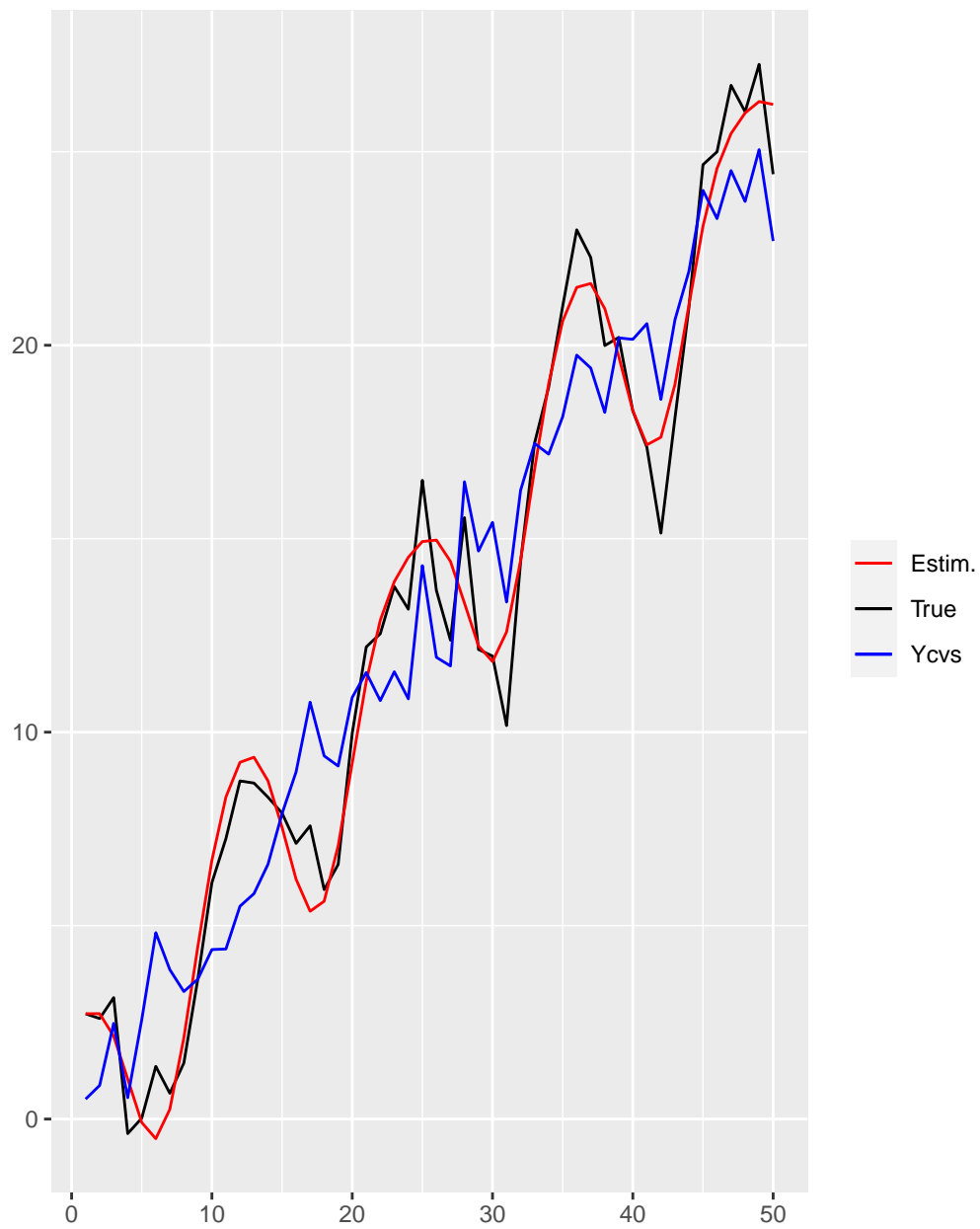


Figure 2.8: Illustration de l'estimation de la tendance et saisonnalité par moindres carrés.

Si on trouve une telle moyenne mobile M , MY_t estime la tendance m_t et on travaille ensuite sur $\hat{Y}_t^{\text{CT}} = Y_t - MY_t$ pour estimer la saisonnalité.

2.3.2.1 Cas d'une périodicité impaire

On suppose que la série temporelle admet une saisonnalité de période impaire $r = 2q + 1$ et on observe cette série temporelle aux instants $t = 1, \dots, n$.

On reprend la moyenne mobile M_{2q+1} (voir Exemple 2.2) symétrique, normalisée et d'ordre $2q + 1$:

$$M_{2q+1} = \frac{1}{2q+1} \sum_{h=-q}^q B^h.$$

Cette moyenne mobile M_{2q+1} appliquée sur la série $(Y_t)_{t \in T}$ donne directement une estimation de la tendance et annule la saisonnalité. En effet,

$$M_{2q+1}Y_t = \frac{1}{2q+1} \sum_{h=-q}^q m_{t-h} + \frac{1}{2q+1} \sum_{h=-q}^q s_{t-h} + \frac{1}{2q+1} \sum_{h=-q}^q X_{t-h}$$

où

- le premier terme donne une estimation de la tendance si celle-ci est assez lisse
- le second terme est nul par l'hypothèse $\sum_{j=1}^r s_{t+j} = 0, \forall t \in T$
- le dernier terme est quasiment nul si q est assez grand (convergence vers $\mathbb{E}[X_t] = 0$)

Une fois que l'on a estimé la tendance $\hat{m}_t = M_{2q+1}Y_t = \frac{1}{2q+1} \sum_{h=-q}^q Y_{t-h}$, on considère la série corrigée de la tendance $\hat{Y}_t^{\text{CT}} = Y_t - \hat{m}_t$, sur laquelle on va estimer la saisonnalité.

Pour $k = 1, \dots, r$, on calcule la moyenne ω_k des valeurs de la série corrigée de la tendance sur tous les points à une même distance r :

$$\omega_k = \frac{1}{\text{Card}\{j : q < k + jr \leq n - q\}} \sum_{\{j : q < k + jr \leq n - q\}} \hat{Y}_{k+jr}^{\text{CT}}$$

Mais comme les termes ω_k ne sont pas nécessairement de somme nulle, on définit comme estimateur de la composante saisonnière

$$\hat{s}_k = \begin{cases} \omega_k - \frac{1}{r} \sum_{i=1}^r \omega_i & \text{pour } k = 1, \dots, r \\ \hat{s}_{k-r} & \text{pour } k > r. \end{cases}$$

Ainsi on assure que $(\hat{s}_t)_{t \in T}$ est bien r -périodique.

2.3.2.2 Cas d'une périodicité paire

Si la périodicité est paire $r = 2q$, on peut utiliser une approche similaire en considérant la moyenne mobile symétrique normalisée :

$$M_{2q}Y_t = \frac{1}{2q} \left(Y_{t-q+\frac{1}{2}} + \dots + Y_{t-\frac{1}{2}} + Y_{t+\frac{1}{2}} + \dots + Y_{t+q-\frac{1}{2}} \right),$$

où $Y_{t-\frac{1}{2}} = \frac{1}{2} (Y_{t-1} + Y_t)$ pour estimer la tendance et éliminer la saisonnalité.

Pour se convaincre que la saisonnalité est éliminée, on peut réécrire $M_{2q}Y_t$:

$$\begin{aligned} M_{2q}Y_t &= \frac{1}{2q} \sum_{h=-(q-1)}^q Y_{(t+h)-\frac{1}{2}} \\ &= \frac{1}{2q} \sum_{h=-(q-1)}^q \frac{1}{2} (Y_{t+h-1} + Y_{t+h}) \\ &= \frac{1}{4q} \left\{ \sum_{h=-q}^{q-1} Y_{t+h} + \sum_{h=-q+1}^q Y_{t+h} \right\} \\ &= \frac{1}{2r} \left\{ Y_{t-q} + 2 \sum_{h=-q+1}^{q-1} Y_{t+h} + Y_{t+q} \right\} \end{aligned}$$

Ainsi $M_{2q}s_t = \frac{1}{2r} \left\{ \sum_{h=-q}^{q-1} s_{t+h} + \sum_{h=-q+1}^q s_{t+h} \right\} = 0$ car sommes de r termes consécutifs de s_t .

On poursuit ensuite le même raisonnement que dans le cas d'une période impaire.

En général, on ré-estime la tendance en considérant les données corrigées des variations saisonnières et en appliquant une des techniques vues précédemment. On obtient alors \tilde{m}_t . La série des bruits est alors

$$X_t = Y_t - \tilde{m}_t - \hat{s}_t, \forall t \in T$$

Exemple 2.8. On s'intéresse à la consommation trimestrielle en électricité d'une entreprise de 1997 à 1999. Les valeurs observées sont données dans la Table 2.1 et représentées en Figure 2.9. On observe une périodicité par trimestre $r = 2q = 4$.

Table 2.1: Consommation électrique par trimestre de 1997 à 1999

	1997	1998	1999
T1	4.5	5.5	7.2
T2	4.1	4.9	6.4

	1997	1998	1999
T3	3.7	4.4	4.8
T4	5.1	6.5	6.8

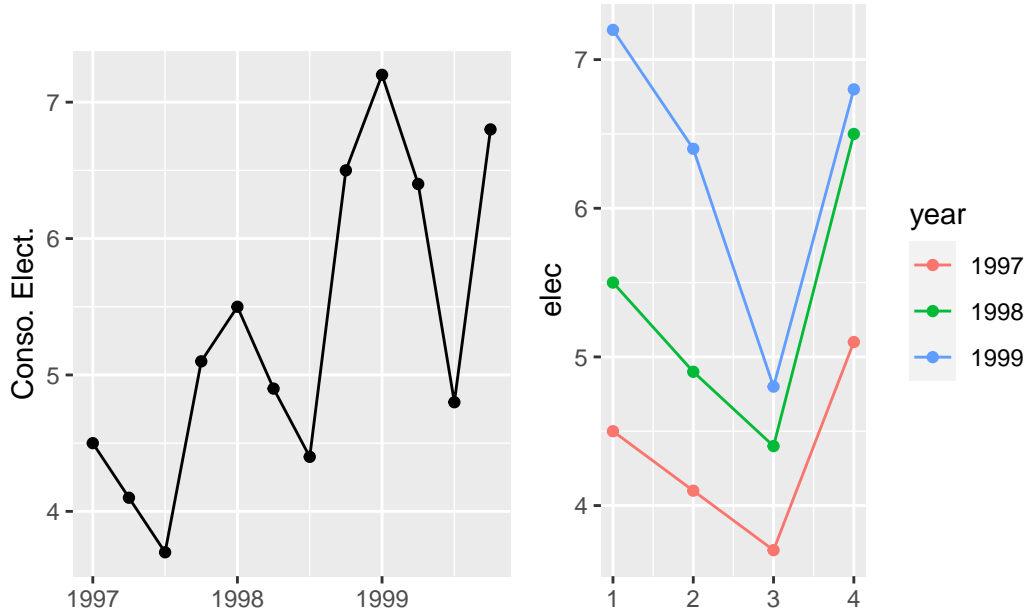


Figure 2.9: Représentation de la consommation électrique par trimestre

On commence donc par appliquer la moyenne mobile M_4 sur la série Y_t pour estimer la tendance : $\hat{m}_t = M_4 Y_t$

	Qtr1	Qtr2	Qtr3	Qtr4
1997	NA	NA	4.4750	4.7000
1998	4.8875	5.1500	5.5375	5.9375
1999	6.1750	6.2625	NA	NA

On forme alors la série corrigée de la tendance $\hat{Y}^{\text{CT}} = Y_t - \hat{m}_t$:

	Qtr1	Qtr2	Qtr3	Qtr4
1997	NA	NA	-0.7750	0.4000
1998	0.6125	-0.2500	-1.1375	0.5625
1999	1.0250	0.1375	NA	NA

On calcule alors les termes ω_k

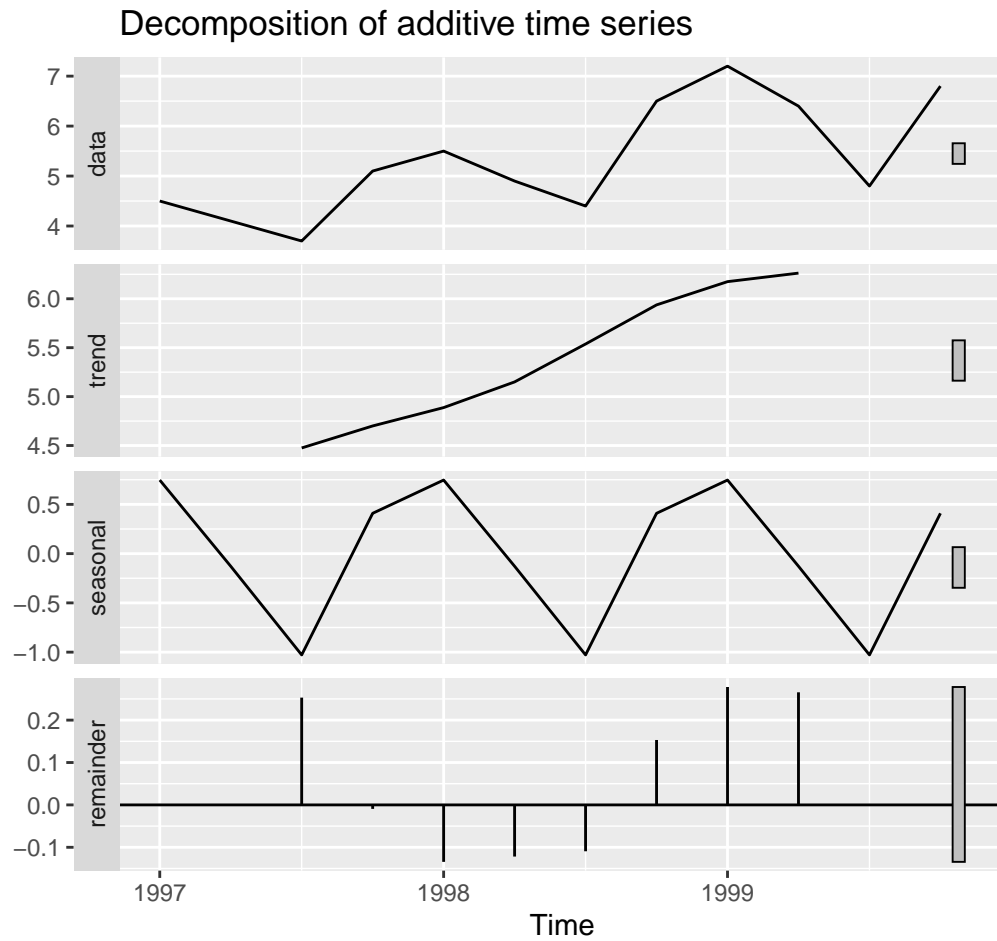
$$\begin{aligned}\omega_1 &= \frac{1}{2}(\hat{Y}_5^{\text{CT}} + \hat{Y}_9^{\text{CT}}) \approx 0.8188 \\ \omega_2 &= \frac{1}{2}(\hat{Y}_6^{\text{CT}} + \hat{Y}_{10}^{\text{CT}}) \approx -0.0562 \\ \omega_3 &= \frac{1}{2}(\hat{Y}_3^{\text{CT}} + \hat{Y}_7^{\text{CT}}) \approx -0.9562 \\ \omega_4 &= \frac{1}{2}(\hat{Y}_4^{\text{CT}} + \hat{Y}_8^{\text{CT}}) \approx 0.4812\end{aligned}$$

et $\sum_{k=1}^4 \omega_k = 0.0719$. On en déduit alors la saisonnalité estimée \hat{s}_t :

	Qtr1	Qtr2	Qtr3	Qtr4
1997	0.746875	-0.128125	-1.028125	0.409375
1998	0.746875	-0.128125	-1.028125	0.409375
1999	0.746875	-0.128125	-1.028125	0.409375

Cette procédure est faite par la fonction `decompose()` :

```
elec<-c(4.5,4.1,3.7,5.1,5.5,4.9,4.4,6.5,7.2,6.4,4.8,6.8)
autoplot(decompose(ts(data=elec,start=c(1997,1),end=c(1999,4),frequency=4),type="additive"
```

Example 2.9. On reprend la série *AirPassengers* (voir Example 1.4). La Figure 2.10 donne la sortie de la fonction `decompose()` appliquée sur la série logtransformée. On retrouve la saisonnalité (de période une année) et une tendance croissante pratiquement linéaire.

```
autoplot(decompose(log(AirPassengers),type="additive"))
```

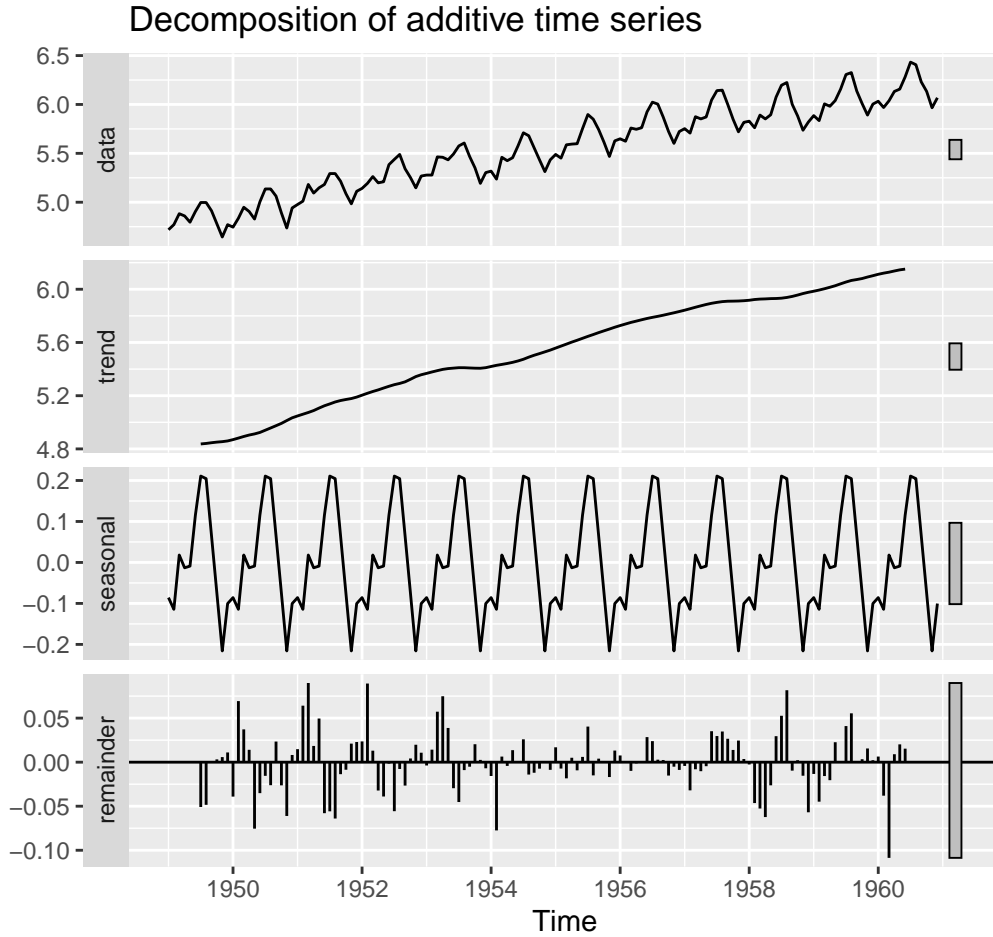


Figure 2.10: Sortie de la fonction `decompose()` sur la série `AirPassengers` logtransformée.

2.3.3 Elimination de la saisonnalité par différenciation

Pour supprimer une saisonnalité de périodicité r , on peut utiliser une méthode de différenciation en utilisant l'opérateur $\nabla_r = I - B^r$ à la série temporelle $(Y_t)_{t \in T}$ afin d'obtenir la série

$$\nabla_r Y_t = Y_t - Y_{t-r}, \quad \forall t \geq r$$

Cette série ne contient plus de saisonnalité car $\nabla_r Y_t = m_t - m_{t-r} + X_t - X_{t-r}$.

On peut alors utiliser une des techniques vues précédemment pour estimer / éliminer la tendance en l'absence de saisonnalité sur la série désaisonnalisée $(\nabla_r Y_t)_{t \geq r}$.

⚠ $\nabla_r \neq \nabla^r$

Attention à ne pas confondre l'opérateur ∇_r et l'opérateur ∇^k (voir Définition 2.5).

Exemple 2.10. On applique l'opérateur de différenciation ∇_{12} à la série *AirPassengers* log-transformée. Figure 2.11 montre la série initiale à gauche et la série différenciée $\nabla_{12}Y_t$ à droite. Cette dernière ne présente plus de saisonnalité.

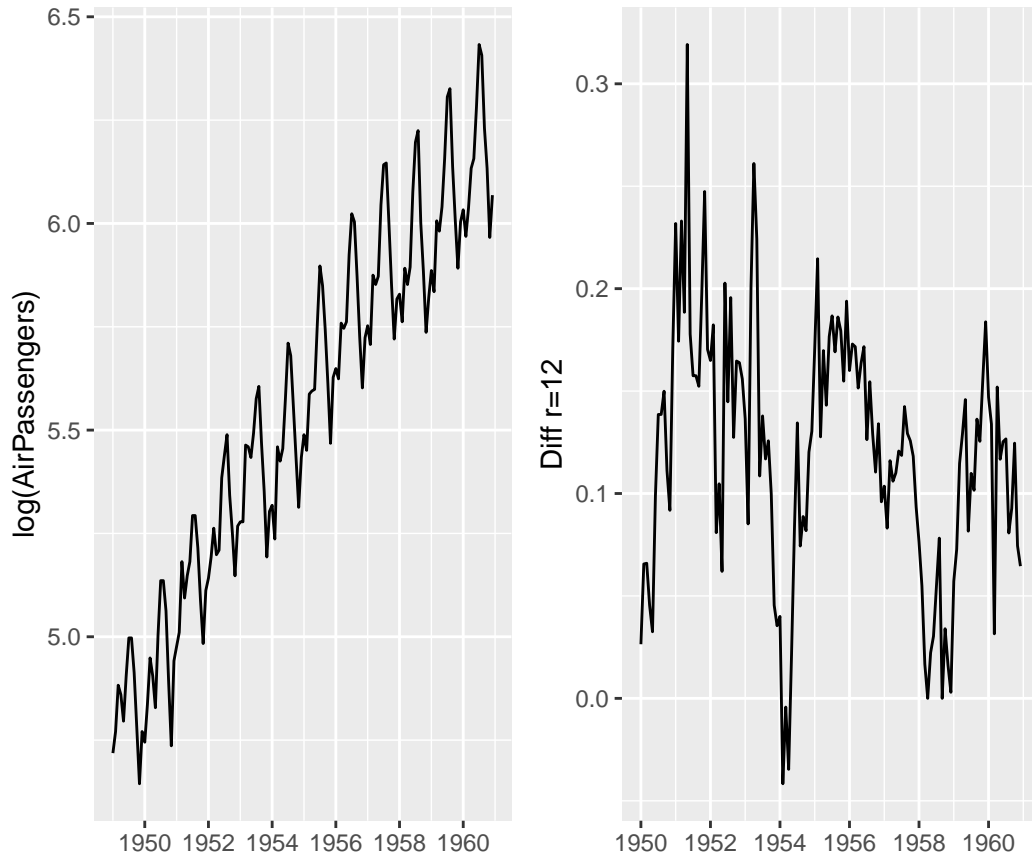


Figure 2.11: Illustration de la différenciation de période 12 sur la série *AirPassengers* logtransformée.

3 Modélisation aléatoire des séries temporelles

Dans la suite du cours, nous allons aller vers une modélisation aléatoire des séries temporelles. Une série temporelle sera vue comme une réalisation d'un processus stochastique avec certaines propriétés. Ce chapitre pose donc les définitions et propriétés importantes pour la suite autour des notions de processus du second ordre et de processus stationnaire.

3.1 Processus stochastiques

3.1.1 Définition

Definition 3.1. On appelle **processus stochastique** toute famille de variables aléatoires (v.a.) $(X_t)_{t \in T}$ d'un espace probabilisé (Ω, \mathcal{A}, P) vers un espace probabilisable (E, \mathcal{E})

$$\forall t \in T, X_t \text{ est une v.a. de } (\Omega, \mathcal{A}, P) \text{ vers } (E, \mathcal{E}).$$

L'ensemble T est appelé **espace des temps** et E **espace des états**. Chacun de ces espaces peut être discret ou continu.

Pour $\omega \in \Omega$, on appelle **trajectoire du processus** la fonction (déterministe) :

$$t \mapsto X_t(\omega).$$

Dans ce cours, les séries temporelles sont vues comme des réalisations d'un processus stochastique à **espace des temps discret**, à savoir $T = \mathbb{N}$, ou pour des raisons mathématiques $T = \mathbb{Z}$. Elles ne sont observées que sur un intervalle de temps fini (on observe qu'une partie de la trajectoire du processus). En général, on considère $E = \mathbb{R}$ comme espace des états et la tribu borélienne $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$.

Dans ce cours on n'abordera pas les **processus spatiaux** ($\dim(T) \geq 2$), ni les séries temporelles **multivariées** ($E = \mathbb{R}^n$).

3.1.2 Premiers exemples de processus

3.1.2.1 Bruit blanc fort / faible

Definition 3.2. Un **bruit blanc fort** est une suite de variables indépendantes et identiquement distribuées (i.i.d.) $(X_t)_{t \in T}$ centrées et de variance σ^2 .

On note $(X_t)_{t \in T} \sim \text{IID}(0, \sigma^2)$.

Dans le cas d'un bruit blanc fort, il n'y a aucune dépendance entre les observations. En particulier, la connaissance de X_1, \dots, X_n n'informe en rien sur la valeur (future) de X_{n+h} . Si la loi commune des v.a. est gaussienne, on parle de bruit blanc gaussien (la Figure 3.1 donne un exemple de simulation d'une partie de trajectoire).

Definition 3.3. On appelle **bruit blanc gaussien**, tout bruit blanc fort pour lequel la loi commune des v.a.r. $(X_t)_{t \in T}$ est une $\mathcal{N}(0, \sigma^2)$.

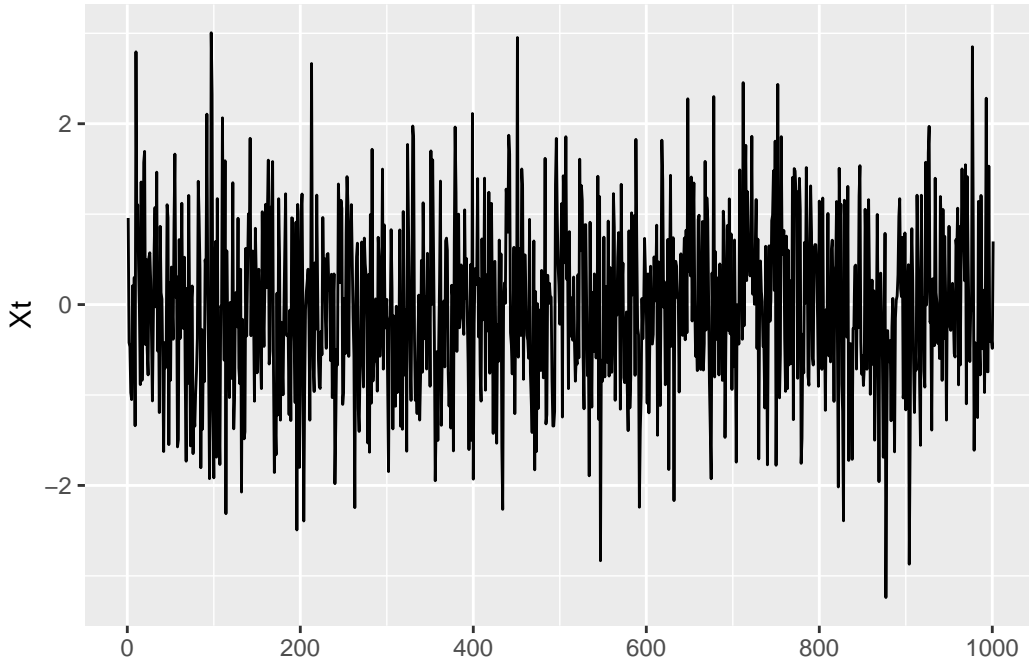


Figure 3.1: Trajectoire d'un bruit blanc gaussien $\mathcal{N}(0, 1)$

L'hypothèse d'i.i.d est très forte, on peut considérer une version faible d'un bruit blanc en ne demandant que la non-corrélation.

Definition 3.4. On appelle **bruit blanc faible** toute suite de v.a.r. $(X_t)_{t \in T}$

- centrées et de variance σ^2
- **non corrélées**

On note $(X_t)_{t \in T} \sim \text{WN}(0, \sigma^2)$

En tant que tels, ces processus n'ont pas d'intérêt direct pour un objectif de prévision. Mais nous verrons qu'ils jouent un rôle important dans la modélisation de séries temporelles plus complexes.

3.1.2.2 Marche aléatoire

A partir d'un bruit blanc fort, on peut construire une marche aléatoire.

Définition 3.5. Une **marche aléatoire** $(S_t)_{t \in \mathbb{N}}$ est obtenue par

$$S_t = X_1 + \dots + X_t, \quad \forall t \in \mathbb{N}$$

où $(X_t)_{t \in \mathbb{N}}$ est un bruit blanc fort.

Si ce dernier est un processus binaire ($\mathbb{P}(X_t = 1) = \mathbb{P}(X_t = -1) = \frac{1}{2}$) alors la marche aléatoire $(S_t)_{t \in \mathbb{N}}$ est dite **symétrique simple**.

La Figure 3.2 montre des exemples de trajectoires de marches aléatoires.

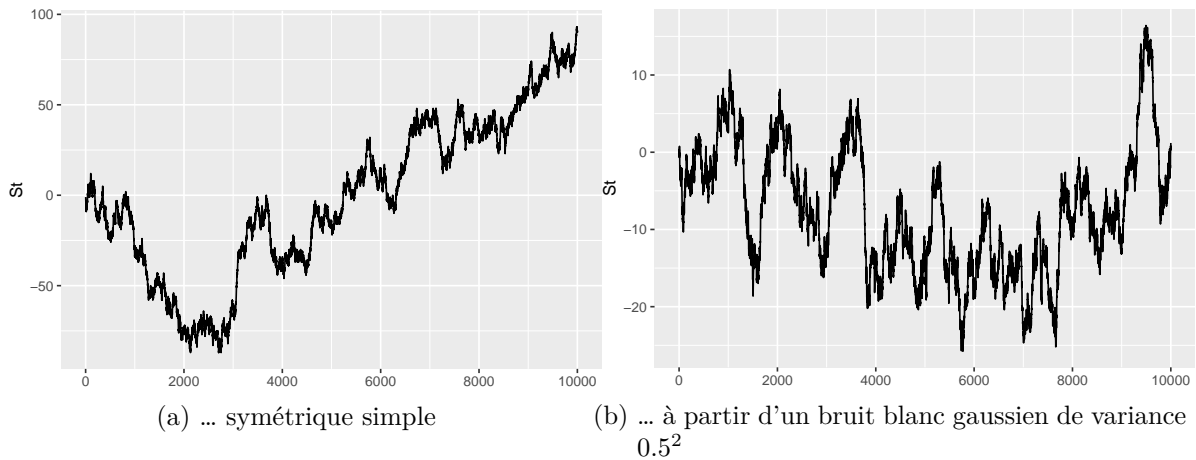


Figure 3.2: Trajectoire d'une marche aléatoire ...

3.1.2.3 Processus gaussien

Definition 3.6. Un **processus gaussien** à temps discret $(X_t)_{t \in \mathbb{Z}}$ est une série temporelle telle que la loi de n'importe quel vecteur extrait est gaussien

$$\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n) \in \mathbb{Z}^n : (X_{t_1}, \dots, X_{t_n}) \text{ est un vecteur gaussien.}$$

Un bruit blanc gaussien est donc un processus gaussien.

3.1.2.4 Processus MA(1)

Definition 3.7. On appelle **processus moyenne mobile d'ordre 1** toute série temporelle $(X_t)_{t \in \mathbb{Z}}$ définie par :

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \forall t \in \mathbb{Z},$$

où $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et $\theta \in \mathbb{R}$.

On note $(X_t)_{t \in \mathbb{Z}} \sim \text{MA}(1)$.

La Figure 3.3 donne des exemples de trajectoires de processus MA(1). Cette notion de processus moyenne mobile sera étendue à des ordres supérieurs et fera l'objet d'une étude approfondie dans la Section 5.3 du Chapitre 5.

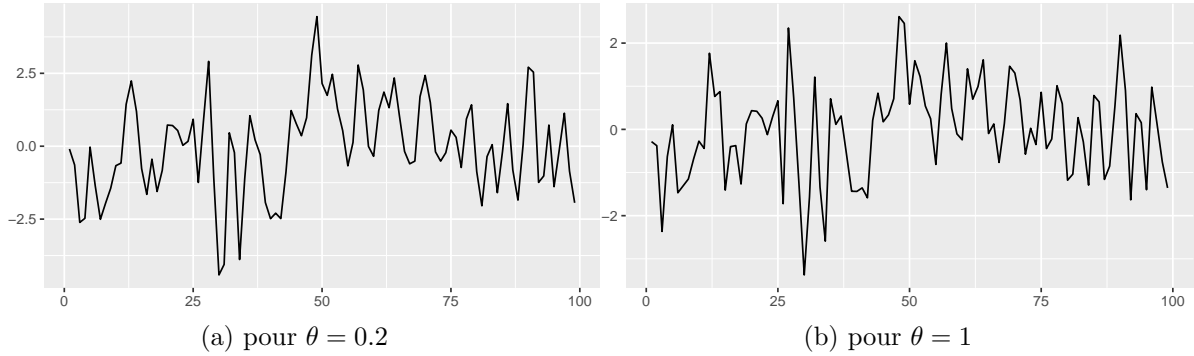


Figure 3.3: Exemple de trajectoires d'un processus MA(1)

3.1.2.5 Processus AR(1)

Definition 3.8. On appelle **processus autorégressif d'ordre 1** toute série temporelle $(X_t)_{t \in \mathbb{Z}}$ définie par :

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varepsilon_t, \quad \forall t \in \mathbb{Z},$$

où $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et $(\varphi_0, \varphi_1) \in \mathbb{R}^2$.

On note $(X_t)_{t \in \mathbb{Z}} \sim \text{AR}(1)$.

La Figure 3.4 donne des exemples de trajectoires de processus AR(1). Cette notion de processus autorégressif sera étendue à des ordres supérieurs et fera l'objet d'une étude approfondie dans la Section 5.2 du Chapter 5.

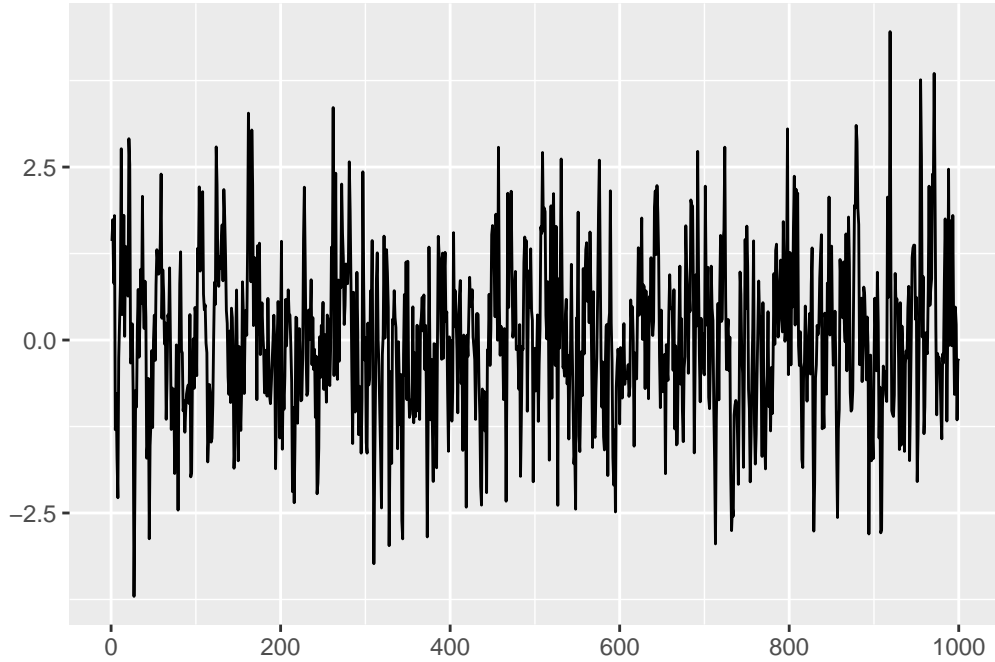


Figure 3.4: Exemple de trajectoire d'un processus AR(1)

3.2 Rappels sur l'espace L^2 et les processus du second ordre

3.2.1 Espace L^2

Definition 3.9. On dit qu'une v.a.r. X est dans l'espace $L^2(\Omega, \mathcal{A}, P)$ si l'on a $\mathbb{E}[X^2] < +\infty$. Cet espace $L^2(\Omega, \mathcal{A}, P)$ est un espace de Hilbert muni du produit scalaire

$$\langle X, Y \rangle_{L^2} = \mathbb{E}[XY]$$

et de la norme

$$\|X\|_{L^2} = \sqrt{\mathbb{E}[X^2]}.$$

Dans cet espace L^2 , on a donc que X et Y sont orthogonaux ($X \perp Y$) si $\mathbb{E}[XY] = 0$. Si de plus les v.a. X et Y sont centrées, cela revient à avoir $\text{Cov}(X, Y) = 0$.

Theorem 3.1 (Inégalité de Cauchy-Schwarz). Si X et Y sont deux v.a. dans $L^2(\Omega, \mathcal{A}, P)$ alors la v.a. XY est dans $L^1(\Omega, \mathcal{A}, P)$ et on a :

$$\|XY\|_{L^1} = \mathbb{E}[|XY|] \leq \|X\|_{L^2} \|Y\|_{L^2}.$$

Definition 3.10. Un processus $(X_t)_{t \in T}$ est dit du **second ordre** si la v.a. X_t est dans $L^2(\Omega, \mathcal{A}, P)$ pour tout t dans T .

3.2.2 Convergence dans L^2

Definition 3.11. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. dans $L^2(\Omega, \mathcal{A}, P)$ et X une v.a. dans $L^2(\Omega, \mathcal{A}, P)$. On dit que $(X_n)_{n \in \mathbb{N}}$ **converge dans L^2** vers X si l'on a :

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_{L^2} = 0.$$

Notation : $X_n \xrightarrow[n \rightarrow +\infty]{L^2} X$

La proposition suivante nous sera très utile dans la suite pour autoriser l'inversion entre le signe somme et l'espérance.

Proposition 3.1. Si $\sum_{i=0}^n X_i$ converge dans L^2 vers $\sum_{i=0}^{+\infty} X_i$ alors on a

$$\mathbb{E} \left[\sum_{i=0}^{+\infty} X_i \right] = \sum_{i=0}^{+\infty} \mathbb{E}[X_i].$$

Preuve

Notons $U_n = \sum_{i=0}^n X_i$ et $U = \sum_{i=0}^{+\infty} X_i$. L'inégalité classique entre les normes L^1 et L^2 (obtenue par l'inégalité de Jensen par exemple) nous donne les inégalités :

$$|\mathbb{E}[U_n] - \mathbb{E}[U]| \leq \mathbb{E}[|U_n - U|] = \|U_n - U\|_{L^1} \leq \|U_n - U\|_{L^2}.$$

L'hypothèse que (U_n) converge dans L^2 vers U nous assure donc que $\mathbb{E}[U_n] = \sum_{i=0}^n \mathbb{E}[X_i]$ converge vers $\mathbb{E}[U] = \mathbb{E}[\sum_{i=0}^{+\infty} X_i]$, dont découle l'égalité annoncée.

La Proposition 3.1 reste valable pour une série indexée dans les deux directions :

si $\sum_{i=-m}^n X_i$ converge dans L^2 vers $\sum_{i=-\infty}^{+\infty} X_i$ quand n et m tendent vers $+\infty$, on a

$$\mathbb{E} \left[\sum_{i=-\infty}^{+\infty} X_i \right] = \sum_{i=-\infty}^{+\infty} \mathbb{E}[X_i].$$

Proposition 3.2. *Si les séries $\sum_{i=0}^n X_i$ et $\sum_{j=0}^n Y_j$ convergent dans L^2 vers $\sum_{i=0}^{+\infty} X_i$ et $\sum_{j=0}^{+\infty} Y_j$ respectivement alors on a*

$$\mathbb{E} \left[\sum_{i=0}^{+\infty} X_i \cdot \sum_{j=0}^{+\infty} Y_j \right] = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \mathbb{E}[X_i Y_j].$$

Corollary 3.1. *Si les séries $\sum_{i=0}^n X_i$ et $\sum_{j=0}^n Y_j$ convergent dans L^2 vers $\sum_{i=0}^{+\infty} X_i$ et $\sum_{j=0}^{+\infty} Y_j$ respectivement alors on a*

$$\text{Cov} \left(\sum_{i=0}^{+\infty} X_i, \sum_{j=0}^{+\infty} Y_j \right) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \text{Cov}(X_i, Y_j).$$

3.2.3 Projection orthogonale

Les projections dans l'espace L^2 s'appuient sur la propriété d'Hilbert de cet espace. Rappelons d'abord la définition d'un sous-espace vectoriel fermé d'un espace de Hilbert dans notre cadre.

Définition 3.12. Un **sous-espace vectoriel** \mathcal{H} de l'espace $L^2(\Omega, \mathcal{A}, P)$ est dit **fermé** s'il contient toutes ses limites

$$\left(X_n \in \mathcal{H}, \forall n \in \mathbb{N}, \text{ et } X_n \xrightarrow[n \rightarrow +\infty]{L^2} X \right) \implies X \in \mathcal{H}.$$

Theorem 3.2 (Théorème de projection). *Soit X une v.a. dans $L^2(\Omega, \mathcal{A}, P)$ et \mathcal{H} un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{A}, P)$. Il existe alors une unique v.a. \hat{X} dans \mathcal{H} telle que :*

$$\|X - \hat{X}\|_{L^2} = \min_{Y \in \mathcal{H}} \|X - Y\|_{L^2}.$$

*La v.a. \hat{X} correspond à la **projection orthogonale** de X sur l'espace \mathcal{H} et est donc telle que :*

$$\hat{X} \in \mathcal{H} \text{ et } X - \hat{X} \perp \mathcal{H}.$$

3.3 Processus stationnaires

On a vu dans le premier chapitre qu’une série temporelle peut souvent être décomposée sous la forme

$$Y_t = m_t + s_t + X_t,$$

où

- m_t est une fonction à variation lente appelée tendance
- s_t une fonction périodique (de somme nulle) appelée saisonnalité.

Le terme restant, le processus X_t , est donc supposé être “plus stable” dans un sens que l’on va ici définir. C’est la notion de **stationnarité** que l’on va expliquer dans cette section.

Definition 3.13. Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit **stationnaire au sens fort** si la loi de tout vecteur $(X_{t_1}, \dots, X_{t_n})$ est invariante par translation temporelle :

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \dots, X_{t_n+h}), \quad \forall (t_1, \dots, t_n) \in \mathbb{Z}^n \text{ et } h \in \mathbb{Z}.$$

Par conséquent, sous la stationnarité forte, toutes les v.a.r X_t ont la même loi. On a immédiatement qu’un bruit blanc fort est stationnaire au sens fort.

Cette hypothèse de stationnarité forte est très contraignante et peu réaliste en pratique. Aussi on va lui préférer la notion de **stationnarité faible**. Pour introduire la notion de stationnarité faible pour les processus du second ordre, nous avons besoin de définir tout d’abord la **fonction moyenne** et la **fonction covariance** des processus du second ordre.

Definition 3.14. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus du second ordre.

La **fonction moyenne** de $(X_t)_{t \in \mathbb{Z}}$ est définie par

$$t \in \mathbb{Z} \mapsto \mathbb{E}[X_t].$$

La **fonction covariance** de $(X_t)_{t \in \mathbb{Z}}$ est définie par :

$$\begin{aligned} (s, t) \in \mathbb{Z}^2 \mapsto \text{Cov}(X_s, X_t) &= \mathbb{E}[(X_s - \mathbb{E}[X_s])(X_t - \mathbb{E}[X_t])] \\ &= \langle X_s - \mathbb{E}[X_s], X_t - \mathbb{E}[X_t] \rangle_{L^2} \\ &= \mathbb{E}[X_s X_t] - \mathbb{E}[X_s] \mathbb{E}[X_t] \end{aligned}$$

Definition 3.15. Un processus du second ordre $(X_t)_{t \in \mathbb{Z}}$ est dit **faiblement stationnaire** si

- sa fonction moyenne est constante $\mathbb{E}[X_t] = \mu_X, \quad \forall t \in \mathbb{Z}$;
- sa fonction covariance ne dépend que de la différence en temps

$$\begin{aligned} \text{Cov}(X_s, X_t) &= \text{Cov}(X_{s+u}, X_{t+u}), \quad \forall (s, t, u) \in \mathbb{Z}^3 \\ &= \text{Cov}(X_0, X_{t-s}) = f(t-s) \end{aligned}$$

Remarques

- Un processus fortement stationnaire est stationnaire au sens faible. Mais la réciproque est fausse !
- Si le processus est faiblement stationnaire alors $\text{Var}(X_t) = \text{Var}(X_0)$, $\forall t \in \mathbb{Z}$.

Exemple 3.1 (Bruit blanc fort / faible). Si le processus $(X_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort dans L^2 alors il est faiblement stationnaire puisque :

$$\mu_X(t) = 0, \forall t \in \mathbb{Z} \text{ et } \text{Cov}(X_{t+h}, X_t) = \sigma^2 \mathbb{1}_{h=0} = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}, \forall t \in \mathbb{Z}$$

Un bruit blanc faible $\text{WN}(0, \sigma^2)$ est également faiblement stationnaire.

Exemple 3.2 (Marche aléatoire non stationnaire). Soit $(S_t)_{t \in \mathbb{N}}$ la marche aléatoire $S_t = X_1 + \dots + X_t$ avec $(X_t)_{t \in \mathbb{Z}} \sim \text{IID}(0, \sigma^2)$. On a

$$\begin{aligned} \mathbb{E}[S_t] &= \mathbb{E} \left[\sum_{k=1}^t X_k \right] = \sum_{k=1}^t \mathbb{E}[X_k] = 0 \\ \text{Var}(S_t) &= \text{Var} \left(\sum_{k=1}^t X_k \right) = \sum_{k=1}^t \text{Var}(X_k) = t\sigma^2. \end{aligned}$$

\Rightarrow le processus $(S_t)_{t \in \mathbb{N}}$ **n'est pas stationnaire** puisque sa variance n'est pas constante.

On peut le vérifier plus largement sur la fonction covariance : pour $h \in \mathbb{N}$,

$$\begin{aligned} \text{Cov}(S_{t+h}, S_t) &= \text{Cov}(S_t + X_{t+1} + \dots + X_{t+h}, S_t) \\ &= \text{Var}(S_t) + \sum_{k=1}^h \text{Cov}(X_{t+k}, S_t) = \text{Var}(S_t) = t\sigma^2. \end{aligned}$$

3.4 Fonctions d'autocovariance et d'autocorrélation

3.4.1 Fonction d'autocovariance (ACVF)

Définition 3.16. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus faiblement stationnaire.

Sa **fonction d'autocovariance** $\gamma_X(\cdot)$ est la fonction définie par

$$\begin{aligned} \gamma_X : \mathbb{Z} &\rightarrow \mathbb{R} \\ h &\mapsto \gamma_X(h) = \text{Cov}(X_h, X_0) = \text{Cov}(X_{t+h}, X_t), \forall t. \end{aligned}$$

Le tracé de la fonction $h \in \mathbb{N} \mapsto \gamma_X(h)$ est appelé **autocovariogramme**.

Proposition 3.3. *La fonction d'autocovariance $\gamma_X(\cdot)$ vérifie les propriétés suivantes :*

- $\gamma_X(0) \geq 0$
- $|\gamma_X(h)| \leq \gamma_X(0)$ pour tout h donc l'ACVF est bornée
- $\gamma_X(h) = \gamma_X(-h)$ pour tout h donc l'ACVF est une fonction paire

Preuve

- On a $\gamma_X(0) = \text{Var}(X_t) \geq 0$.
- Comme une corrélation est toujours comprise dans l'intervalle $[-1, 1]$, on a donc $|\gamma_X(h)| \leq \gamma_X(0)$.
- On peut écrire $\gamma_X(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma_X(-h)$.

3.4.2 Fonction d'autocorrélation (ACF)

Definition 3.17. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus faiblement stationnaire.

Sa fonction **d'autocorrélation** $\rho_X(\cdot)$ est définie par

$$\begin{aligned} \rho_X : \mathbb{Z} &\rightarrow [-1, 1] \\ h &\mapsto \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Corr}(X_{t+h}, X_t) \\ &= \frac{\text{Cov}(X_{t+h}, X_t)}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}}, \forall t. \end{aligned}$$

Le tracé de la fonction $h \in \mathbb{N} \mapsto \rho_X(h)$ est appelé **autocorrélogramme**.

Ainsi les fonctions ACVF et ACF mesurent le degré de dépendance entre les valeurs d'une série temporelle à des instants différents. Ce sont des notions très importantes pour la suite du cours.

3.4.3 Exemples

Example 3.3 (Bruit blanc gaussien). On considère la série $(\varepsilon_t)_{t \in T}$ avec les ε_t i.i.d $\mathcal{N}(0, 1)$. La trajectoire observée de cette série temporelle est représentée en Figure 3.5a. En traçant le nuage de points de coordonnées (X_t, X_{t+1}) (voir Figure 3.5b), on constate qu'il n'y a pas de dépendance entre deux temps successifs. Sur l'autocorrélogramme empirique (Figure 3.5c), on a un pic à 1 pour $h = 0$ et des valeurs proches de 0 pour $h \geq 1$.

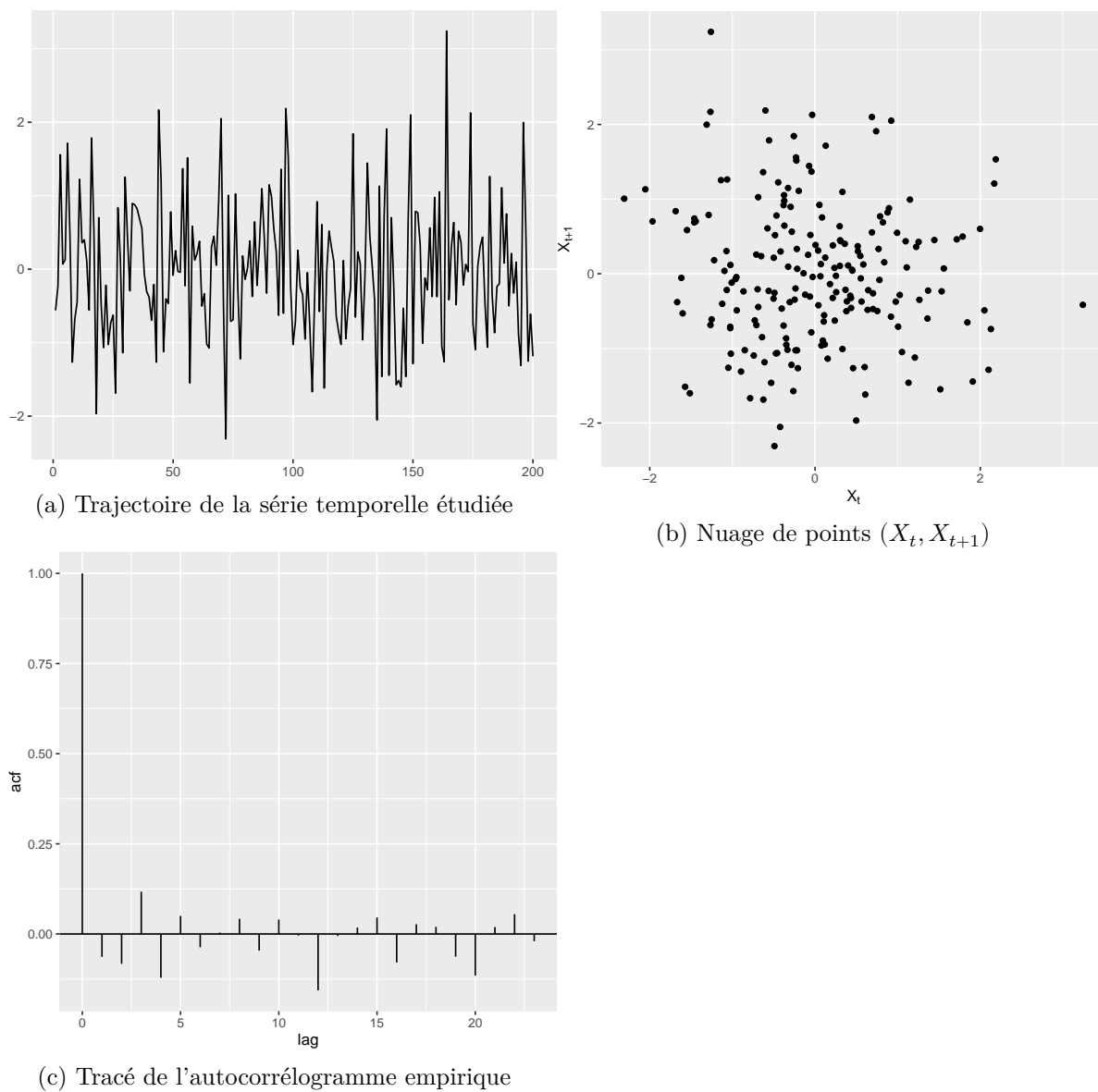


Figure 3.5: Exemple du bruit blanc gaussien

Exemple 3.4 (Moyenne mobile d'ordre 1 MA(1)). Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle MA(1) définie par

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_t \sim \text{WN}(0, \sigma^2).$$

C'est un processus stationnaire au sens faible car

- $\mathbb{E}[X_t] = \mathbb{E}[\varepsilon_t] + \theta \mathbb{E}[\varepsilon_{t-1}] = 0$
- La fonction d'autocovariance vaut

$$\begin{aligned} \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}[(\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1})] \\ &= \mathbb{E}[\varepsilon_t \varepsilon_{t+h}] + \mathbb{E}[\varepsilon_t \theta \varepsilon_{t+h-1}] \\ &\quad + \mathbb{E}[\theta \varepsilon_{t-1} \varepsilon_{t+h}] + \mathbb{E}[\theta \varepsilon_{t-1} \theta \varepsilon_{t+h-1}] \\ &= \sigma^2(\mathbb{1}_{h=0} + \theta \mathbb{1}_{h=1} + \theta \mathbb{1}_{h=-1} + \theta^2 \mathbb{1}_{h=0}) \\ &= \begin{cases} \sigma^2(1 + \theta^2) & \text{si } h = 0 \\ \theta \sigma^2 & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases} \end{aligned}$$

On en déduit l'expression de la fonction d'autocorrélation :

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

Pour illustrer, on observe la trajectoire d'une série temporelle issue du modèle suivant

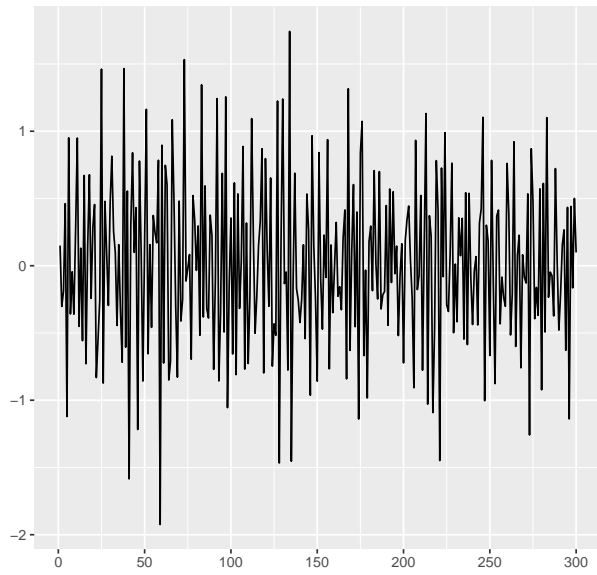
$$X_t = \varepsilon_t - 0.7\varepsilon_{t-1}, \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_t \sim \text{WN}(0, 0.5^2)$$

La trajectoire observée de cette série temporelle est représentée en Figure 3.6a. On constate une dépendance linéaire quand on trace le nuage de points de coordonnées (X_t, X_{t+1}) (voir Figure 3.6b) et aucune pour le nuage de points de coordonnées (X_t, X_{t+2}) (voir Figure 3.6c). Sur l'autocorrélogramme empirique (Figure 3.6d), on a un pic à 1 pour $h = 0$, un pic proche de $\frac{-0.7}{1+(-0.7)^2} \approx -0.47$ pour $h = 1$, et des valeurs proches de 0 pour $h \geq 2$.

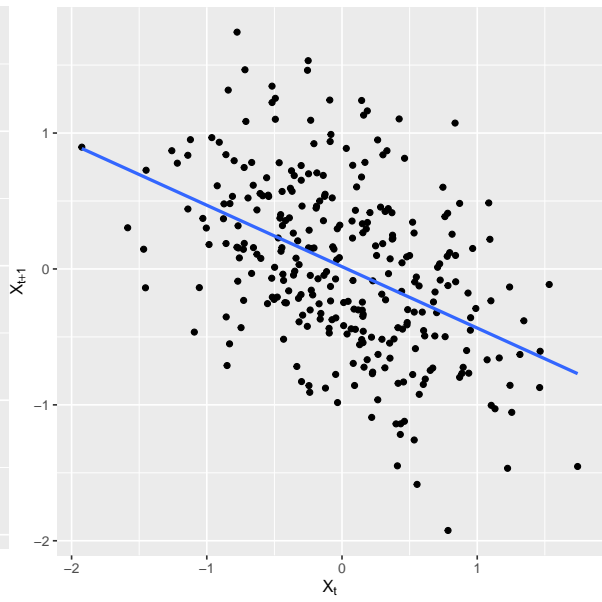
Comme second exemple, on observe la trajectoire d'une série temporelle issue du modèle suivant

$$X_t = \varepsilon_t + \varepsilon_{t-1}, \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_t \sim \text{WN}(0, 0.5^2)$$

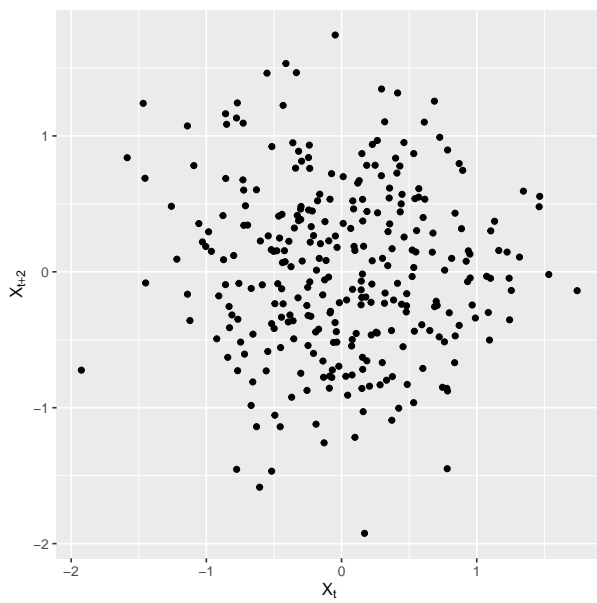
La trajectoire observée de cette série temporelle est représentée en Figure 3.7a. On constate une dépendance linéaire avec pente positive quand on trace le nuage de points de coordonnées (X_t, X_{t+1}) (voir Figure 3.7b) et aucune pour le nuage de points de coordonnées (X_t, X_{t+2}) (voir Figure 3.7c). Sur l'autocorrélogramme empirique (Figure 3.7d), on a un pic à 1 pour $h = 0$, un pic proche de $\frac{1}{1+(1)^2} = \frac{1}{2}$ pour $h = 1$, et des valeurs proches de 0 pour $h \geq 2$.



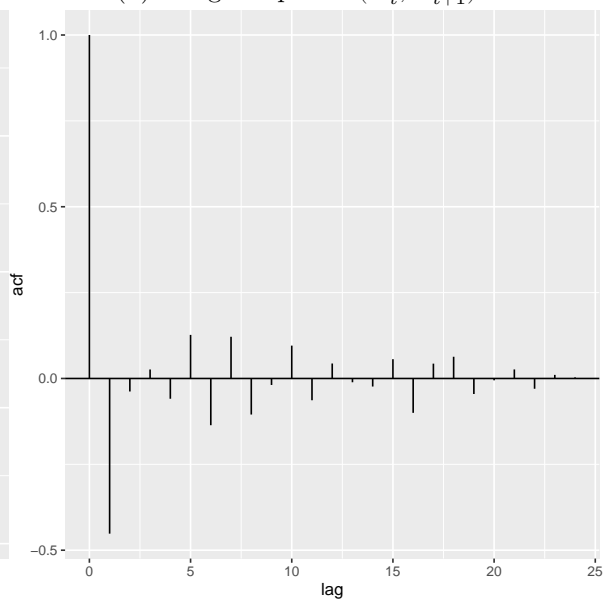
(a) Trajectoire de la série temporelle étudiée



(b) Nuage de points (X_t, X_{t+1})

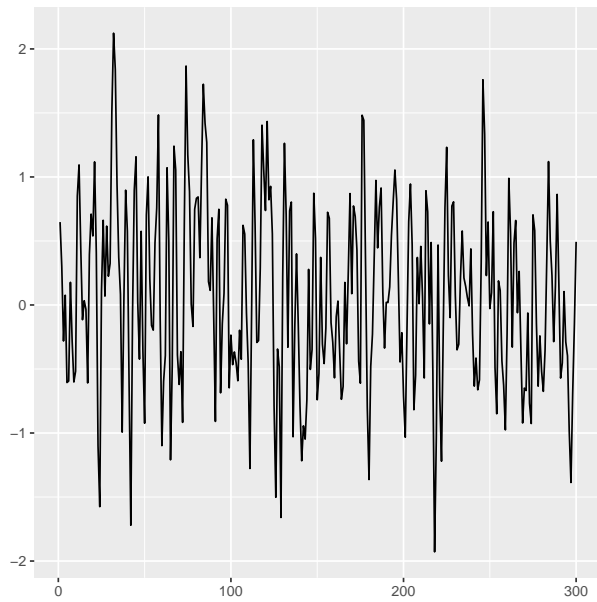


(c) Nuage de points (X_t, X_{t+2})

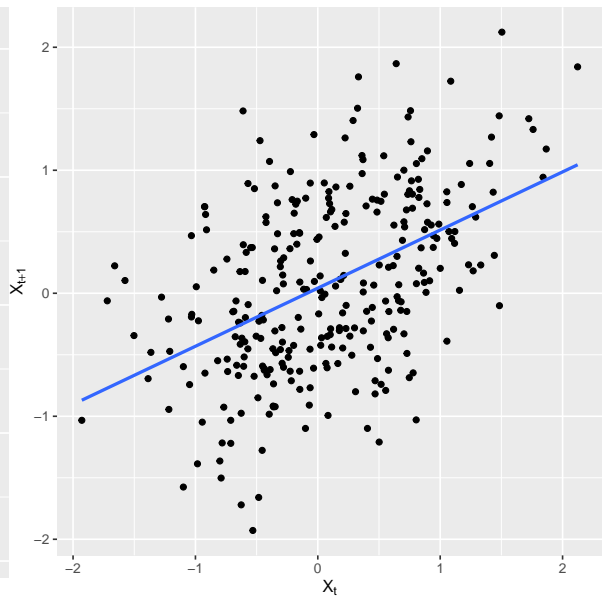


(d) Tracé de l'autocorrélogramme empirique

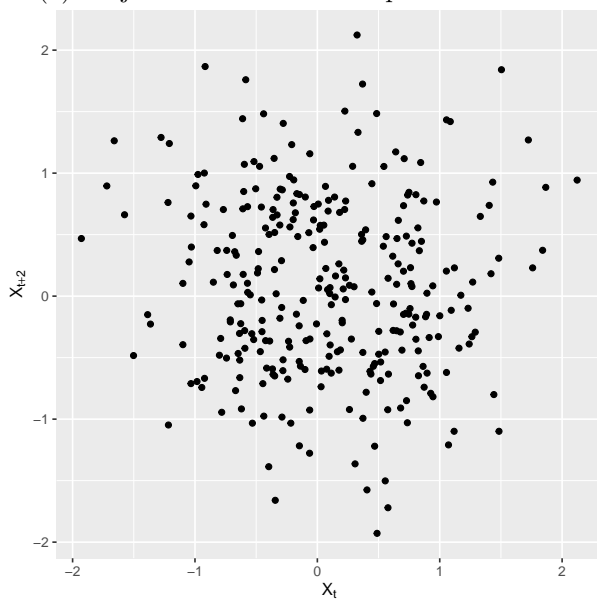
Figure 3.6: Résultats pour la série MA(1) $X_t = \varepsilon_t - 0.7\varepsilon_{t-1}$



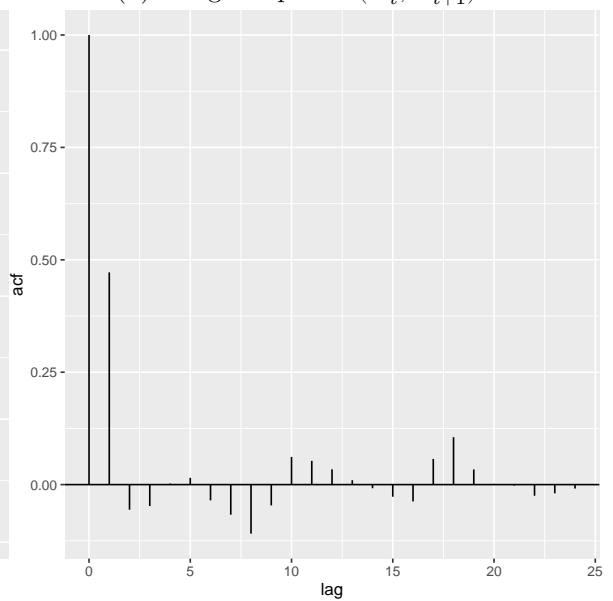
(a) Trajectoire de la série temporelle étudiée



(b) Nuage de points (X_t, X_{t+1})



(c) Nuage de points (X_t, X_{t+2})



(d) Tracé de l'autocorrélogramme empirique

Figure 3.7: Résultats pour la série MA(1) $X_t = \varepsilon_t + \varepsilon_{t-1}$

Exemple 3.5 (Processus autorégressif d'ordre 1 AR(1)). Soit $(X_t)_{t \in \mathbb{Z}}$ le processus stationnaire défini par

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t) \sim \text{WN}(0, \sigma^2) \text{ et } 0 < |\phi| < 1$$

On suppose que le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ est tel que, pour tout t ,

$$\begin{aligned} \text{Cov}(\varepsilon_t, X_s) = 0, \quad \forall s < t &\iff \langle \varepsilon_t, X_s \rangle_{L^2} = 0, \quad \forall s < t \\ &\iff \varepsilon_t \perp \mathcal{H}_{t-1} := \text{sp}\{X_{t-1}, X_{t-2}, \dots\}. \end{aligned}$$

Le processus X_t est centré car : $\mathbb{E}[X_t] = \phi \mathbb{E}[X_{t-1}] + \mathbb{E}[\varepsilon_t] = \phi \mathbb{E}[X_{t-1}] + 0$. Or le processus $(X_t)_{t \in \mathbb{Z}}$ est stationnaire donc $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}]$. On obtient donc que $\mathbb{E}[X_t] = \phi \mathbb{E}[X_t]$ d'où $\mathbb{E}[X_t] = 0$ car $\phi \neq 1$.

Déterminons maintenant la fonction d'autocovariance / autocorrélation : soit $h \in \mathbb{N}$,

$$\begin{aligned} \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}[X_t X_{t+h}] \\ &= \mathbb{E}[(\phi X_{t+h-1} + \varepsilon_{t+h}) X_t] \\ &= \phi \text{Cov}(X_{t+h-1}, X_t) + \text{Cov}(\varepsilon_{t+h}, X_t). \end{aligned}$$

Or par hypothèse, $\varepsilon_t \perp \mathcal{H}_{t-1}$ donc

$$\gamma_X(h) = \phi \gamma_X(h-1) = \dots = \phi^h \gamma_X(0).$$

D'où $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^h$, $\forall h > 0$. Finalement, comme $\gamma_X(h) = \gamma_X(-h)$, $\rho_X(h) = \phi^{|h|}$, $\forall h \in \mathbb{Z}$.

Pour illustrer ce résultat, on observe une série temporelle simulée selon

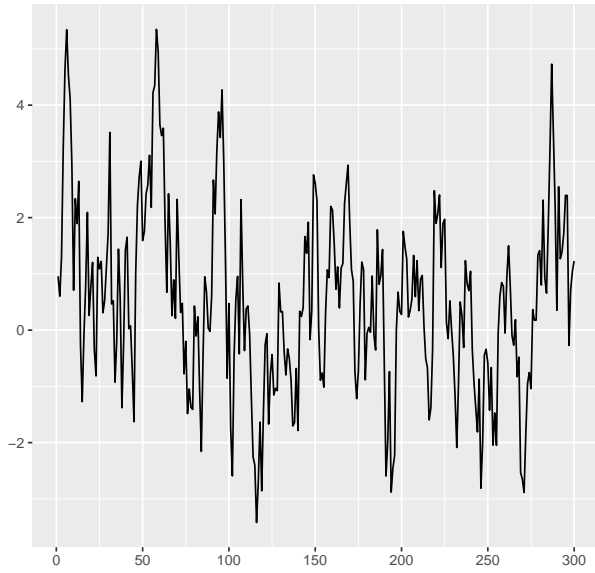
$$X_t = 0.8X_{t-1} + \varepsilon_t, \quad \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_t \sim \text{WN}(0, 0.5^2)$$

La trajectoire observée de cette série temporelle est représentée en Figure 3.8a. On constate des dépendances linéaires quand on trace le nuage de points de coordonnées (X_t, X_{t+1}) (voir Figure 3.8b) et le nuage de points de coordonnées (X_t, X_{t+2}) (voir Figure 3.8c). Sur l'autocorrélogramme empirique (Figure 3.8d), on constate la décroissance exponentielle vers 0 ($\rho_X(h) = e^{|h| \ln(0.8)} \approx e^{-0.22 |h|}$).

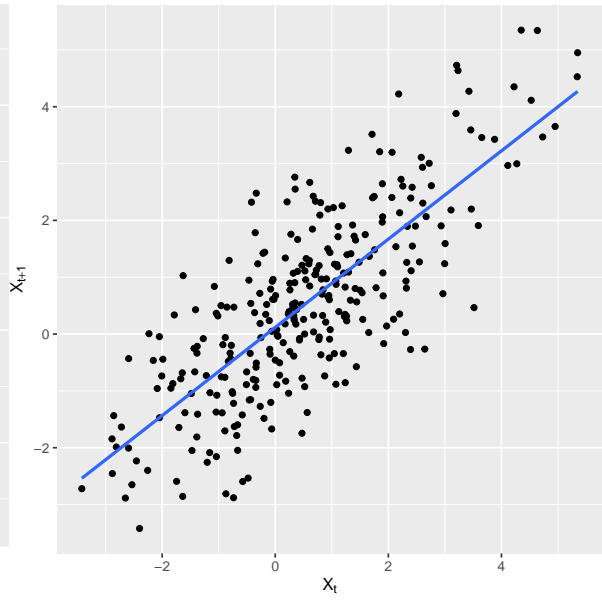
Pour la seconde illustration, on considère la série temporelle

$$X_t = -0.8X_{t-1} + \varepsilon_t, \quad \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_t \sim \text{WN}(0, 0.5^2)$$

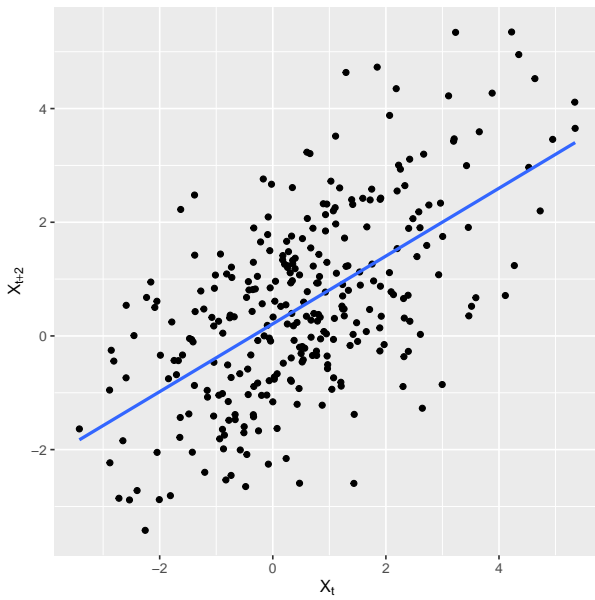
La trajectoire observée de cette série temporelle est représentée en Figure 3.9a. On constate des dépendances linéaires quand on trace le nuage de points de coordonnées (X_t, X_{t+1}) (voir Figure 3.9b) et le nuage de points de coordonnées (X_t, X_{t+2}) (voir Figure 3.9c), avec des pentes qui changent de signe car $\phi < 0$. Sur l'autocorrélogramme empirique (Figure 3.9d), on constate la décroissance vers 0 avec alternance du signe ($\rho_X(h) = (-1)^{|h|} e^{|h| \ln(0.8)} \approx (-1)^{|h|} e^{-0.22 |h|}$).



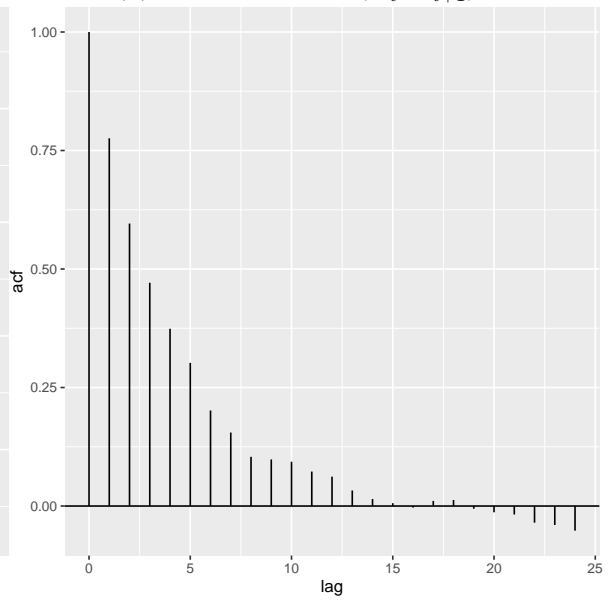
(a) Trajectoire de la série temporelle étudiée



(b) Nuage de points (X_t, X_{t+1})

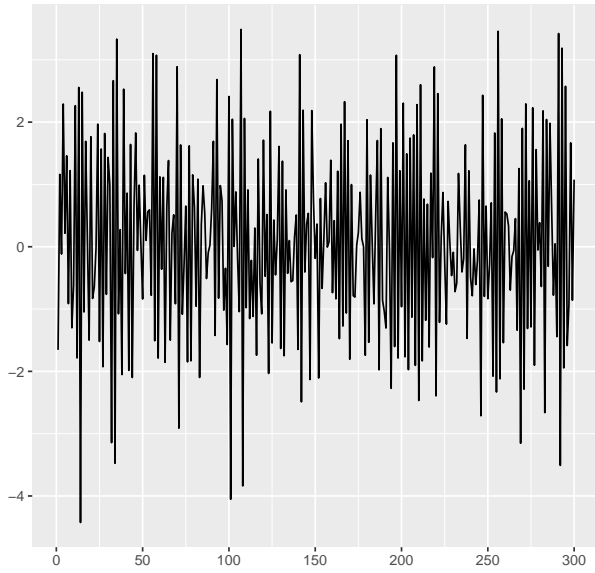


(c) Nuage de points (X_t, X_{t+2})

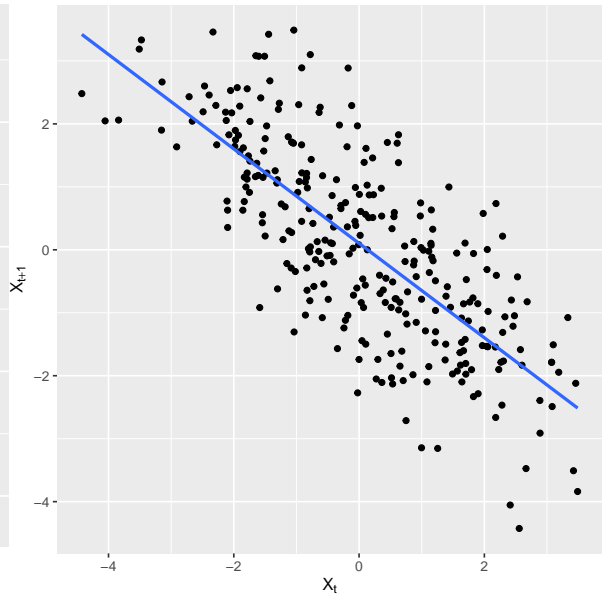


(d) Tracé de l'autocorrélogramme empirique

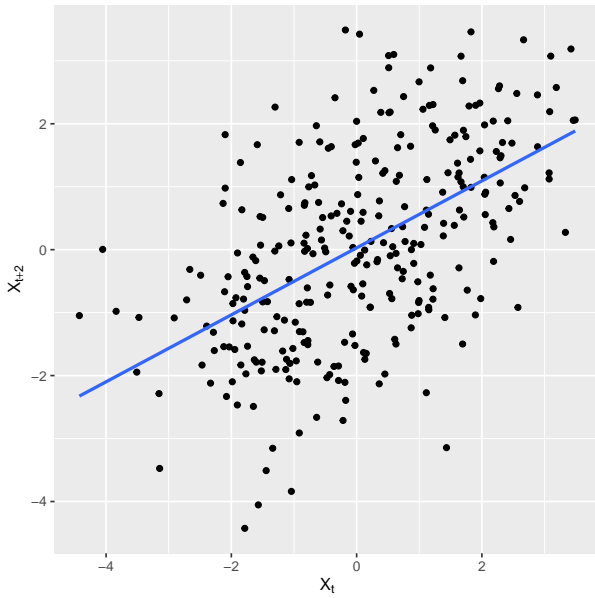
Figure 3.8: Résultats pour la série AR(1) $X_t = 0.8X_{t-1} + \varepsilon_t$



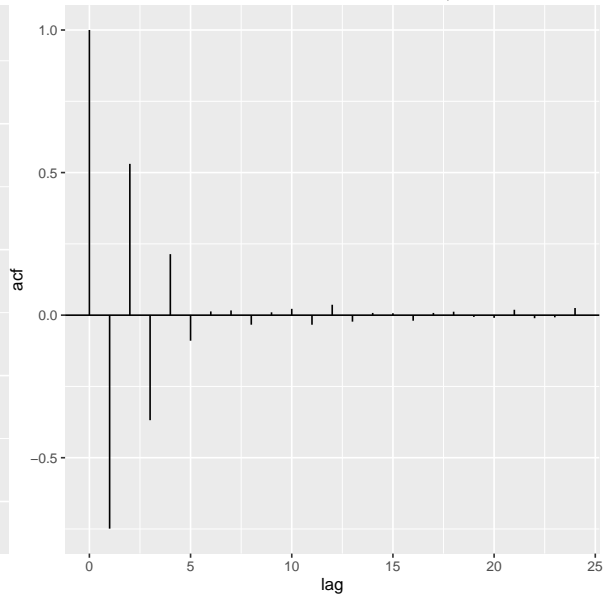
(a) Trajectoire de la série temporelle étudiée



(b) Nuage de points (X_t, X_{t+1})



(c) Nuage de points (X_t, X_{t+2})



(d) Tracé de l'autocorrélogramme empirique

Figure 3.9: Résultats pour la série AR(1) $X_t = -0.8X_{t-1} + \varepsilon_t$

3.4.4 CNS pour une fonction d'autocovariance

Definition 3.18. Une fonction $K : \mathbb{Z} \rightarrow \mathbb{R}$ est dite **semi-définie positive** si l'on a

$$\sum_{i,j=1}^n a_i a_j K(i-j) \geq 0,$$

pour tout n et tout vecteur $(a_1, \dots, a_n) \in \mathbb{R}^n$.

Theorem 3.3. Une fonction réelle définie sur \mathbb{Z} est une fonction d'autocovariance d'une série temporelle **si et seulement si** elle est paire et semi-définie positive.

Preuve

\Rightarrow) : Soit $\gamma_X(\cdot)$ la fonction d'autocovariance de $(X_t)_{t \in \mathbb{Z}}$. On a déjà montré précédemment qu'elle est paire. Montrons qu'elle est semi-définie positive: soit $(a_1, \dots, a_n) \in \mathbb{R}^n$,

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_X(i-j) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_{t-i}, X_{t-j}) \\ &= \text{Cov} \left(\sum_{i=1}^n a_i X_{t-i}, \sum_{j=1}^n a_j X_{t-j} \right) \\ &= \text{Var} \left(\sum_{i=1}^n a_i X_{t-i} \right) \geq 0. \end{aligned}$$

\Leftarrow) Soit $K : \mathbb{Z} \rightarrow \mathbb{R}$ paire et demi-définie positive. On veut montrer qu'il existe un processus stationnaire tel que $\gamma_X(\cdot) = K(\cdot)$.

$\forall n \in \mathbb{N}^*, \forall \mathbf{t} = (t_1, \dots, t_n) \in \mathbb{Z}^n$. Soit $F_{\mathbf{t}}$ la loi sur \mathbb{R}^n de fonction caractéristique

$$\Phi_{\mathbf{t}}(\mathbf{u}) = \exp \left[-\frac{1}{2} \mathbf{u}' \mathbb{K} \mathbf{u} \right]$$

où $\mathbf{u} = (u_1, \dots, u_n)' \in \mathbb{R}^n$ et la matrice $\mathbb{K} = (K(t_i - t_j))_{i,j=1,\dots,n}$. Comme K est semi-définie positive, \mathbb{K} est une matrice semi-définie positive et donc $\Phi_{\mathbf{t}}(\cdot)$ est la fonction caractéristique d'un vecteur gaussien $\mathcal{N}(0, \mathbb{K})$.

On peut alors vérifier que

$$\lim_{u_i \rightarrow 0} \Phi_{\mathbf{t}}(\mathbf{u}) = \Phi_{\mathbf{t}_{-i}}(\mathbf{u}_{-i}) = \exp \left[-\frac{1}{2} \mathbf{u}_{-i}' \mathbb{K}_{\mathbf{t}_{-i}} \mathbf{u}_{-i} \right]$$

où $\mathbf{t}_{-i} = (t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)'$ et $\mathbf{u}_{-i} = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n)'$. On conclut la preuve en utilisant le théorème d'existence de Kolmogorov : $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$ sont les fonctions de distribution d'un processus stochastique si et seulement si pour tout n ,

$\mathbf{t} = (t_1, \dots, t_n) \in \mathcal{T}$, $\mathbf{x} \in \mathbb{R}^n$ et $1 \leq k \leq n$,

$$\lim_{x_k \rightarrow \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}_{-k}}(\mathbf{x}_{-k})$$

où $\mathbf{t}_{-k} = (t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n)'$ et $\mathbf{x}_{-k} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)'$.

Remarque : Pour la fonction d'autocorrélation, on a les mêmes propriétés + $\rho_X(0) = 1$.

3.4.5 Matrice d'autocorrélation

Definition 3.19. Soit $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire du second ordre. On appelle **matrice d'autocorrélation** de (X_t, \dots, X_{t+h-1}) pour $h \in \mathbb{N}^*$

$$R_{X,h} = \begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_X(h-1) \\ \rho_X(1) & 1 & \ddots & \rho_X(h-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho_X(h-1) & \cdots & \rho_X(1) & 1 \end{pmatrix}.$$

Cette matrice d'autocorrélation est une **matrice de Toeplitz** (matrice à diagonales constantes) et

$$R_{X,h} = \left(\begin{array}{ccc|c} & & & \rho_X(h-1) \\ & & & \rho_X(h-2) \\ & & & \vdots \\ & & & \rho_X(1) \\ \hline \rho_X(h-1) & \cdots & \rho_X(1) & 1 \end{array} \right)$$

La fonction d'autocorrélation vérifie la même propriété de semi-définie positivité que la fonction d'autocovariance. La proposition suivante donne en outre une condition équivalente en terme du déterminant des matrices $R_{X,h}$.

Proposition 3.4. *La fonction d'autocorrélation est également une fonction semi-définie positive. D'après le critère de Sylvester, cette propriété est équivalente à la positivité de tous les déterminants des mineurs principaux*

$$\det(R_{X,h}) \geq 0, \text{ pour tout } h$$

Cette propriété fixe une infinité de contraintes sur les corrélations

- $\det(R_{X,2}) \geq 0$ donne $\rho_X^2(1) \leq 1$
- $\det(R_{X,3}) \geq 0$ donne $(1 - \rho_X(2))(1 + \rho_X(2) - 2\rho_X^2(1)) \geq 0$, et donc $1 + \rho_X(2) - 2\rho_X^2(1) \geq 0$,
- et ainsi de suite...

3.4.6 Premiers pas vers les processus ARMA

Les processus ARMA seront au coeur du Chapter 5. Nous les abordons ici très rapidement pour parler de **filtrage linéaire**.

Dans la modélisation ARMA, on utilise un opérateur de série en B de coefficients $(\psi_j)_{j \in \mathbb{Z}}$

$$\psi(B) = \sum_{j \in \mathbb{Z}} \psi_j B^j$$

que l'on applique sur un processus stationnaire. On parle alors de **filtrage linéaire**. Si $(X_t)_{t \in \mathbb{Z}}$ est une série temporelle stationnaire alors

$$Y_t = \psi(B)X_t = \sum_{j \in \mathbb{Z}} \psi_j X_{t-j}$$

On reviendra dans le Chapter 5 sur cette notion de série en B pour mieux la définir et établir des résultats. Nous allons ici nous intéresser à la propriété de maintien de la propriété de stationnarité par le filtrage linéaire sous une condition de sommabilité de ses coefficients.

Proposition 3.5 (Filtrage linéaire). *Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle **stationnaire** de fonction moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$.*

Si la suite de réels $(\psi_j)_{j \in \mathbb{Z}}$ est sommable ($\sum_{j=-\infty}^{+\infty} |\psi_j| < +\infty$) alors la série $(Y_t)_{t \in \mathbb{Z}}$ obtenue par filtrage linéaire de $(X_t)_{t \in \mathbb{Z}}$,

$$Y_t = \psi(B)X_t = \sum_{j=-\infty}^{+\infty} \psi_j X_{t-j},$$

*existe et est également **stationnaire**.*

La moyenne est donnée par la relation

$$\mu_Y = \mu_X \sum_{j=-\infty}^{+\infty} \psi_j$$

et la fonction d'autocovariance vaut

$$\gamma_Y(h) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \psi_i \psi_j \gamma_X(h+i-j), \quad \forall h \in \mathbb{Z}$$

Preuve

- Existence de la série $(Y_t)_{t \in \mathbb{Z}}$:

On a $\sum_{j \in \mathbb{Z}} \|\psi_j X_{t-j}\|_{L^2} = \sum_{j \in \mathbb{Z}} |\psi_j| \|X_{t-j}\|_{L^2}$. Or $\|X_{t-j}\|_{L^2}^2 = \mathbb{E}[X_{t-j}^2] = \text{Var}(X_{t-j}) + \mathbb{E}[X_{t-j}]^2 = \gamma_X(0) + \mu_X^2$. Ainsi $\sum_{j \in \mathbb{Z}} \|\psi_j X_{t-j}\|_{L^2} = \sqrt{\gamma_X(0) + \mu_X^2} \sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$. Donc la série $(Y_t)_{t \in \mathbb{Z}}$ est bien définie.

- Série du second ordre :

$$\|Y_t\|_{L^2} = \left\| \sum_{j \in \mathbb{Z}} \psi_j X_{t-j} \right\|_{L^2} \leq \sum_{j \in \mathbb{Z}} \|\psi_j X_{t-j}\|_{L^2} < +\infty$$

- La fonction moyenne de $(Y_t)_{t \in \mathbb{Z}}$: $\mathbb{E}[Y_t] = \mathbb{E} \left[\sum_{j \in \mathbb{Z}} \psi_j X_{t-j} \right] = \sum_{j \in \mathbb{Z}} \psi_j \mathbb{E}[X_{t-j}] = \mu_X \left(\sum_{j \in \mathbb{Z}} \psi_j \right)$ car $\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$. Donc la fonction moyenne est constante.
- La fonction d'autocovariance de $(Y_t)_{t \in \mathbb{Z}}$:

$$\begin{aligned} \text{Cov}(Y_t, Y_{t+h}) &= \text{Cov} \left(\sum_{i \in \mathbb{Z}} \psi_i X_{t-i}, \sum_{j \in \mathbb{Z}} \psi_j X_{t+h-j} \right) \\ &= \sum_{j \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \psi_i \psi_j \text{Cov}(X_{t-i}, X_{t+h-j}) \\ &= \sum_{j \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \psi_i \psi_j \gamma_X(h+i-j) \\ &= f(h) \end{aligned}$$

car $\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$ (échange $\text{Cov}(\cdot, \cdot)$ et \sum , voir Corollary 3.1).

Ainsi $(Y_t)_{t \in \mathbb{Z}}$ est un processus faiblement stationnaire.

Exemple 3.6 (Filtrage linéaire d'un bruit blanc). Dans cet exemple, on considère que le processus initial est un bruit blanc

$$(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$$

qui est donc stationnaire.

Le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \sum_{j \in \mathbb{Z}} \psi_j \varepsilon_{t-j}$$

avec $\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$ est bien défini et stationnaire. La fonction moyenne de $(Y_t)_{t \in \mathbb{Z}}$ est

$$\mu_Y = \mu_\varepsilon \sum_{j \in \mathbb{Z}} \psi_j = 0$$

et de fonction d'autocovariance

$$\begin{aligned} \gamma_Y(h) &= \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \psi_i \psi_j \gamma_\varepsilon(h + i - j) \\ &= \sigma^2 \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \psi_i \psi_j \mathbb{1}_{h+i-j=0} \\ &= \sigma^2 \sum_{i \in \mathbb{Z}} \psi_i \psi_{h+i} \end{aligned}$$

car $\gamma_\varepsilon(u) = \sigma^2 \mathbb{1}_{u=0}$.

Et pour une somme finie ?

Soit $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \sum_{i=0}^t \varepsilon_{t-i}$$

Le processus $(Y_t)_{t \in \mathbb{Z}}$ existe et il est du second ordre.

Le processus reste centré $\mathbb{E}[Y_t] = \sum_{i=0}^t \mathbb{E}[\varepsilon_{t-i}] = 0$.

Fonction de covariance : soit $h \in \mathbb{N}$,

$$\begin{aligned} \text{Cov}(Y_t, Y_{t+h}) &= \mathbb{E} \left[\sum_{i=0}^t \varepsilon_{t-i} \times \sum_{j=0}^{t+h} \varepsilon_{t+h-j} \right] \\ &= \sum_{i=0}^t \mathbb{E}[\varepsilon_i^2] = \sigma^2 (1 + t). \end{aligned}$$

\Rightarrow **Le processus n'est donc pas stationnaire!** Ainsi, de manière paradoxale, les sommes finies peuvent poser plus de “problèmes” que les sommes infinies !

3.5 Densité spectrale

Jusqu'ici, on a abordé les processus stationnaires du second ordre via leur représentation temporelle. On peut également s'intéresser à leur représentation dans le domaine des fréquences. On aborde alors la notion de **densité spectrale** qui est une fonction contenant la même information que la fonction d'autocovariance.

Définition 3.20. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de fonction d'autocovariance $\gamma_X(\cdot)$. On appelle **densité spectrale**, quand elle existe, la fonction $f_X(\cdot)$ définie sur \mathbb{R} par

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-i\omega h}.$$

On peut reconnaître que cette densité spectrale revient à la transformée de Fourier discrète de la fonction $\gamma_X(\cdot)$ définie sur \mathbb{Z} .

La proposition suivante donne une propriété d'existence de la densité spectrale.

Proposition 3.6 (Existence de la densité spectrale). *La densité spectrale d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ de fonction d'autocovariance $\gamma_X(\cdot)$ existe dès que l'on a :*

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty.$$

Proposition 3.7. *La densité spectrale d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est une fonction réelle, continue, positive, paire et 2π -périodique.*

Preuve

En réécrivant la densité spectrale, on a

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-i\omega h} \\ &= \frac{\gamma_X(0)}{2\pi} + \frac{1}{2\pi} \sum_{h=-\infty}^{-1} \gamma_X(h) e^{-i\omega h} + \frac{1}{2\pi} \sum_{h=1}^{+\infty} \gamma_X(h) e^{-i\omega h} \\ &= \frac{\gamma_X(0)}{2\pi} + \frac{1}{2\pi} \sum_{h=1}^{+\infty} \gamma_X(h) (e^{-i\omega h} + e^{i\omega h}) \\ &= \frac{\gamma_X(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{+\infty} \gamma_X(h) \cos(\omega h) \end{aligned}$$

donc $f_X(\cdot)$ est réelle, continue, paire et 2π -périodique.

Il reste à montrer que $f_X(\cdot)$ est positive. Pour $N \in \mathbb{N}$, on définit

$$\begin{aligned}
f_N(\omega) &= \frac{1}{2\pi N} \mathbb{E} \left[\left| \sum_{j=1}^N (X_j - \mu_X) e^{-i\omega j} \right|^2 \right] \\
&= \frac{1}{2\pi N} \mathbb{E} \left[\sum_{j=1}^N (X_j - \mu_X) e^{-i\omega j} \sum_{k=1}^N (X_k - \mu_X) e^{i\omega k} \right] \\
&= \frac{1}{2\pi N} \sum_{j=1}^N \sum_{k=1}^N e^{-i\omega(j-k)} \mathbb{E} [(X_j - \mu_X)(X_k - \mu_X)] \\
&= \frac{1}{2\pi N} \sum_{j=1}^N \sum_{k=1}^N e^{-i\omega(j-k)} \gamma_X(j-k) \\
&= \frac{1}{2\pi N} \sum_{h; |h| < N} \sum_{j=|h|+1}^N e^{-i\omega h} \gamma_X(h) \\
&= \frac{1}{2\pi N} \sum_{h; |h| < N} (N - |h|) e^{-i\omega h} \gamma_X(h) \\
&= \frac{1}{2\pi} \sum_{h; |h| < N} e^{-i\omega h} \gamma_X(h) - \frac{1}{2\pi} \sum_{h; |h| < N} \frac{|h|}{N} e^{-i\omega h} \gamma_X(h).
\end{aligned}$$

Comme $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$, on a pour le premier terme que

$$\frac{1}{2\pi} \sum_{h; |h| < N} e^{-i\omega h} \gamma_X(h) \xrightarrow{N \rightarrow +\infty} \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-i\omega h} \gamma_X(h) = f_X(\omega).$$

Pour le second terme, on utilise le résultat suivant :

Soit $(a_h)_{h \in \mathbb{Z}}$ une suite de réels positifs telle que $\sum_{h \in \mathbb{Z}} a_h < +\infty$. On a alors

$$\lim_{N \rightarrow +\infty} \sum_{h; |h| < N} \frac{|h|}{N} a_h = 0.$$

Ici, on a

$$0 \leq \left| \frac{1}{2\pi} \sum_{h; |h| < N} \frac{|h|}{N} e^{-i\omega h} \gamma_X(h) \right| \leq \frac{1}{2\pi} \sum_{h; |h| < N} \frac{|h|}{N} |\gamma_X(h)| \xrightarrow{N \rightarrow +\infty} 0.$$

Au final, on a que $f_N(\omega) \xrightarrow{N \rightarrow +\infty} f_X(\omega)$, $\forall \omega \in \mathbb{R}$ et $\forall \omega \in \mathbb{R}$, $f_N(\omega) \geq 0$ ce qui conclut à la positivité de $f_X(\cdot)$.

Theorem 3.4. *Il est équivalent de connaître la fonction d'autocovariance ou la densité spec-*

trale, quand elle existe, d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$:

$$\gamma_X(h) = \int_{-\pi}^{\pi} f_X(\omega) \cos(\omega h) d\omega = \int_{-\pi}^{\pi} f_X(\omega) e^{i\omega h} d\omega.$$

Preuve

On peut écrire :

$$\begin{aligned} \int_{-\pi}^{\pi} f_X(\omega) \cos(\omega h) d\omega &= \int_{-\pi}^{\pi} f_X(\omega) \left(\frac{e^{i\omega h} + e^{-i\omega h}}{2} \right) d\omega \\ &= \frac{1}{2} \int_{-\pi}^{\pi} f_X(\omega) e^{i\omega h} d\omega + \frac{1}{2} \int_{-\pi}^{\pi} f_X(\omega) e^{-i\omega h} d\omega \\ &= \int_{-\pi}^{\pi} f_X(\omega) e^{i\omega h} d\omega \quad \text{par parité de la densité spectrale} \\ &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) e^{-i\omega k} e^{i\omega h} d\omega \\ &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) \int_{-\pi}^{\pi} e^{i\omega(h-k)} d\omega \end{aligned}$$

par la convergence normale de la série. Or

$$\int_{-\pi}^{\pi} e^{i\omega(h-k)} d\omega = \begin{cases} 2\pi & \text{si } h = k \\ \left[\frac{e^{i\omega(h-k)}}{h-k} \right]_{-\pi}^{\pi} = 0 & \text{si } k \neq h \end{cases}$$

Ainsi,

$$\int_{-\pi}^{\pi} f_X(\omega) \cos(\omega h) d\omega = \gamma_X(h).$$

Exemple 3.7 (Bruit blanc faible). Soit un bruit blanc faible $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$. Rappelons que sa fonction d'autocovariance vaut $\gamma_{\varepsilon}(h) = \sigma^2 \mathbb{1}_{h=0}$. Sa densité spectrale vaut donc

$$f_{\varepsilon}(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_{\varepsilon}(h) e^{-i\omega h} = \frac{\sigma^2}{2\pi}, \quad \forall \omega \in \mathbb{R}.$$

Et réciproquement tout processus stationnaire de densité spectrale constante est un bruit blanc faible.

La proposition suivante concerne la densité spectrale d'un filtrage linéaire.

Proposition 3.8. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de densité spectrale $f_X(\cdot)$. Soit $(Y_t)_{t \in \mathbb{Z}}$ un filtrage linéaire de la série $(X_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \psi(B)X_t = \sum_{j=-\infty}^{+\infty} \psi_j X_{t-j} \text{ avec } \sum_{j=-\infty}^{+\infty} |\psi_j| < +\infty.$$

Le processus $(Y_t)_{t \in \mathbb{Z}}$ est donc stationnaire de densité spectrale

$$\forall \omega \in \mathbb{R}, f_Y(\omega) = f_X(\omega) \left| \sum_{j=-\infty}^{+\infty} \psi_j e^{-i\omega j} \right|^2.$$

Preuve

- Existence de la densité spectrale de $(Y_t)_{t \in \mathbb{Z}}$:

Comme $\sum_{j=-\infty}^{+\infty} |\psi_j| < +\infty$, d'après la Proposition 3.5, on a

$$\gamma_Y(h) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_j \psi_k \gamma_X(h + j - k).$$

Donc

$$\begin{aligned} \sum_{h \in \mathbb{Z}} |\gamma_Y(h)| &= \sum_{h \in \mathbb{Z}} \left| \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_j \psi_k \gamma_X(h + j - k) \right| \\ &\leq \sum_{h \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |\psi_j| |\psi_k| |\gamma_X(h + j - k)| \\ &= \sum_{j \in \mathbb{Z}} |\psi_j| \cdot \sum_{k \in \mathbb{Z}} |\psi_k| \cdot \sum_{h \in \mathbb{Z}} |\gamma_X(h + j - k)|, \end{aligned}$$

où $\sum_{h \in \mathbb{Z}} |\gamma_X(h + j - k)| < +\infty$ par hypothèse d'existence de la densité spectrale du processus $(X_t)_{t \in \mathbb{Z}}$ et $\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$ par hypothèse. Ainsi la densité spectrale du processus $(Y_t)_{t \in \mathbb{Z}}$ existe d'après la Proposition 3.6.

- Calcul de $f_Y(\cdot)$: Maintenant, on peut écrire

$$\begin{aligned}
f_Y(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_Y(h) e^{-i\omega h} \\
&= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_j \psi_k \gamma_X(h + j - k) e^{-i\omega h} \\
&= \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} \psi_j e^{i\omega j} \sum_{k \in \mathbb{Z}} \psi_k e^{-i\omega k} \sum_{h \in \mathbb{Z}} \gamma_X(h + j - k) e^{-i\omega(h + j - k)} \\
&= \frac{1}{2\pi} \left(\sum_{j \in \mathbb{Z}} \psi_j e^{i\omega j} \right) \left(\sum_{k \in \mathbb{Z}} \psi_k e^{-i\omega k} \right) (2\pi f_X(\omega)) \\
&= f_X(\omega) \left| \sum_{j \in \mathbb{Z}} \psi_j e^{-i\omega j} \right|^2.
\end{aligned}$$

Example 3.8 (Densité spectrale du filtrage linéaire d'un bruit blanc). Soit $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et $(Y_t)_{t \in \mathbb{Z}}$ le filtrage linéaire de ce bruit blanc défini par

$$Y_t = \sum_{j \in \mathbb{Z}} \psi_j \varepsilon_{t-j}$$

avec $\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$. D'après la Proposition 3.8, la densité spectrale de ce processus est alors égale à :

$$f_Y(\omega) = f_\varepsilon(\omega) \cdot \left| \sum_{j=-\infty}^{+\infty} \psi_j e^{-i\omega j} \right|^2 = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{+\infty} \psi_j e^{-i\omega j} \right|^2.$$

4 Statistique des processus stationnaires du second ordre

Dans ce chapitre, nous abordons quelques points d'inférence statistique dans le cadre de l'étude des séries temporelles. Nous allons nous intéresser à la construction d'estimateurs pour la moyenne, la fonction d'autocovariance, la fonction d'autocorrélation et la densité spectrale d'un processus stationnaire. Nous aborderons ensuite la problématique de la prévision des valeurs futures d'une série temporelle. Enfin, nous verrons quelques tests statistiques qui permettent de se prononcer sur la stationnarité d'un processus.

Dans tout ce chapitre, nous supposons que l'on observe le processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ sur les instants $t = 1, \dots, n$.

4.1 Quelques estimateurs

4.1.1 Estimateur de la moyenne du processus stationnaire

Rappelons que la fonction moyenne μ_X est constante pour un processus stationnaire. Ainsi, on l'estime facilement par la moyenne empirique

$$\hat{\mu}_X = \bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t.$$

Cet estimateur est sans biais et L^2 -consistant d'après le théorème suivant.

Theorem 4.1. *Si $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$ alors :*

- si $\gamma_X(h) \xrightarrow{h \rightarrow +\infty} 0$ alors $\text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} 0$
- si de plus $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$ alors

$$n \text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} \sum_{h \in \mathbb{Z}} \gamma_X(h) = 2\pi f_X(0)$$

Preuve

On peut écrire

$$\begin{aligned}
n\text{Var}(\bar{X}_n) &= \text{Cov}\left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{j=1}^n X_j\right) \\
&= \frac{1}{n}\sum_{i,j=1}^n \text{Cov}(X_i, X_j) \\
&= \frac{1}{n}\sum_{i,j=1}^n \gamma_X(i-j) \\
&= \frac{1}{n}\sum_{h:|h|<n} (n-|h|)\gamma_X(h) \\
&= \sum_{h:|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \\
&\leq \sum_{h:|h|<n} |\gamma_X(h)| = |\gamma_X(0)| + 2\sum_{h=1}^n |\gamma_X(h)|.
\end{aligned}$$

Par le théorème de Césaro, on sait que l'on a la convergence de la moyenne de Césaro

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{h=1}^n |\gamma_X(h)| = \lim_{h \rightarrow +\infty} |\gamma_X(h)|,$$

dès lors que cette dernière limite existe. Donc comme par hypothèse $\gamma_X(h) \xrightarrow{h \rightarrow +\infty} 0$, on obtient que $\text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} 0$.

Maintenant, sous l'hypothèse

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty,$$

on peut appliquer le théorème de convergence dominée pour inverser limite et somme et obtenir

$$\begin{aligned}
\lim_{n \rightarrow +\infty} n\text{Var}(\bar{X}_n) &= \lim_{n \rightarrow +\infty} \sum_{h:|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \\
&= \lim_{n \rightarrow +\infty} \sum_{h \in \mathbb{Z}} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \mathbb{1}_{|h|<n} \\
&= \sum_{h \in \mathbb{Z}} \gamma_X(h),
\end{aligned}$$

car $g_n(h) := \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \mathbb{1}_{|h|<n} \xrightarrow{n \rightarrow +\infty} \gamma_X(h)$.

Enfin, par définition de la densité spectrale, $\sum_{h \in \mathbb{Z}} \gamma_X(h) = 2\pi f_X(0)$.

Le premier résultat prouve la convergence dans L^2 de l'estimateur vers la moyenne μ_X du processus. Le second donne la variance asymptotique de l'estimateur normalisé.

Sous des hypothèses supplémentaires, on peut obtenir le comportement asymptotique gaussien de \bar{X}_n .

Proposition 4.1. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire défini, pour tout $t \in \mathbb{Z}$, par*

$$X_t = \mu_X + \sum_{j \in \mathbb{Z}} \psi_j \varepsilon_{t-j}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un $IID(0, \sigma^2)$ tel que $\mathbb{E}[\varepsilon_t^4] < +\infty$ et où la suite des coefficients $(\psi_j)_{j \in \mathbb{Z}}$ est sommable et de somme non nulle

$$\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty \text{ et } \sum_{j \in \mathbb{Z}} \psi_j \neq 0.$$

On a alors

$$\sqrt{n}(\bar{X}_n - \mu_X) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sum_{h \in \mathbb{Z}} \gamma_X(h)\right).$$

4.1.2 Estimateur de la fonction d'auto-covariance / -corrélation

On rappelle que la fonction d'autocovariance est définie par

$$\forall h \in \mathbb{Z}, \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mathbb{E}[X_t])(X_{t+h} - \mathbb{E}[X_{t+h}])].$$

Donc à partir de X_1, \dots, X_n , on peut estimer, pour les valeurs de h telles que $|h| < n - 1$, l'autocovariance $\gamma_X(h)$ par

$$\begin{cases} \hat{\gamma}_{X,n}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n) & \text{si } h \geq 0 \\ \hat{\gamma}_{X,n}(h) = \hat{\gamma}_{X,n}(-h) & \text{si } h < 0 \end{cases}$$

Proposition 4.2. *L'estimateur $\hat{\gamma}_{X,n}(h)$ est asymptotiquement sans biais et consistant pour $\gamma_X(h)$.*

Remarque

En pratique, on utilise cet estimateur pour $h \leq \frac{n}{4}$.

A partir de l'estimateur de la fonction covariance, on en déduit l'estimateur suivant pour la fonction d'autocorrélation : Pour tout $|h| < n - 1$,

$$\hat{\rho}_{X,n}(h) = \frac{\hat{\gamma}_{X,n}(h)}{\hat{\gamma}_{X,n}(0)} = \frac{\sum_{t=1}^{n-|h|} (X_t - \bar{X}_n) (X_{t+|h|} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2}.$$

Proposition 4.3. *L'estimateur $\hat{\rho}_{X,n}(h)$ est consistant.*

En pratique on estime $\rho_X(h)$ pour $h \leq \frac{n}{4}$.

On a de plus le comportement asymptotique suivant :

Proposition 4.4. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire défini par $X_t = \mu_X + \sum_{j \in \mathbb{Z}} \psi_j \varepsilon_{t-j}$ où $(\varepsilon_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2)$ tel que $\mathbb{E}[\varepsilon_t^4] < +\infty$ et où la suite des coefficients $(\psi_j)_{j \in \mathbb{Z}}$ est sommable ($\sum_{j \in \mathbb{Z}} |\psi_j| < +\infty$) et de somme non nulle ($\sum_{j \in \mathbb{Z}} \psi_j \neq 0$).*

On a alors, pour tout $k > 0$ fixé,

$$\sqrt{n} \begin{pmatrix} \hat{\rho}_{X,n}(1) - \rho_X(1) \\ \vdots \\ \hat{\rho}_{X,n}(k) - \rho_X(k) \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^{[k]}),$$

où $\Sigma^{[k]} = (\Sigma_{ij}^{[k]})_{1 \leq i, j \leq k}$ est la matrice de covariance asymptotique déterminée par

$$\Sigma_{ij}^{[k]} = \sum_{h \in \mathbb{Z}} \{[\rho_X(h+i) + \rho_X(h-i) - 2\rho_X(i)\rho_X(h)] \times [\rho_X(h+j) + \rho_X(h-j) - 2\rho_X(j)\rho_X(h)]\}$$

(formule de Bartlett).

4.1.3 Estimateur de la matrice d'auto-covariance / -corrélation

On s'intéresse maintenant à l'estimation des matrices d'autocovariance et d'autocorrélation.

Définition 4.1. La matrice

$$\hat{\Gamma}_{X,n} := \begin{pmatrix} \hat{\gamma}_{X,n}(0) & \hat{\gamma}_{X,n}(1) & \dots & \hat{\gamma}_{X,n}(n-1) \\ \hat{\gamma}_{X,n}(1) & \hat{\gamma}_{X,n}(0) & \dots & \hat{\gamma}_{X,n}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{X,n}(n-1) & \hat{\gamma}_{X,n}(n-2) & \dots & \hat{\gamma}_{X,n}(0) \end{pmatrix}$$

est un estimateur de la matrice d'autocovariance.

La matrice $\hat{R}_{X,n} = \frac{\hat{\Gamma}_{X,n}}{\hat{\gamma}_{X,n}(0)}$ est un estimateur de la matrice d'autocorrélation $R_{X,n}$.

Proposition 4.5. Les matrices $\hat{\Gamma}_{X,n}$ et $\hat{R}_{X,n}$ sont des matrices semi-définies positives.

Preuve

On commence par remarquer que l'on peut écrire $\hat{\Gamma}_{X,n} = \frac{1}{n}TT'$ avec

$$T = \begin{pmatrix} 0 & \dots & \dots & \dots & \dots & 0 & Y_1 & Y_2 & 0 & \dots & 0 & Y_n \\ 0 & \dots & \dots & \dots & 0 & Y_1 & Y_2 & 0 & \dots & \dots & Y_n & 0 \\ \vdots & & & & & & \vdots & & & & & \vdots \\ 0 & Y_1 & Y_2 & 0 & \dots & 0 & Y_n & 0 & \dots & \dots & \dots & 0 \end{pmatrix} \in \mathcal{M}_{n \times 2n}(\mathbb{R})$$

et $Y_t = X_t - \bar{X}_n$, $t \in \{1, \dots, n\}$. Alors $\forall a \in \mathbb{R}^n$, $a' \hat{\Gamma}_{X,n} a = \frac{1}{n}(a'T)(a'T)' \geq 0$.

4.1.4 Estimateur de la densité spectrale

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$ telle que $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$

Sous ces hypothèses, on a vu (voir Definition 3.20 et Proposition 3.6) que la densité spectrale existe et est définie par

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-ih\omega}, \quad \forall \omega \in \mathbb{R}$$

Pour rappel, $f_X(\cdot)$ est une fonction paire, 2π -périodique, continue, positive.

Pour construire un estimateur de la densité spectrale, on commence par définir le périodogramme.

Definition 4.2. Le **périodogramme** associé à (X_1, \dots, X_n) est défini par

$$I_n(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{-it\omega_j} \right|^2, \quad \forall \omega_j \in \Omega_n := \left\{ \omega_j = \frac{2\pi j}{n}; \omega_j \in]-\pi, \pi] \right\}$$

Nous allons maintenant relier le périodogramme et l'estimateur $\hat{\gamma}_{X,n}$.

Proposition 4.6.

$$\begin{cases} I_n(0) = n|\bar{X}_n|^2 \\ I_n(\omega_j) = \sum_{|h| < n} \hat{\gamma}_{X,n}(h) e^{-ih\omega_j} \quad \text{si } \omega_j \in \Omega_n, \omega_j \neq 0 \end{cases}$$

Preuve

- Pour $\omega_j = 0$: $I_n(0) = \frac{1}{n} \left| \sum_{t=1}^n X_t \right|^2 = n |\bar{X}_n|^2$
- Soit $\omega_j \neq 0$. On a

$$\begin{aligned}
 I_n(\omega_j) &= \frac{1}{n} \left| \sum_{t=1}^n X_t e^{-it\omega_j} \right|^2 \\
 &= \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n X_s X_t e^{-it\omega_j} e^{is\omega_j} \\
 &= \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n (X_s - \bar{X}_n)(X_t - \bar{X}_n) e^{-i(t-s)\omega_j} \\
 &\quad + \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n \bar{X}_n X_t e^{-i(t-s)\omega_j} \\
 &\quad + \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n X_s \bar{X}_n e^{-i(t-s)\omega_j} \\
 &\quad - \frac{1}{n} (\bar{X}_n)^2 \sum_{s=1}^n \sum_{t=1}^n e^{-i(t-s)\omega_j}.
 \end{aligned}$$

Or $\sum_{s=1}^n e^{is\omega_j} = \sum_{t=1}^n e^{-it\omega_j} = 0$ si $\omega_j \neq 0$, donc les trois dernières sommes sont nulles. Ainsi

$$\begin{aligned}
 I_n(\omega_j) &= \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n (X_s - \bar{X}_n)(X_t - \bar{X}_n) e^{-i(t-s)\omega_j} \\
 &= \sum_{|h| < n} \frac{1}{n} \sum_{s=1}^{n-|h|} (X_s - \bar{X}_n)(X_{s+h} - \bar{X}_n) e^{-ih\omega_j} \\
 &= \sum_{|h| < n} \hat{\gamma}_{X,n}(h) e^{-ih\omega_j}
 \end{aligned}$$

On peut alors proposer naturellement d'estimer

- $f_X(\omega_j) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-ih\omega_j}$ par $\hat{f}_X(\omega_j) := \frac{I_n(\omega_j)}{2\pi}$ pour $\omega_j \neq 0$.
- $f_X(0) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h)$ par $\hat{f}_X(0) = \frac{1}{2\pi} \sum_{|h| < n} \hat{\gamma}_{X,n}(0)$.

On cherche ensuite à étendre \hat{f}_X à tout l'intervalle $[-\pi, \pi]$ pour estimer $f_X(\cdot)$ (qui est paire et

2π -périodique). On peut le faire en obtenant un estimateur $\hat{f}_X(\cdot)$ contant par morceaux

$$\hat{f}_X(\omega) = \begin{cases} \hat{f}_X(\omega_j) & \text{si } \omega_j - \frac{\pi}{n} < \omega \leq \omega_j + \frac{\pi}{n}, \omega \in [0, \pi] \\ \hat{f}_X(-\omega) & \text{si } \omega \in [-\pi, 0[\end{cases}$$

Proposition 4.7. *Si $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$ telle que $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$, alors*

- $\mathbb{E}[I_n(0)] - n\mu_X^2 \xrightarrow{n \rightarrow +\infty} 2\pi f_X(0)$
- $\mathbb{E}[\hat{f}_X(\omega)] \xrightarrow{n \rightarrow +\infty} f_X(\omega)$ si $\omega \neq 0$

Preuve

- $\mathbb{E}[I_n(0)] - n\mu_X^2 = \mathbb{E}[n\bar{X}_n^2] - n\mu_X^2 = n\text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} \sum_{h \in \mathbb{Z}} \gamma_X(h) = 2\pi f(0)$ d'après Theorem 4.1.
- Soit $\omega \in]0, \pi]$. Soit $g(n, \omega)$ le multiple de $\frac{2\pi}{n}$ le plus proche de ω . Pour n assez

grand, on a que $g(n, \omega) \neq 0$. Alors

$$\begin{aligned}
2\pi \mathbb{E} [\hat{f}_X(\omega)] &= \mathbb{E} [I_n(\omega)] \\
&= \mathbb{E} [I_n(g(n, \omega))] \\
&= \mathbb{E} \left[\sum_{|h| < n} \hat{\gamma}_{X,n}(h) e^{-ihg(n, \omega)} \right] \\
&= \mathbb{E} \left[\sum_{|h| < n} \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n) e^{-ihg(n, \omega)} \right] \\
&= \mathbb{E} \left[\sum_{|h| < n} \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \mu)(X_{t+h} - \mu) e^{-ihg(n, \omega)} \right] \\
&= \sum_{|h| < n} \mathbb{E} [(X_t - \mu)(X_{t+h} - \mu)] e^{-ihg(n, \omega)} \\
&= \sum_{|h| < n} \frac{1}{n} \sum_{t=1}^{n-|h|} \gamma_X(h) e^{-ihg(n, \omega)} \\
&= \sum_{|h| < n} \frac{n-|h|}{n} \gamma_X(h) e^{-ihg(n, \omega)} \\
&= \sum_{|h| < n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) e^{-ihg(n, \omega)}.
\end{aligned}$$

Comme $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$, par convergence dominée

$$\sum_{h \in \mathbb{Z}} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) e^{-ih\lambda} 1_{|h| < n} \xrightarrow{n \rightarrow +\infty} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-ih\lambda} = 2\pi f(\lambda),$$

et $g(n, \omega) \xrightarrow{n \rightarrow +\infty} \omega$. Donc

$$\mathbb{E} [I_n(\omega)] \xrightarrow{n \rightarrow +\infty} 2\pi f(\omega).$$

Une autre stratégie est d'utiliser un noyau sur les $I_n(\omega_j)$ pour construire un estimateur lissé.

Soit $(m_n)_n$ une suite d'entiers positifs tels que $m_n \xrightarrow{n \rightarrow +\infty} +\infty$ et $\frac{m_n}{n} \xrightarrow{n \rightarrow +\infty} 0$. Soit un noyau $W_n(\cdot)$ tel que

- $W_n(-j) = W_n(j) \geq 0, \forall j$
- $\sum_{|j| \leq m_n} W_n(j) = 1$
- $\sum_{|j| \leq m_n} W_n(j)^2 \xrightarrow{n \rightarrow +\infty} 0$

Alors on définit l'estimateur

$$\tilde{f}_X(\omega) = \frac{1}{2\pi} \sum_{|j| \leq m_n} W_n(j) I_n \left(g(n, \omega) + \frac{2\pi j}{n} \right) \text{ où } g(n, \omega) = \operatorname{argmin}_{\omega_k \in \Omega_n} |\omega_k - \omega|.$$

Proposition 4.8. *Par les hypothèses sur la suite $(m_n)_n$ et le noyau W_n , l'estimateur lissé $\tilde{f}_X(\cdot)$ est asymptotiquement sans biais.*

$$\forall \omega, \mathbb{E} [\tilde{f}_X(\omega)] \xrightarrow{n \rightarrow +\infty} f_X(\omega).$$

Preuve

$$\begin{aligned} |\mathbb{E}[\tilde{f}_X(\omega) - f_X(\omega)]| &= \left| \sum_{|j| \leq m_n} \frac{W_n(j)}{2\pi} \mathbb{E} \left[I_n \left(g(n, \omega) + \frac{2\pi j}{n} \right) \right] - f_X(\omega) \right| \\ &= \left| \sum_{|j| \leq m_n} W_n(j) \left\{ \frac{\mathbb{E} [I_n (g(n, \omega) + \frac{2\pi j}{n})]}{2\pi} - f_X \left(g(n, \omega) + \frac{2\pi}{j} \right) + f_X \left(g(n, \omega) + \frac{2\pi}{j} \right) - f_X(\omega) \right\} \right| \end{aligned}$$

car $\sum_{|j| \leq m_n} W_n(j) = 1$.

L'hypothèse sur $(m_n)_n$ implique que

$$\max_{|j| \leq m_n} \left| g(n, \omega) + \frac{2\pi j}{n} - \omega \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Pour tout $\varepsilon > 0$, avec la continuité de $f_X(\cdot)$, pour n assez grand

$$\max_{|j| \leq m_n} \left| f_X \left(g(n, \omega) + \frac{2\pi j}{n} \right) - f_X(\omega) \right| \leq \frac{\varepsilon}{2}.$$

De plus, $\mathbb{E}[I_n(\omega)] \xrightarrow{n \rightarrow +\infty} 2\pi f_X(\omega)$, $\forall \omega \neq 0$ donc pour n assez grand

$$\max_{|j| \leq m_n} \left| \frac{\mathbb{E} [I_n (g(n, \omega) + \frac{2\pi j}{n})]}{2\pi} - f_X \left(g(n, \omega) + \frac{2\pi}{j} \right) \right| \leq \frac{\varepsilon}{2}.$$

Ainsi, pour n assez grand,

$$\begin{aligned} |\mathbb{E}[\tilde{f}_X(\omega) - f_X(\omega)]| &\leq \sum_{|j| \leq m_n} W_n(j) \left\{ \left| \frac{\mathbb{E} [I_n (g(n, \omega) + \frac{2\pi j}{n})]}{2\pi} - f_X \left(g(n, \omega) + \frac{2\pi}{j} \right) \right| + \left| f_X \left(g(n, \omega) + \frac{2\pi}{j} \right) - f_X(\omega) \right| \right\} \\ &\leq \sum_{|j| \leq m_n} W_n(j) \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \right) \leq \varepsilon. \end{aligned}$$

Donc

$$\mathbb{E} [\tilde{f}_X(\omega)] \xrightarrow{n \rightarrow +\infty} f_X(\omega).$$

Exemple 4.1. Soit le processus $(X_t)_{t \in \mathbb{Z}}$ défini par $X_t = \varepsilon_t + 0.7\varepsilon_{t-1}$, $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$.

D'après Exemple 3.4, la fonction d'autocovariance de ce processus MA(1) vaut

$$\gamma_X(h) = \begin{cases} (1 + 0.7^2) & \text{si } h = 0 \\ 0.7 & \text{si } |h| = 1 \\ 0 & \text{si } |h| > 1 \end{cases}$$

La densité spectrale de ce processus vaut alors

$$f_X(\omega) = \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-i\omega h} = 1 + (0.7)^2 + 2 \times 0.7 \times \cos(\omega), \quad \forall \omega \in \mathbb{R}$$

Sur la Figure 4.1, la densité spectrale du processus est représentée en rouge et deux estimateurs par lissage (obtenus pour deux paramètres différents de lissage de la fonction `smooth.periodogram`) en bleu et en magenta. Les points correspondent aux valeurs de $2\pi I_n(\omega_j)$.

4.2 Prédiction linéaire optimale

Un objectif important dans l'étude des séries temporelles est de pouvoir **prédire les valeurs non encore observées** de la série et cela à des horizons plus ou moins éloignés. Nous allons dans cette section introduire la méthode de prédiction la plus couramment utilisée appelée **la prédiction linéaire optimale**. Cette méthode s'appuie sur la propriété d'espace de Hilbert de $L^2(\Omega, \mathcal{A}, P)$.

4.2.1 Espaces linéaires engendrés par un processus du second ordre

Définition 4.3. Soit $(X_t)_{t \in \mathbb{Z}}$, un processus du second ordre. On appelle **espace vectoriel fermé engendré** par une famille $(X_t)_{t \in I}$, où $I \subset \mathbb{Z}$, le plus petit sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{A}, P)$ qui contient tous les X_t pour $t \in I$. On le note $\overline{\text{Vect}}(X_t, t \in I)$.

Proposition 4.9. *Le sous-espace vectoriel fermé engendré par une famille **finie** $(X_i)_{i \in I}$, où $I \subset \mathbb{Z}$, est l'ensemble de toutes les combinaisons linéaires, i.e. l'ensemble des v.a. Y de la forme $Y = \sum_{i \in I} \alpha_i X_i$.*

Définition 4.4. Soit l'espace vectoriel fermé engendré par une famille **finie** $(X_i)_{i \in I}$ de v.a. de $L^2(\Omega, \mathcal{A}, P)$, où $I \subset \mathbb{Z}$ **fini** : $\mathcal{H} = \overline{\text{Vect}}(X_i, i \in I)$.

La **projection orthogonale** d'une v.a. X sur \mathcal{H} est l'unique élément

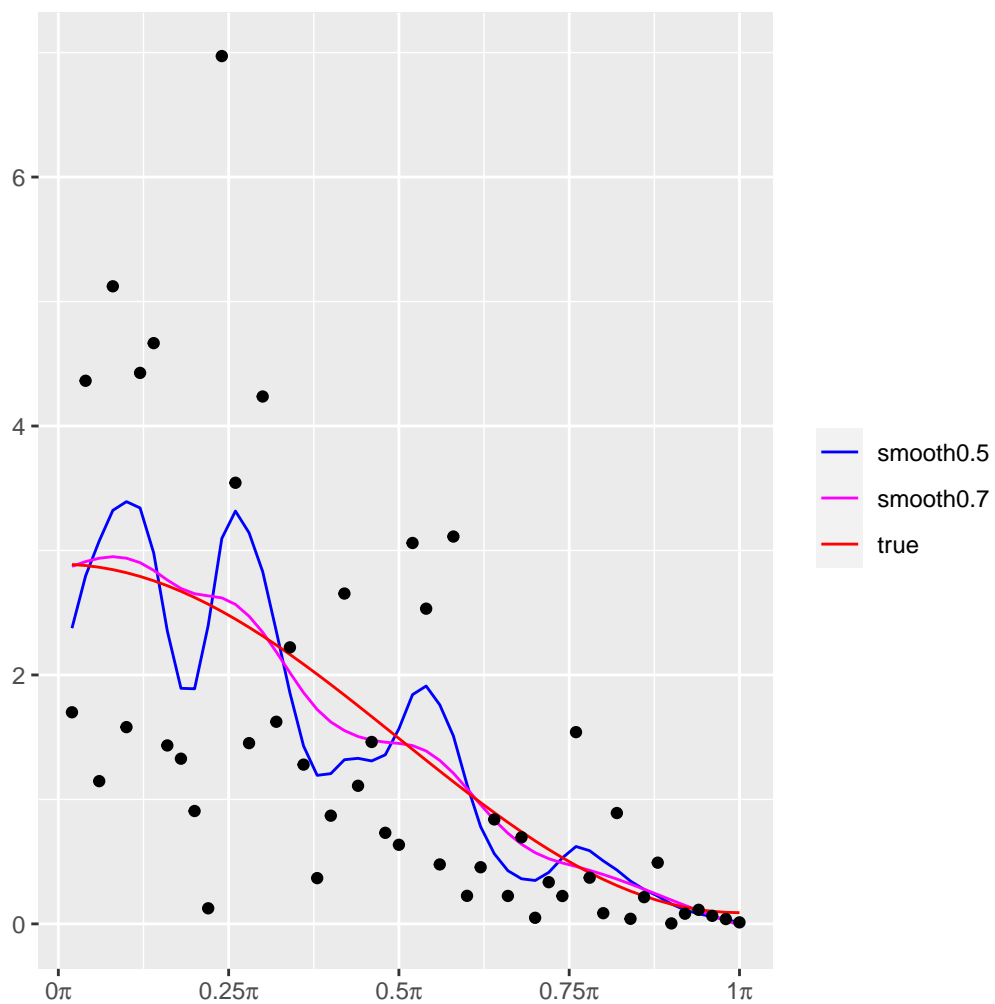


Figure 4.1: Illustration de l'estimation de la densité spectrale d'un processus MA(1)

$$\hat{X} = P_{\mathcal{H}} X = \sum_{i \in I} \alpha_i X_i \text{ tel que } \langle X - \hat{X}, Z \rangle_{L^2} = 0, \quad \forall Z \in \mathcal{H}.$$

On a que $\langle \hat{X}, X_i \rangle_{L^2} = \langle X, X_i \rangle_{L^2}$, pour tout $i \in I$.

Dans la suite, on note

- $\mathcal{H}_1^n = \overline{\text{Vect}}(1, X_1, \dots, X_n)$
- $\mathcal{H}_{-\infty}^n = \overline{\text{Vect}}(1, X_t, t \leq n)$.

4.2.2 Régression linéaire

Pour prédire X_{n+1} (ou X_{n+h}) à partir de l'observation des X_1, \dots, X_n , on pourrait s'appuyer sur la projection dans L^2 de X_{n+1} (ou X_{n+h}) sur le sous-espace vectoriel fermé des fonctions $\sigma(X_1, \dots, X_n)$ -mesurables

$$\begin{aligned} \hat{X} &= \mathbb{E}[X | \sigma(X_1, \dots, X_n)] \\ &= P_{\mathcal{M}(X_1, \dots, X_n)}(X) \\ &= \arg \inf_{Y \in \mathcal{M}(X_1, \dots, X_n)} \|X - Y\|_{L^2} \end{aligned}$$

où $\mathcal{M}(X_1, \dots, X_n) = \{g(X_1, \dots, X_n); g \text{ fonction borélienne de } \mathbb{R}^n \text{ vers } \mathbb{R}\}$ et $\sigma(X_1, \dots, X_n)$ est la tribu engendrée par les v.a. X_1, \dots, X_n . Mais une telle espérance conditionnelle est souvent difficilement calculable. Aussi l'idée est de se restreindre à un espace plus simple, inclus dans $\mathcal{M}(X_1, \dots, X_n)$ pour lequel l'espérance conditionnelle est accessible. On va donc ici se restreindre à projeter sur l'espace vectoriel fermé $\mathcal{H}_1^n := \overline{\text{Vect}}(1, X_1, \dots, X_n) \subset \mathcal{M}(X_1, \dots, X_n)$. Ainsi on cherche une v.a. \hat{X} comme une combinaison linéaire des v.a. $1, X_1, \dots, X_n$ plutôt qu'une fonction mesurable quelconque de ces variables.

Définition 4.5. On appelle **régression linéaire** d'une v.a. Y de $L^2(\Omega, \mathcal{A}, P)$ sur $\mathcal{H}_1^n = \overline{\text{Vect}}(1, X_1, \dots, X_n)$, la projection orthogonale, au sens de la norme L^2 , de Y sur cet espace. On la note $EL(Y | \mathcal{H}_1^n)$.

Proposition 4.10 (Caractérisation). *Soit $Y \in L^2(\Omega, \mathcal{A}, P)$. La régression linéaire $\hat{Y} = EL(Y | \mathcal{H}_1^n)$ est la v.a.*

$$\hat{Y} = \alpha_0 + \sum_{t=1}^n \alpha_t X_t$$

telle que $\mathbb{E}[\hat{Y}] = \mathbb{E}[Y]$ et $\mathbb{E}[\hat{Y} X_t] = \mathbb{E}[Y X_t]$ pour $t = 1, \dots, n$.

Preuve

Pour la preuve, il suffit de traduire que $Y - \hat{Y}$ est orthogonal à $1, X_1, \dots, X_n$ donc

$$\begin{aligned}\langle Y - \hat{Y}, 1 \rangle_{L^2} &= \mathbb{E}[(Y - \hat{Y})1] = 0 \Leftrightarrow \mathbb{E}[Y] = \mathbb{E}[\hat{Y}] \\ \langle Y - \hat{Y}, X_t \rangle_{L^2} &= \mathbb{E}[(Y - \hat{Y})X_t] = 0 \Leftrightarrow \mathbb{E}[YX_t] = \mathbb{E}[\hat{Y}X_t] \text{ pour } t = 1, \dots, n\end{aligned}$$

4.2.3 Prévision linéaire optimale

Definition 4.6. Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle stationnaire. La **prévision linéaire optimale** de X_{n+1} sachant son passé observé est

- $\hat{X}_{n+1} = \text{EL}(X_{n+1} | \mathcal{H}_1^n)$ dans le cas d'un passé fini
- $\hat{X}_{n+1} = \text{EL}(X_{n+1} | \mathcal{H}_{-\infty}^n)$ dans le cas d'un passé infini

Definition 4.7. Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle stationnaire et, pour tout $t \in \mathbb{Z}$, la prévision linéaire optimale $\hat{X}_t = \text{EL}(X_t | \mathcal{H}_{-\infty}^{t-1})$ de X_t sachant le passé (infini) du processus.

On appelle **processus des innovations** le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ des erreurs de prévision successives

$$\varepsilon_t = X_t - \hat{X}_t, \quad \forall t \in \mathbb{Z}.$$

Proposition 4.11. *Le processus des innovations $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.*

Preuve

- Par définition de $t - 1_{-\infty}$ et de \hat{X}_t , on a $\langle X_t - \hat{X}_t, 1 \rangle_{L^2} = 0$ donc $\mathbb{E}[X_t] = \mathbb{E}[\hat{X}_t]$, ce qui prouve que le processus des innovations est centré.

- Soit $h \in \mathbb{N}^*$. On a aussi

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \mathbb{E}[\varepsilon_t \varepsilon_{t+h}] = \langle \varepsilon_t, \varepsilon_{t+h} \rangle.$$

Or $\varepsilon_{t+h} \perp \mathcal{H}_{-\infty}^{t+h-1}$ et $\varepsilon_t = X_t - \text{EL}(X_t | \mathcal{H}_{-\infty}^{t-1}) \subset \mathcal{H}_{-\infty}^{t+h-1}$. Donc $\text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \langle \varepsilon_t, \varepsilon_{t+h} \rangle = 0$.

- Montrons que la variance du processus des innovations est constante :

Tout élément de $t - 1_{-\infty}$ s'écrit sous la forme $\sum_{i=1}^{+\infty} \alpha_i X_{t-i}$ donc $\varepsilon_t = X_t - \sum_{i=1}^{+\infty} \alpha_i X_{t-i}$.

Ainsi,

$$\begin{aligned}
\text{Var}(\varepsilon_t) &= \text{Var}(X_t - \text{EL}(X_t | t-1, \dots, -\infty)) \\
&= \text{Var}\left(X_t - \sum_{i=1}^{+\infty} \alpha_i X_{t-i}\right) \\
&= \text{Var}\left(-\sum_{i=0}^{+\infty} \alpha_i X_{t-i}\right) \quad \text{avec } \alpha_0 = -1 \\
&= \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \alpha_i \alpha_j \text{Cov}(X_{t-i}, X_{t-j}) \\
&= \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \alpha_i \alpha_j \gamma_X(i-j)
\end{aligned}$$

donc la variance est constante.

4.2.4 Prédiction linéaire optimale dans le cas d'un passé fini

On se place ici dans le cas particulier où l'on observe qu'un passé fini X_1, \dots, X_n et on souhaite prédire des valeurs futures.

Proposition 4.12. Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle stationnaire de moyenne μ_X et d'ACVF $\gamma_X(\cdot)$. La **prédiction linéaire optimale** de X_{n+h} , pour $h \in \mathbb{N}^*$, ayant observé le passé X_1, \dots, X_n , est

$$\hat{X}_n(h) = \hat{X}_{n+h} = \alpha_0 + \sum_{t=1}^n \alpha_t X_t,$$

où les coefficients $\alpha_0, \dots, \alpha_n$ sont donnés par

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \Gamma_{X,n}^{-1} \begin{pmatrix} \gamma_X(n+h-1) \\ \vdots \\ \gamma_X(h) \end{pmatrix}$$

avec la matrice de covariance du vecteur (X_1, \dots, X_n) , supposée inversible,

$$\Gamma_{X,n} = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(n-1) \\ \gamma_X(1) & \gamma_X(0) & \ddots & \gamma_X(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_X(n-1) & \cdots & \gamma_X(1) & \gamma_X(0) \end{pmatrix}$$

et

$$\mu_X = \alpha_0 + \mu_X \sum_{t=1}^n \alpha_t.$$

Notons que si le processus $(X_t)_{t \in \mathbb{Z}}$ est centré, alors le premier coefficient α_0 est nul et

$$\hat{X}_n(h) = \hat{X}_{n+h} = \sum_{t=1}^n \alpha_t X_t.$$

Preuve

Par définition de la prévision linéaire optimale et grâce à la Proposition 4.10, on a que

$$\begin{aligned} \mathbb{E}[\hat{X}_n(h)] = \mathbb{E}[X_{n+h}] &\Leftrightarrow \mathbb{E}\left[\alpha_0 + \sum_{t=1}^n \alpha_t X_t\right] = \mathbb{E}[X_{n+h}] \\ &\Leftrightarrow \alpha_0 + \mu_X \sum_{t=1}^n \alpha_t = \mu_X \end{aligned}$$

car le processus $(X_t)_{t \in \mathbb{Z}}$ est stationnaire. De plus, pour tout $i = 1, \dots, n$, $\mathbb{E}[X_{n+h} X_i] = \mathbb{E}[\hat{X}_n(h) X_i]$ se réécrit

$$\mathbb{E}[X_{n+h} X_i] = \alpha_0 \mathbb{E}[X_i] + \sum_{t=1}^n \alpha_t \mathbb{E}[X_t X_i].$$

On a ainsi, pour tout $i = 1, \dots, n$,

$$\begin{aligned} \mathbb{E}[X_{n+h} X_i] - \mathbb{E}[X_{n+h}] \mathbb{E}[X_i] &= \alpha_0 \mathbb{E}[X_i] + \sum_{t=1}^n \alpha_t \mathbb{E}[X_t X_i] - \mu_X \mathbb{E}[X_i] \\ &= \alpha_0 \mathbb{E}[X_i] + \sum_{t=1}^n \alpha_t \mathbb{E}[X_t X_i] - \mathbb{E}[X_i] \left\{ \alpha_0 + \sum_{t=1}^n \alpha_t \mu_X \right\} \\ &= \sum_{t=1}^n \alpha_t (\mathbb{E}[X_t X_i] - \mu_X^2) = \sum_{t=1}^n \alpha_t \text{Cov}(X_t, X_i). \end{aligned}$$

On a donc obtenu, pour tout $i = 1, \dots, n$, $\gamma_X(n+h-i) = \sum_{t=1}^n \alpha_t \gamma_X(t-i)$, ce qui donne l'équation matricielle

$$\begin{pmatrix} \gamma_X(n+h-1) \\ \vdots \\ \gamma_X(h) \end{pmatrix} = \Gamma_{X,n} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

L'hypothèse d'inversibilité de la matrice $\Gamma_{X,n}$ permet d'obtenir l'équation donnée dans la proposition.

Exemple 4.2 (Processus autorégressif d'ordre 1 AR(1)). Soit $X_t = \phi X_{t-1} + \varepsilon_t$, $\forall t \in \mathbb{Z}$, où $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$ et $0 < |\phi| < 1$. On rappelle que les ε_t sont indépendants du passé de la série

temporelle et $\gamma_X(h) = \phi^{|h|}\gamma_X(0)$, $\forall h \in \mathbb{Z}$.

La prévision linéaire optimale de X_{n+1} sur la base des observations de X_1, \dots, X_n est de la forme

$$\hat{X}_n(1) = \hat{X}_{n+1} = \sum_{t=1}^n \alpha_t X_t,$$

avec

$$\begin{pmatrix} 1 & \phi & \dots & \phi^{n-1} \\ \phi & 1 & \ddots & \phi^{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ \phi^{n-1} & \dots & \phi & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \phi^n \\ \phi^{n-1} \\ \vdots \\ \phi \end{pmatrix}.$$

Comme $(\alpha_1, \dots, \alpha_n) = (0, \dots, 0, \phi)$ est une solution de cette équation et l'unicité de la projection orthogonale donne

$$\hat{X}_n(1) = \hat{X}_{n+1} = \phi X_n.$$

4.2.5 Evolution des prévisions linéaires optimales en fonction de la taille de la mémoire

Soit un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ tel que ses matrices d'autocorrélation $R_X(h)$ sont inversibles pour tout h dans \mathbb{N} . On s'intéresse ici à la prévision linéaire optimale de X_t pour une taille de mémoire k , c'est-à-dire en fonction de l'observation des v.a. X_{t-1}, \dots, X_{t-k} .

Sans perte de généralité, on suppose le processus est centré.

D'après la Proposition 4.12, on a vu comment obtenir les coefficients $\alpha_1(k), \dots, \alpha_k(k)$ de la prévision linéaire optimale de X_t en fonction du passé observé X_{t-1}, \dots, X_{t-k} :

$$\text{EL}(X_t | \mathcal{H}_{t-k}^{t-1}) = \alpha_1(k)X_{t-1} + \dots + \alpha_k(k)X_{t-k}$$

avec

$$\begin{pmatrix} \alpha_1(k) \\ \vdots \\ \alpha_k(k) \end{pmatrix} = R_X(k)^{-1} \begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(k) \end{pmatrix}$$

Si on augmente la taille de la mémoire, il faut à chaque instant d'observation supplémentaire inverser la matrice de corrélation d'après l'expression précédente. Nous allons donc chercher une **méthode itérative** permettant de déterminer les nouveaux coefficients (avec une mémoire de taille $k+1$) en fonction des anciens (avec une mémoire de taille k). Pour cela, nous avons besoin de quelques lemmes techniques.

Lemma 4.1. *Les coefficients de la régression de X_t sur le passé de taille de mémoire k sont les mêmes que ceux de la régression de X_t sur les k prochaines variables du processus :*

$$EL(X_t | \mathcal{H}_{t-k}^{t-1}) = \sum_{i=1}^k \alpha_i(k) X_{t-i} \implies EL(X_t | \mathcal{H}_{t+1}^{t+k}) = \sum_{i=1}^k \alpha_i(k) X_{t+i}$$

Preuve

On note $\beta_1(k), \dots, \beta_k(k)$ les coefficients de la régression de X_t sur \mathcal{H}_{t+1}^{t+k} (engendré par les k variables futures). Par le même raisonnement, ces coefficients vérifient

$$\begin{pmatrix} \beta_1(k) \\ \vdots \\ \beta_k(k) \end{pmatrix} = R_X(k)^{-1} \begin{pmatrix} \rho_X(-1) \\ \vdots \\ \rho_X(-k) \end{pmatrix},$$

puisque la matrice de corrélation ne dépend que des écarts temporels. Par parité de la fonction d'autocorrélation, le vecteur des corrélations à droite n'est autre que le vecteur $(\rho_X(1), \dots, \rho_X(k))$. Ainsi les $\beta_i(k)$ et les $\alpha_i(k)$ satisfont la même équation donc $\alpha_i(k) = \beta_i(k)$, pour $i = 1, \dots, k$.

Lemma 4.2. *On a l'équation récursive suivante exprimant les coefficients pour une mémoire de taille k en fonction de ceux d'une mémoire de taille $k-1$*

$$\alpha_i(k) = \alpha_i(k-1) - \alpha_k(k) \alpha_{k-i}(k-1), \quad \forall i = 1, \dots, k-1.$$

Preuve

On commence par remarquer que

$$\mathcal{H}_{t-(k-1)}^{t-1} = \overline{\text{Vect}}(X_{t-(k-1)}, \dots, X_{t-1}) \subset \overline{\text{Vect}}(X_{t-(k-1)}, \dots, X_t) = \mathcal{H}_{t-(k-1)}^t$$

donc en terme de projection orthogonale, on a que

$$P_{t-(k-1)}^{t-1}(X_t) = P_{t-(k-1)}^{t-1} \circ P_{t-k}^{t-1}(X_t).$$

Ainsi, on a :

$$\begin{aligned} EL(X_t | t - (k-1)^{t-1}) &= P_{t-(k-1)}^{t-1} (\alpha_1(k) X_{t-1} + \dots + \alpha_{k-1}(k) X_{t-(k-1)} + \alpha_k(k) x_{t-k}) \\ &= \alpha_1(k) X_{t-1} + \dots + \alpha_{k-1}(k) X_{t-(k-1)} + \alpha_k(k) EL(X_{t-k} | t - (k-1)^{t-1}). \end{aligned}$$

Mais on a aussi que

$$EL(X_t | t - (k-1)^{t-1}) = \alpha_1(k-1) X_{t-1} + \dots + \alpha_{k-1}(k-1) X_{t-(k-1)}.$$

Or d'après le Lemma 4.1,

$$\begin{aligned}
\text{EL}(X_{t-k}|t - (k-1)^{t-1}) &= \text{EL}(X_{t-k}|t - k + 1^{(t-k)+(k-1)}) \\
&= \sum_{i=1}^{k-1} \alpha_i(k-1)X_{(t-k)+i} \\
&= \alpha_1(k-1)X_{t-k+1} + \alpha_2(k-1)X_{t-k+2} + \dots + \alpha_{k-1}(k-1)X_{t-1}
\end{aligned}$$

En rassemblant les expressions, on a

$$\begin{aligned}
&\alpha_1(k-1)X_{t-1} + \dots + \alpha_{k-1}(k-1)X_{t-(k-1)} \\
&= \alpha_1(k)X_{t-1} + \dots + \alpha_{k-1}(k)X_{t-(k-1)} \\
&\quad + \alpha_k(k) \{ \alpha_1(k-1)X_{t-k+1} + \alpha_2(k-1)X_{t-k+2} + \dots + \alpha_{k-1}(k-1)X_{t-1} \}
\end{aligned}$$

Donc

$$\begin{cases} \alpha_1(k-1) &= \alpha_1(k) + \alpha_k(k)\alpha_{k-1}(k-1) \\ \alpha_2(k-1) &= \alpha_2(k) + \alpha_k(k)\alpha_{k-2}(k-1) \\ &\vdots \\ \alpha_{k-1}(k-1) &= \alpha_{k-1}(k) + \alpha_k(k)\alpha_1(k-1) \end{cases}$$

Du Lemma 4.2, on peut constater que l'on peut obtenir un algorithme récursif pour calculer les coefficients si l'on est capable d'exprimer le dernier terme $\alpha_k(k)$ en fonction des $\alpha_i(k-1)$. C'est l'objectif du lemme suivant.

Lemma 4.3. *On a la relation*

$$\alpha_k(k) = \frac{\rho_X(k) - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(k-i)}{1 - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(i)} \quad \forall k \geq 2$$

Preuve

On revient à l'équation matricielle :

$$\begin{aligned}
\begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(k) \end{pmatrix} &= R_X(k) \begin{pmatrix} \alpha_1(k) \\ \vdots \\ \alpha_k(k) \end{pmatrix} \\
&= \begin{pmatrix} 1 & \rho_X(1) & \dots & \rho_X(k-1) \\ \rho_X(1) & 1 & \ddots & \rho_X(k-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho_X(k-1) & \dots & \rho_X(1) & 1 \end{pmatrix} \begin{pmatrix} \alpha_1(k) \\ \vdots \\ \alpha_k(k) \end{pmatrix}
\end{aligned}$$

La dernière ligne de ce produit matriciel nous donne l'équation suivante :

$$\rho_X(k) = \sum_{i=1}^{k-1} \rho_X(k-i) \alpha_i(k) + \alpha_k(k).$$

Donc, en utilisant le Lemma 4.2,

$$\begin{aligned} \alpha_k(k) &= \rho_X(k) - \sum_{i=1}^{k-1} \rho_X(k-i) \alpha_i(k) \\ &= \rho_X(k) - \sum_{i=1}^{k-1} \rho_X(k-i) (\alpha_i(k-1) - \alpha_k(k) \alpha_{k-i}(k-1)) \\ &= \rho_X(k) - \sum_{i=1}^{k-1} \rho_X(k-i) \alpha_i(k-1) + \alpha_k(k) \sum_{i=1}^{k-1} \rho_X(k-i) \alpha_{k-i}(k-1) \\ &= \rho_X(k) - \sum_{i=1}^{k-1} \rho_X(k-i) \alpha_i(k-1) + \alpha_k(k) \sum_{u=1}^{k-1} \rho_X(u) \alpha_u(k-1) \end{aligned}$$

ce qui, en isolant le terme $\alpha_k(k)$, donne la formule voulue.

Avec le Lemma 4.2 et le Lemma 4.3, on en déduit l'algorithme de Durbin-Levinson.

Proposition 4.13 (Algorithme de Durbin-Levinson). *Les coefficients de la régression linéaire $EL(X_t | \mathcal{H}_{t-k}^{t-1})$ pour une mémoire de taille k s'obtiennent en fonction de ceux de la régression linéaire $EL(X_t | \mathcal{H}_{t-(k-1)}^{t-1})$ pour une mémoire de taille $k-1$ grâce aux formules récursives suivantes*

$$\begin{cases} \alpha_i(k) &= \alpha_i(k-1) - \alpha_k(k) \alpha_{k-i}(k-1), \\ \alpha_k(k) &= \frac{\rho_X(k) - \sum_{i=1}^{k-1} \alpha_i(k-1) \rho_X(k-i)}{1 - \sum_{i=1}^{k-1} \alpha_i(k-1) \rho_X(i)}, \quad \forall k \geq 2 \\ \alpha_1(1) &= \rho_X(1). \end{cases}$$

4.3 Autocorrélations partielles

Comme nous avons pu le voir dans la section précédente, le coefficient $\alpha_k(k)$ devant X_{t-k} dans la prévision linéaire optimale de X_t en fonction du passé fini \mathcal{H}_{t-k}^{t-1} de la série temporelle $(X_t)_{t \in \mathbb{Z}}$ joue un rôle particulier. Ces coefficients $\alpha_k(k)$ sont appelés **autocorrélations partielles**. Ils vont être au coeur de cette section.

Proposition 4.14. Le coefficient $\alpha_k(k)$ défini dans

$$EL(X_t|\mathcal{H}_{t-k}^{t-1}) = \alpha_1(k)X_{t-1} + \dots + \alpha_k(k)X_{t-k}$$

correspond au **coefficient de corrélation** entre les variables $X_t - EL(X_t|\mathcal{H}_{t-(k-1)}^{t-1})$ et $X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-(k-1)}^{t-1})$.

Preuve

Par définition de la prévision linéaire optimale, on a

$$EL(X_t|t - k^{t-1}) = \sum_{i=1}^k \alpha_i(k)X_{t-i}.$$

On a vu précédemment au début de la preuve du Lemma 4.2 l'égalité

$$EL(X_t|t - k + 1^{t-1}) = \sum_{i=1}^{k-1} \alpha_i(k)X_{t-i} + \alpha_k(k)EL(X_{t-k}|t - (k-1)^{t-1}).$$

De ces deux équations, on obtient

$$EL(X_t|t - k^{t-1}) - EL(X_t|t - (k-1)^{t-1}) = \alpha_k(k) [X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1})].$$

On en déduit que

$$\begin{aligned} & \text{Cov} \left(EL(X_t|t - k^{t-1}) - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) \\ &= \alpha_k(k) \text{Var} \left(X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right). \end{aligned}$$

Par ailleurs,

$$\begin{aligned} & \text{Cov} \left(EL(X_t|t - k^{t-1}) - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) \\ &= \text{Cov} \left(X_t - EL(X_t|t - (k-1)^{t-1}) - (X_t - EL(X_t|t - k^{t-1})), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) \\ &= \text{Cov} \left(X_t - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) \\ & \quad - \text{Cov} \left(X_t - EL(X_t|t - k^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) \\ &= \text{Cov} \left(X_t - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right) - 0, \end{aligned}$$

car la seconde covariance est nulle puisque l'on a $X_t - EL(X_t|t - k^{t-1}) \perp t - k^{t-1}$ et $X_{t-k} - EL(X_{t-k}|t - k + 1^{t-1}) \in t - k^{t-1}$.

Des deux dernières égalités on en déduit que

$$\begin{aligned} \alpha_k(k) &= \frac{\text{Cov} \left(X_t - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right)}{\text{Var} \left(X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right)} \\ &= \frac{\text{Cov} \left(X_t - EL(X_t|t - (k-1)^{t-1}), X_{t-k} - EL(X_{t-k}|t - k + 1^{t-1}) \right)}{\sqrt{\text{Var} \left(X_t - EL(X_t|t - (k-1)^{t-1}) \right)} \sqrt{\text{Var} \left(X_{t-k} - EL(X_{t-k}|t - (k-1)^{t-1}) \right)}}, \end{aligned}$$

la dernière égalité étant justifiée par l'invariance par translation temporelle des covariances.

Definition 4.8. Ce coefficient de corrélation $\alpha_k(k)$ est appelé **autocorrélation partielle d'ordre k** et est noté $r_X(k)$.

L'autocorrélation partielle s'interprète donc comme la corrélation entre X_t et X_{t-k} quand on leur a retiré leurs meilleures explications données par les variables intermédiaires.

Proposition 4.15. *Il est équivalent de connaître le vecteur $(\rho_X(1), \dots, \rho_X(k))$ ou le vecteur $(r_X(1), \dots, r_X(k))$.*

Exemple 4.3. On donne ici l'exemple des autocorrélations et autocorrélations partielles empiriques de quelques séries temporelles simulées.

```
MA1<-arima.sim(n=500,list(ma=c(0.7)))
autoplplot(acf(MA1,plot=FALSE))
autoplplot(pacf(MA1,plot=FALSE))
```

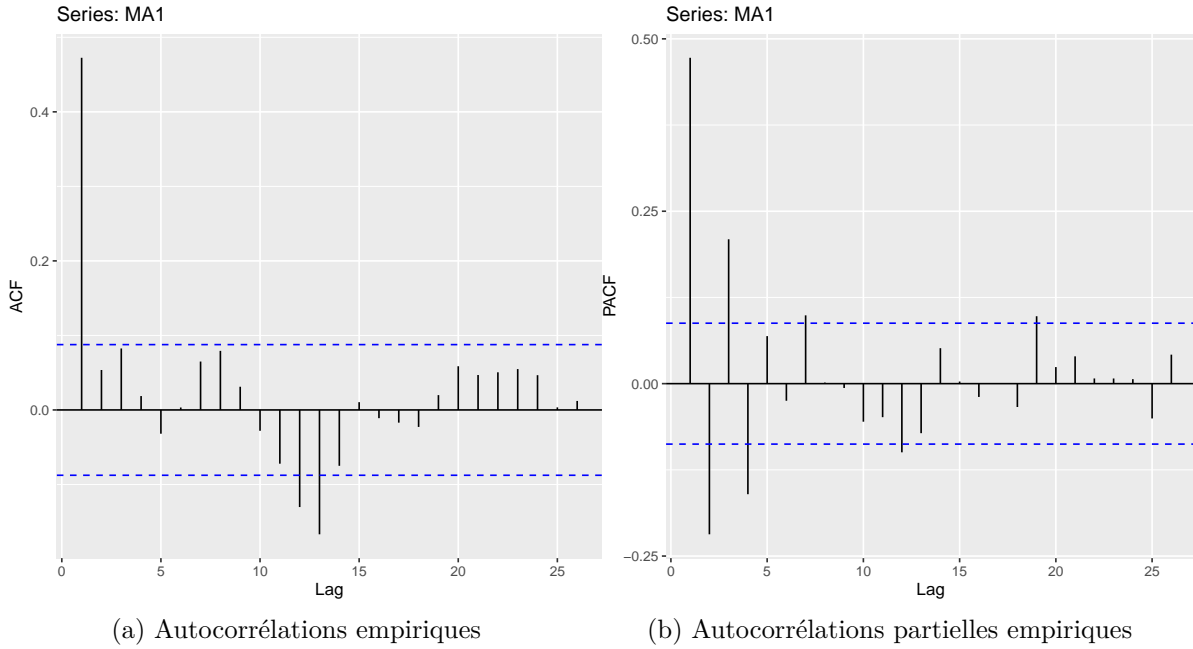


Figure 4.2: Résultats pour la série MA(1) $X_t = \varepsilon_t - 0.7\varepsilon_{t-1}$, $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$

```
AR1<-arima.sim(n=500,list(ar=c(0.7)))
autoplot(acf(AR1,plot=FALSE))
autoplot(pacf(AR1,plot=FALSE))
```

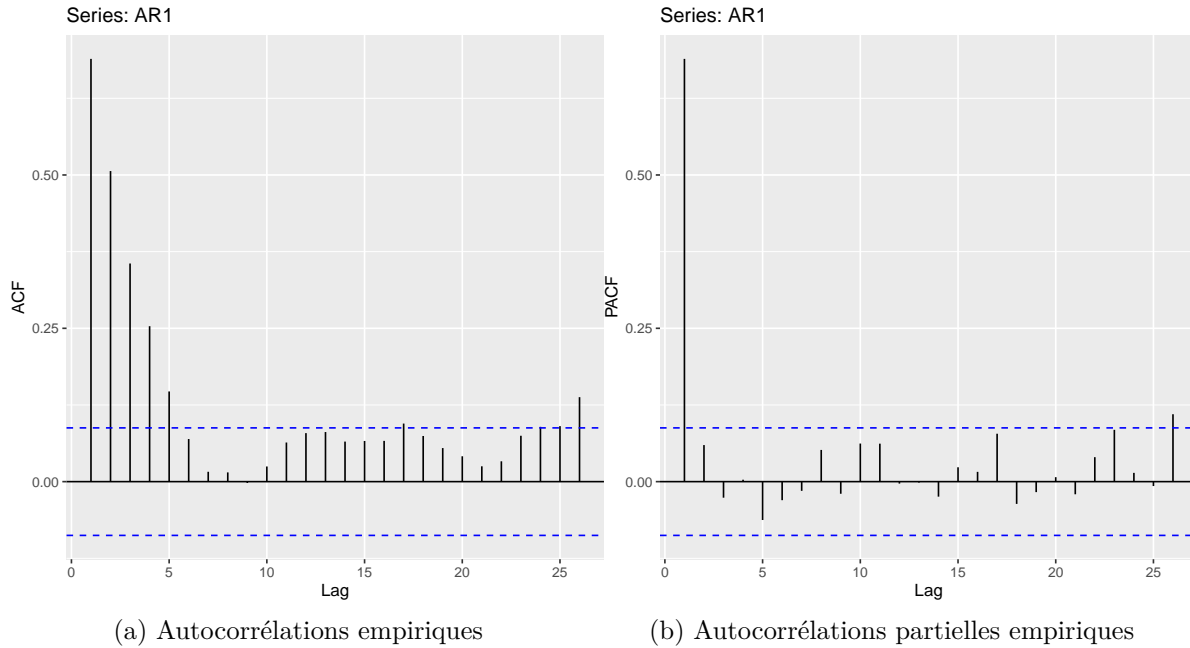


Figure 4.3: Résultats pour la série AR(1) $X_t - 0.7X_{t-1} = \varepsilon_t$, $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$

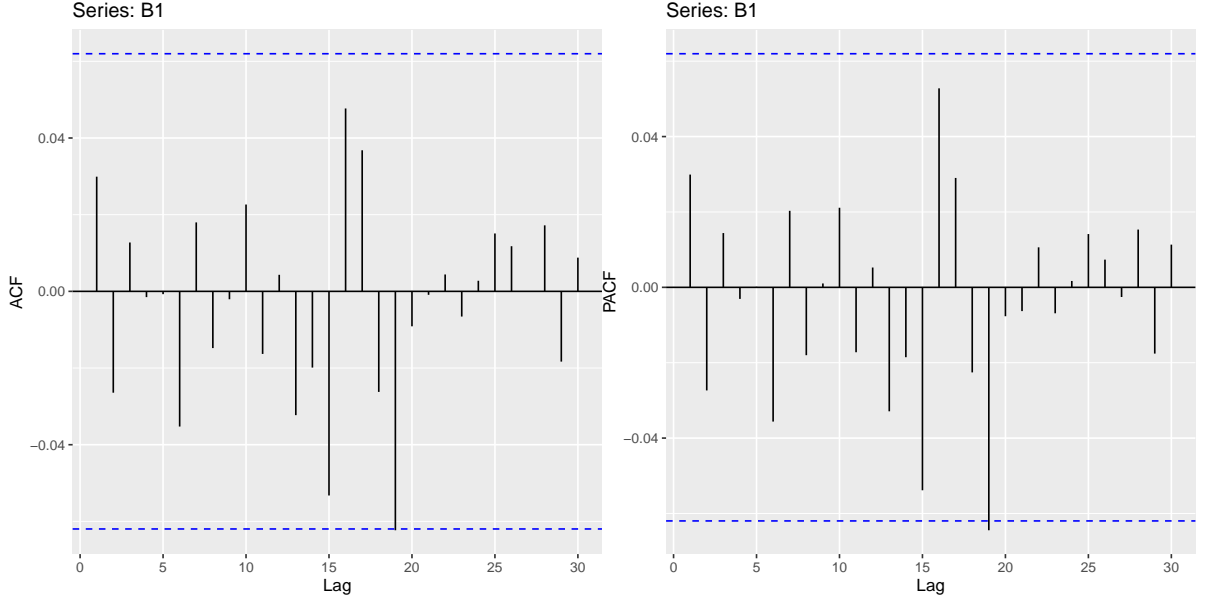
```
B1<-rnorm(1000,0,1)
autoplot(acf(B1,plot=FALSE))
autoplot(pacf(B1,plot=FALSE))
```

4.4 Tests de blancheur d'un processus

Quand on va aborder les modèles ARMA, on va chercher à décomposer la partie stationnaire de la série (après avoir estimé ou éliminé tendance et saisonnalité) en une partie exploitable pour la prévision et une partie de bruit blanc. Aussi on a besoin de pouvoir tester la blancheur des résidus.

On souhaite donc ici tester l'hypothèse

$$\mathcal{H}_0 : (X_t)_{t \in \mathbb{Z}} \text{ est un bruit blanc}$$



(a) Autocorrélations empiriques

(b) Autocorrélations partielles empiriques

Figure 4.4: Résultats pour un bruit blanc gaussien $\mathcal{N}(0, 1)$

contre

$\mathcal{H}_1 : (X_t)_{t \in \mathbb{Z}}$ n'est pas un bruit blanc.

Pour réaliser ce test, on suppose que l'on observe X_1, \dots, X_n . Le théorème suivant présente des statistiques de test pour répondre à ce test de blancheur.

Theorem 4.2. • *Statistique de **Portmanteau***

$$Q_k = n \sum_{h=1}^k \hat{\rho}_{X,n}(h)^2 \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k)$$

• *Statistique de **Ljung-Box***

$$Q_k^* = n(n+2) \sum_{h=1}^k \frac{\hat{\rho}_{X,n}(h)^2}{n-h} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k)$$

Sous l'hypothèse \mathcal{H}_0 (bruit blanc), on a vu que les autocorrélations sont données par $\rho_X(h) = \mathbb{1}_{h=0}$, $\forall h \in \mathbb{Z}$. Ainsi les statistiques Q_k et Q_k^* ont tendance à être faibles sous \mathcal{H}_0 et élevées sous \mathcal{H}_1 . On va donc rejeter \mathcal{H}_0 si ces statistiques sont élevées.

Proposition 4.16 (Procédures de test). *On rejette \mathcal{H}_0 au risque α si Q_k (resp. Q_k^*) est supérieure à $q_{1-\alpha,k}$, le $(1-\alpha)$ -quantile de la loi $\chi^2(k)$*

Zone de rejet :

$$\mathcal{R}_\alpha^{(k)} = \{Q_k^* > q_{1-\alpha,k}\} \quad (\text{resp. } \mathcal{R}_\alpha^{(k)} = \{Q_k > q_{1-\alpha,k}\}).$$

Remarquons que l'on a un test pour chaque valeur de k !

Exemple 4.4. Pour illustrer ces tests, on considère un bruit blanc gaussien. On simule une série de taille 300 et on utilise la fonction `Box.test` pour tester la blancheur de cette série.

```
Xt<-rnorm(300,0,1)
# Box.test(Xt,lag=1,type="Box-Pierce")
Box.test(Xt,lag=10,type="Ljung-Box")
```

Box-Ljung test

```
data:  Xt
X-squared = 21.354, df = 10, p-value = 0.01876
```

On ne rejette donc pas \mathcal{H}_0 pour $k = 10$. La Figure 4.5 représente les p-values du même test pour k variant de 1 à 10. Sur la Figure 4.6, on observe que les autocorrélations empiriques sont proche de 0 pour tout $|h| > 0$ donc proche du comportement théorique des autocorrélations d'un bruit blanc.

```
Box.Ljung.Test(Xt)
ggacf(Xt)
```

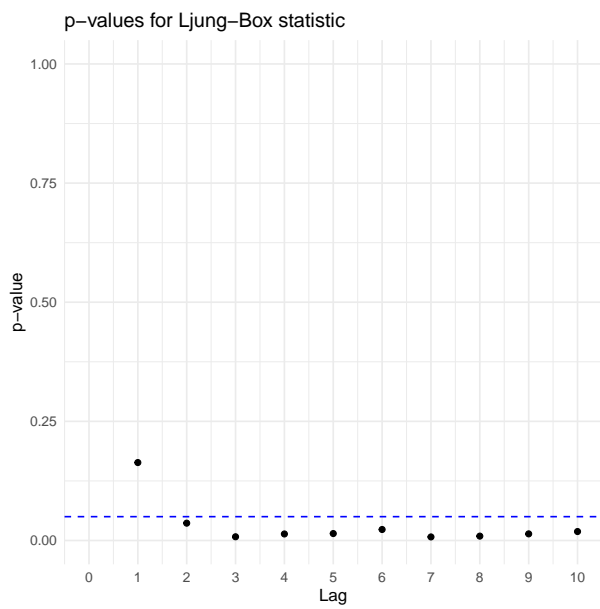


Figure 4.5: p-values du test de Ljung-Box pour plusieurs valeurs de k

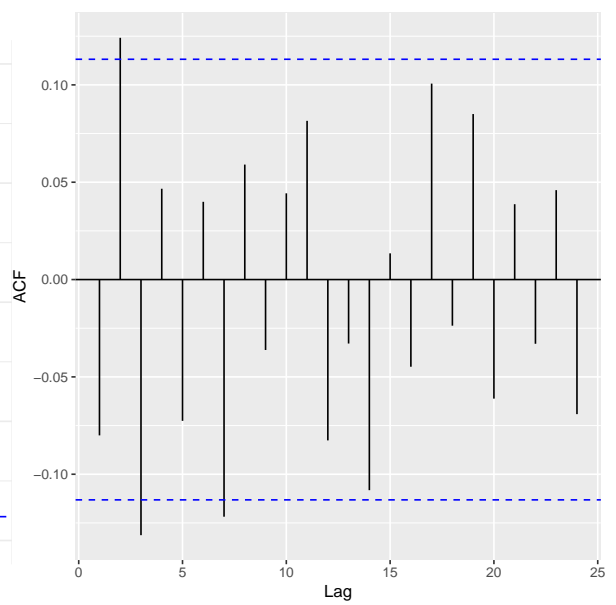


Figure 4.6: Autocorrélations empiriques de la série de bruit blanc gaussien

5 Les modèles ARMA

Dans ce chapitre, nous allons nous intéresser aux modèles ARMA qui sont très couramment utilisés dans l'étude des séries temporelles. Ce sont des modèles paramétriques linéaires de séries temporelles proposés par Box et Jenkins. Leurs écriture et analyse utilisent abondamment les opérateurs retard B et avance F . Aussi, nous allons débiter ce chapitre en étudiant quelques propriétés sur des polynômes ou séries “en B ” que nous exploiterons ensuite dans le cadre des processus ARMA.

5.1 Polynômes et séries en B

5.1.1 Définitions

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire du second ordre. Rappelons que l'opérateur retard B est défini par $BX_t = X_{t-1}$ et l'opérateur avance F par $FX_t = X_{t+1}$. Par composition, on a alors que $B^k X_t = X_{t-k}$ et $F^k X_t = X_{t+k}$. Puis par extension, un **polynôme en B** (ou F) est défini par

$$\left(\sum_{i=1}^p a_i B^i \right) X_t = \sum_{i=1}^p a_i X_{t-i}.$$

On a déjà utilisé ce type d'opérateur pour définir les moyennes mobiles (voir Chapter 2). Nous avons également considéré la notion de série en B quand nous avons introduit la notion de filtrage linéaire (voir Chapter 4). Rappelons que nous avons vu que si $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire, alors le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}$$

est également stationnaire si $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$.

Définition 5.1. Soit $(a_i)_{i \in \mathbb{Z}}$ une famille absolument sommable de réels $(\sum_{i \in \mathbb{Z}} |a_i| < +\infty)$. On appelle **série en B** de coefficients $(a_i)_{i \in \mathbb{Z}}$ l'opérateur

$$P(B) = \sum_{i \in \mathbb{Z}} a_i B^i$$

sur les processus stationnaires qui transforme le processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ en un processus stationnaire $(Y_t)_{t \in \mathbb{Z}}$ défini par , pour tout $t \in \mathbb{Z}$, par

$$Y_t = P(B)X_t = \left(\sum_{i \in \mathbb{Z}} a_i B^i \right) X_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}, \quad \forall t \in \mathbb{Z}$$

La manipulation des séries en B se fait comme celle des séries réelles, en particulier pour les combinaisons linéaires et composées de séries.

Proposition 5.1. *On a les résultats suivants :*

- La **combinaison linéaire** de deux séries en B est une série en B .
- La **composée** de deux séries en B est une série en B .

Preuve

Soient $P(B) = \sum_{i \in \mathbb{Z}} a_i B^i$ et $Q(B) = \sum_{i \in \mathbb{Z}} b_i B^i$ deux séries en B , donc telles que $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$ et $\sum_{i \in \mathbb{Z}} |b_i| < +\infty$. Considérons l'opérateur $\lambda P(B) + Q(B)$ et appliquons le à un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$. On obtient, pour tout $t \in \mathbb{Z}$:

$$\begin{aligned} (\lambda P(B) + Q(B)) X_t &= \lambda P(B)X_t + Q(B)X_t = \sum_{i \in \mathbb{Z}} \lambda a_i X_{t-i} + \sum_{i \in \mathbb{Z}} b_i X_{t-i} \\ &= \lim_{n \rightarrow +\infty} \sum_{i=-n}^{i=+n} \lambda a_i X_{t-i} + \lim_{n \rightarrow +\infty} \sum_{i=-n}^{i=+n} b_i X_{t-i} \\ &= \lim_{n \rightarrow +\infty} \sum_{i=-n}^{i=+n} (\lambda a_i X_{t-i} + b_i X_{t-i}) = \sum_{i \in \mathbb{Z}} (\lambda a_i + b_i) X_{t-i}, \end{aligned}$$

où la dernière égalité est justifiée par l'absolue convergence de la série $\sum (\lambda a_i + b_i)$ garantie par l'inégalité

$$\sum_{i \in \mathbb{Z}} |\lambda a_i + b_i| \leq \lambda \sum_{i \in \mathbb{Z}} |a_i| + \sum_{i \in \mathbb{Z}} |b_i| < +\infty.$$

On a bien alors

$$\lambda \sum_{i \in \mathbb{Z}} a_i B^i + \sum_{i \in \mathbb{Z}} b_i B^i = \sum_{i \in \mathbb{Z}} (\lambda a_i + b_i) B^i.$$

Considérons maintenant la composition des deux séries $P(B)$ et $Q(B)$. On a

$$P(B) \circ Q(B) X_t = P(B) Y_t,$$

où

$$Y_t = \sum_{j \in \mathbb{Z}} b_j X_{t-j}.$$

On a donc

$$\begin{aligned}
P(B) \circ Q(B)X_t &= \sum_{i \in \mathbb{Z}} a_i Y_{t-i} = \lim_{n \rightarrow +\infty} \sum_{i=-n}^n a_i Y_{t-i} \\
&= \lim_{n \rightarrow +\infty} \sum_{i=-n}^n a_i \left(\lim_{m \rightarrow +\infty} \sum_{j=-m}^m b_j X_{t-i-j} \right) \\
&= \lim_{n \rightarrow +\infty} \lim_{m \rightarrow +\infty} \sum_{i=-n}^n \sum_{j=-m}^m a_i b_j X_{t-i-j} \\
&= \lim_{n \rightarrow +\infty} \lim_{m \rightarrow +\infty} \sum_{k=-n-m}^{n+m} \left(\sum_{i=\max(-n, k-m)}^{\min(n, k+m)} a_i b_{k-i} \right) X_{t-k},
\end{aligned}$$

où la dernière égalité est obtenue en faisant un changement d'indice de j à $k = i + j$. On sait que les séries $\sum a_i$ et $\sum b_j$ étant absolument sommables, la série produit de Cauchy (appelé aussi parfois convoluée) $\sum c_k$ de ces deux suites, où

$$c_k = \sum_{i \in \mathbb{Z}} a_i b_{k-i},$$

pour $k \in \mathbb{Z}$, l'est aussi. On a donc :

$$P(B) \circ Q(B)X_t = \sum_{k \in \mathbb{Z}} c_k B^k,$$

ce qui montre que la composition des deux séries en B est bien une série en B .

5.1.2 Inversion de $I - \lambda B$

Dans l'étude des modèles ARMA, on va avoir besoin de savoir quand et comment on peut inverser des polynômes en B . On va donc ici commencer par le polynôme de degré 1 en B : $I - \lambda B$.

L'opérateur $P(B) = I - \lambda B$ associe à tout processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ un nouveau processus stationnaire $(Y_t)_{t \in \mathbb{Z}}$ tel que

$$Y_t = (I - \lambda B)X_t = X_t - \lambda X_{t-1}.$$

Se donnant un processus stationnaire $(Y_t)_{t \in \mathbb{Z}}$, le problème de l'inversion du polynôme $I - \lambda B$ revient à déterminer s'il existe un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifiant :

$$Y_t = (I - \lambda B)X_t = X_t - \lambda X_{t-1}$$

et de donner son expression en fonction du processus $(Y_t)_{t \in \mathbb{Z}}$.

On va montrer dans la suite que ce polynôme $P(B) = I - \lambda B$ est inversible seulement si $|\lambda| \neq 1$ et on obtiendra dans ce cas l'expression de son inverse.

5.1.2.1 1er Cas : $|\lambda| < 1$

- **Une solution :**

Soit la suite $(a_i)_{i \in \mathbb{Z}}$ définie par $a_i = 0$ pour $i \in \mathbb{Z} \setminus \mathbb{N}$ et $a_i = \lambda^i$ pour $i \in \mathbb{N}$. On a alors

$$\sum_{i \in \mathbb{Z}} |a_i| = \sum_{i \in \mathbb{N}} |\lambda|^i < +\infty.$$

La série en B de coefficients $(a_i)_{i \in \mathbb{Z}}$ est donc bien définie et on peut écrire

$$\begin{aligned} (I - \lambda B) \left(\sum_{i \in \mathbb{Z}} a_i B^i \right) &= (I - \lambda B) \left(\sum_{i \in \mathbb{N}} \lambda^i B^i \right) \\ &= \sum_{i \in \mathbb{N}} \lambda^i B^i - \sum_{i \in \mathbb{N}} \lambda^{i+1} B^{i+1} = I. \end{aligned}$$

Ainsi, pour un processus stationnaire $(Y_t)_{t \in \mathbb{Z}}$ donné, le processus $(X_t)_{t \in \mathbb{Z}}$ défini par $X_t = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i}$ est solution de l'équation

$$Y_t = (I - \lambda B)X_t.$$

- **Non unicité :**

$(X_t)_{t \in \mathbb{Z}}$ avec $X_t = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i}$ n'est pas l'unique solution. En effet, soit $(X_t^*)_{t \in \mathbb{Z}}$ tel que $(I - \lambda B)X_t^* = 0$ et définissons le processus $(\tilde{X}_t)_{t \in \mathbb{Z}}$ par $\tilde{X}_t = X_t + X_t^*$. Alors on a que $Y_t = (I - \lambda B)\tilde{X}_t$.

Un processus $(X_t^*)_{t \in \mathbb{Z}}$ vérifiant $(I - \lambda B)X_t^* = X_t^* - \lambda X_{t-1}^* = 0$ est de la forme $X_t^* = \lambda^t A$, pour tout $t \in \mathbb{Z}$, où A est une v.a.r. de L^2 . Ainsi, étant donné un processus $(Y_t)_{t \in \mathbb{Z}}$, les solutions de l'équation $Y_t = (I - \lambda B)X_t$, $\forall t \in \mathbb{Z}$ sont de la forme

$$\tilde{X}_t = X_t + X_t^* = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i} + \lambda^t A, \quad t \in \mathbb{Z}$$

Mais le processus $(X_t^*)_{t \in \mathbb{Z}}$ n'est pas stationnaire !

- **Unique solution stationnaire :**

Etant donné le processus $(Y_t)_{t \in \mathbb{Z}}$, la **solution stationnaire** de l'équation $Y_t = (I - \lambda B)X_t$ est le processus $(X_t)_{t \in \mathbb{Z}}$ défini par

$$X_t = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i}.$$

Proposition 5.2. *L'opérateur $(I - \lambda B)$ est inversible dans la classe des processus stationnaires d'inverse $\sum_{i=0}^{+\infty} \lambda^i B^i$ quand $|\lambda| < 1$.*

5.1.2.2 2ème Cas : $|\lambda| > 1$

On peut écrire

$$I - \lambda B = -\lambda B \left(I - \frac{1}{\lambda} F \right)$$

Comme $|\frac{1}{\lambda}| < 1$, par le même raisonnement que précédent, le polynôme $(I - \frac{1}{\lambda} F)$ est inversible dans la classe des processus stationnaires d'inverse $\sum_{i=0}^{+\infty} \lambda^{-i} F^i$. Le processus $-\lambda B$ est inversible d'inverse $(-1/\lambda)F$.

Proposition 5.3. *Quand $|\lambda| > 1$, l'opérateur $I - \lambda B$ est inversible dans la classe des processus stationnaires d'inverse*

$$\left(\frac{-1}{\lambda} F \right) \left(\sum_{i=0}^{+\infty} \frac{1}{\lambda^i} F^i \right) = - \sum_{i=0}^{+\infty} \frac{1}{\lambda^{i+1}} F^{i+1} = - \sum_{j=1}^{+\infty} \frac{1}{\lambda^j} F^j$$

Etant donné un processus stationnaire $(Y_t)_{t \in \mathbb{Z}}$, l'unique solution stationnaire de l'équation $Y_t = (I - \lambda B)X_t$ est donc $(X_t)_{t \in \mathbb{Z}}$ défini par

$$X_t = - \sum_{j=1}^{+\infty} \frac{1}{\lambda^j} Y_{t+j}.$$

Remarque

Dans le cas $|\lambda| > 1$, l'expression de la solution $(X_t)_{t \in \mathbb{Z}}$ est fonction du futur du processus $(Y_t)_{t \in \mathbb{Z}}$ et non en fonction de son passé comme dans le cas $|\lambda| < 1$.

5.1.2.3 3ème Cas : $|\lambda| = 1$

Proposition 5.4. *Si $|\lambda| = 1$, l'opérateur $I - \lambda B$ n'est pas inversible dans la classe des processus stationnaires.*

Preuve

- Soit $(X_t)_{t \in \mathbb{Z}}$ le processus constant $X_t = m$, $\forall t \in \mathbb{Z}$. On a alors $(I - B)X_t = m - m = 0$ ce qui prouve que l'opérateur n'est pas injectif et donc pas bijectif.
- On peut aussi montrer que $(I - B)$ n'est pas surjectif :

si le processus constant $(Y_t)_{t \in \mathbb{Z}}$ égal à $m \neq 0$ avait un antécédent $(X_t)_{t \in \mathbb{Z}}$ stationnaire, ce dernier serait tel que $X_t - X_{t-1} = m$, $\forall t \in \mathbb{Z}$. On arrive alors à une contradiction car $\mathbb{E}[X_t - X_{t-1}] = 0 \neq m$.

5.1.3 Inverse d'un polynôme en B

On considère maintenant un polynôme en B de la forme

$$\Phi(B) = I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p.$$

et on répondre à la question suivante : $\Phi(B)$ est-il inversible ? Se donnant un processus $(Y_t)_{t \in \mathbb{Z}}$, existe-t-il un unique processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ tel que

$$\Phi(B)X_t = Y_t, \quad \forall t \in \mathbb{Z}.$$

Dans les complexes, le polynôme $\Phi(z)$ a p racines non nécessairement distinctes. Comme il est non constant, ce polynôme est scindé. On peut donc alors utiliser les résultats du polynôme $I - \lambda B$. On a alors la discussion de cas suivante :

- Si au moins une des racines est de module 1, alors il n'existe pas de processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ solution de l'équation $\Phi(B)X_t = Y_t$.
- Si toutes les racines sont de module différent de 1, alors il existe une série en B , notée $\Psi(B) = \sum \psi_i B^i$, telle que :
 - $\Phi(B)\Psi(B) = I$
 - $(X_t)_{t \in \mathbb{Z}}$, solution de $\Phi(B)X_t = Y_t$, est stationnaire.
- Si toutes les racines sont à l'extérieur du disque unité, alors l'inverse est une série en puissances positives de B uniquement

$$\Psi(B) = \sum_{i \in \mathbb{N}} \psi_i B^i$$

- Si toutes les racines sont à l'intérieur du disque unité, alors l'inverse est une série en puissances strictement positives de F uniquement

$$\Psi(B) = \sum_{i \in \mathbb{N}^*} \psi_i F^i$$

5.1.3.1 Comment déterminer l'inverse de $\Phi(B)$?

Si l'inverse du polynôme $\Phi(B)$ existe, on peut le déterminer par l'une de ces méthodes

- *Identification* :

Par exemple quand toutes les racines sont à l'extérieur du disque unité, on écrit

$$(1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p) \left(\sum_{i \in \mathbb{N}} \psi_i z^i \right) = 1$$

dont on tire des équations donnant les expressions des coefficients ψ_i en fonction des φ_i .

- *Décomposition en éléments simples de $\Phi(z)$* : On écrit

$$\begin{aligned} \frac{1}{\Phi(z)} &= \frac{1}{\prod_{i=1}^p (1 - \lambda_i z)} \\ &= \sum_{i=1}^p \frac{a_i}{1 - \lambda_i z} \\ &= \sum_{i=1}^p a_i \sum_{j \in \mathbb{Z}} \lambda_i^j z^j \\ &= \sum_{j \in \mathbb{Z}} \left(\sum_{i=1}^p a_i \lambda_i^j \right) z^j = \Psi(z). \end{aligned}$$

5.2 Processus AR

5.2.1 Définition

Dans cette partie, on se restreint à des **processus centrés** sans perte de généralités

Dans de nombreuses situations pratiques la valeur en un instant t d'une série temporelle peut s'écrire comme la somme d'une combinaison linéaire des valeurs précédentes de la série et d'un terme de bruit. Un tel modèle est connu sous le nom d'un **processus AR (AutoRegressive)**.

Definition 5.2 (AR(p)). On dit qu'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation **auto-régressive d'ordre p** (noté AR(p)) s'il vérifie l'équation récurrente :

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p} = \varepsilon_t,$$

où $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$ avec $\varphi_p \neq 0$.

Réécriture : Un processus AR(p) vérifie l'équation : $\Phi(B)X_t = \varepsilon_t$, $\forall t \in \mathbb{Z}$ avec le polynôme en B

$$\Phi(B) = I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p.$$

Cette représentation est dite **canonique** si le bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ est tel que $\varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(X)$ pour tout $t \in \mathbb{Z}$.

Remarque

Comme un processus AR doit être stationnaire par déf., les racines du polynôme $\Phi(z)$ doivent toutes être de module différent de 1 !

Soit un processus AR(1) vérifiant

$$X_t - \varphi X_{t-1} = \varepsilon_t \text{ avec } \varphi = \pm 1.$$

Par itération, on a $X_t = \varphi^t X_0 + \sum_{j=0}^{t-1} \varphi^j \varepsilon_{t-j}$.

$$\text{Var}(X_t) + \text{Var}(\varphi^t X_0) - 2\varphi^t \text{Cov}(X_t, X_0) = \sum_{j=0}^{t-1} \varphi^{2j} \text{Var}(\varepsilon_{t-j}).$$

S'il existe un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifiant cette équation, on doit donc avoir $2 - 2\varphi^t \rho_X(t) = \frac{t\sigma^2}{\gamma_X(0)}$. On aboutit à une contradiction en faisant tendre t vers $+\infty$.

5.2.2 Un exemple instructif : le processus AR(1)

On considère $(X_t)_{t \in \mathbb{Z}}$ un processus AR(1), vérifiant donc l'équation de récurrence

$$X_t - \varphi X_{t-1} = \varepsilon_t, \quad \forall t \in \mathbb{Z}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc $\text{WN}(0, \sigma^2)$ et $\varphi \neq 0$.

D'après l'étude précédente sur l'inversibilité de $I - \varphi B$, trois cas se présentent suivant les valeurs du paramètre φ .

5.2.2.1 Cas $|\varphi| = 1$

On a vu précédemment que dans ce cas, il n'existe pas de processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifiant l'équation $(I - \varphi B)X_t = \varepsilon_t$. Il n'existe donc pas de processus AR(1) de paramètre 1 ou -1.

5.2.2.2 Cas $|\varphi| < 1$

- **Ecriture de X_t en fonction du bruit blanc :**

Pour $|\varphi| < 1$, l'inversibilité de $I - \varphi B$ est assurée dans la classe des processus stationnaires et

$$(I - \varphi B)^{-1} = \sum_{i=0}^{+\infty} \varphi^i B^i.$$

Ainsi le processus AR(1) pour $|\varphi| < 1$ s'écrit, en fonction du bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$, sous la forme :

$$X_t = \sum_{i=0}^{+\infty} \varphi^i \varepsilon_{t-i}, \quad \forall t \in \mathbb{Z}.$$

De cette écriture, on en déduit que $X_t \in \mathcal{H}_{-\infty}^t(\varepsilon)$ et donc $\mathcal{H}_{-\infty}^t(X) \subset \mathcal{H}_{-\infty}^t(\varepsilon)$ pour tout t dans \mathbb{Z} . De plus, par propriété des bruits blancs $\varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(\varepsilon)$. Ainsi $\varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(X)$, donc cette représentation est **canonique**.

- **Prévision linéaire optimale :**

Ayant observé X_1, \dots, X_n , on prédit la valeur suivante de la série X_{n+1} par

$$\hat{X}_n(1) = \text{EL}(X_{n+1} | \mathcal{H}_{-\infty}^n(X)) = \text{EL}(\varphi X_n + \varepsilon_{n+1} | \mathcal{H}_{-\infty}^n(X)) = \varphi X_n$$

car $\varepsilon_{n+1} \perp \mathcal{H}_{-\infty}^n(X)$.

- **Processus des innovations :**

Dans ce cas,

$$X_t - \hat{X}_{t-1}(1) = X_t - \varphi X_{t-1} = \varepsilon_t, \quad \forall t \in \mathbb{Z}$$

C'est donc le bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ dans la représentation AR(1). On verra que ce résultat reste valable pour tout modèle AR de représentation canonique.

5.2.2.3 Cas $|\varphi| > 1$

- **Ecriture (non canonique) de X_t selon le bruit blanc :**

D'après l'étude précédente, le polynôme $I - \varphi B$ est inversible d'inverse

$$(I - \varphi B)^{-1} = - \sum_{i=1}^{+\infty} \frac{1}{\varphi^i} B^i \text{ d'où } X_t = - \sum_{i=1}^{+\infty} \frac{1}{\varphi^i} \varepsilon_{t+i}.$$

Cette représentation du processus AR(1) **n'est pas canonique** car

$$\begin{aligned}
\text{Cov}(\varepsilon_t, X_{t-1}) &= \text{Cov}\left(\varepsilon_t, -\sum_{i=1}^{+\infty} \frac{1}{\varphi^i} \varepsilon_{t-1+i}\right) \\
&= -\frac{1}{\varphi} \text{Var}(\varepsilon_t) = -\frac{\sigma^2}{\varphi} \neq 0.
\end{aligned}$$

- **Prévision linéaire optimale :**

Ayant observé X_1, \dots, X_n , on prédit la valeur suivante de la série X_{n+1} par

$$\begin{aligned}
\hat{X}_n(1) &= \text{EL}(X_{n+1} | \mathcal{H}_{-\infty}^n(X)) \\
&= \text{EL}(\varphi X_n + \varepsilon_{n+1} | \mathcal{H}_{-\infty}^n(X)) \\
&= \varphi X_n + \text{EL}(\varepsilon_{n+1} | \mathcal{H}_{-\infty}^n(X))
\end{aligned}$$

le dernier terme n'ayant aucune raison d'être nul, puisque la représentation n'est pas canonique.

- **Processus des innovations :**

$$\begin{aligned}
X_{t+1} - \hat{X}_t(1) &= \varphi X_t + \varepsilon_{t+1} - \varphi X_t - \text{EL}(\varepsilon_{t+1} | \mathcal{H}_{-\infty}^t(X)) \\
&= \varepsilon_{t+1} - \text{EL}(\varepsilon_{t+1} | \mathcal{H}_{-\infty}^t(X))
\end{aligned}$$

donc non confondus avec le bruit blanc utilisé dans la représentation (non canonique) du processus AR(1).

- **Vers une représentation canonique :**

On va ici montrer que le processus AR(1) pour $|\varphi| > 1$ possède une autre représentation, qui est canonique. Pour cela, on utilise la formule reliant les densités spectrales de deux processus stationnaires dont l'un est le filtrage linéaire d'un autre : partant de $(I - \varphi B)X_t = \varepsilon_t$, on a

$$f_X(\omega) |1 - \varphi e^{-i\omega}|^2 = f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi} \iff f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\omega}|^2}.$$

Soit le processus $(\eta_t)_{t \in \mathbb{Z}}$ défini par $\eta_t = X_t - \frac{1}{\varphi} X_{t-1}$, $\forall t \in \mathbb{Z}$. Alors

$$\begin{aligned}
f_\eta(\omega) &= \hat{f}_X(\omega) \left| 1 - \frac{1}{\varphi} e^{-i\omega} \right|^2 = \frac{\sigma^2}{2\pi} \frac{|1 - \frac{1}{\varphi} e^{-i\omega}|^2}{|1 - \varphi e^{-i\omega}|^2} \\
&= \frac{\sigma^2}{2\pi\varphi^2} \frac{|1 - \frac{1}{\varphi} e^{-i\omega}|^2}{|\frac{1}{\varphi} - e^{-i\omega}|^2} = \frac{\sigma^2}{2\pi\varphi^2} \frac{|1 - \frac{1}{\varphi} e^{-i\omega}|^2}{|\frac{1}{\varphi} e^{i\omega} - 1|^2} = \frac{\sigma^2}{2\pi\varphi^2}.
\end{aligned}$$

On a donc une nouvelle représentation AR(1) pour le processus $(X_t)_{t \in \mathbb{Z}}$:

$$(I - \frac{1}{\varphi} B)X_t = \eta_t \text{ où } (\eta_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2/\varphi^2)$$

et la représentation est canonique car $|1/\varphi| < 1$. On a aussi que le processus $(\eta_t)_{t \in \mathbb{Z}}$ correspond au processus des innovations.

Cette étude du cas particulier du processus AR(1) nous a donc permis de voir que la représentation AR(1) n'est pas unique. En choisissant celle avec le coefficient φ ou $1/\varphi$ de module inférieur à 1, on a la représentation canonique. Ce résultat se généralise aux processus AR d'ordre quelconque.

5.2.3 Propriétés

La proposition suivante généralise les résultats vus sur la surjectivité et l'inversibilité du polynôme en $I - \lambda B$ et décrit le processus AR en fonction du bruit blanc.

Proposition 5.5. *Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc et $\Phi(B)$ un polynôme en B .*

- *Il existe une infinité de processus du second ordre $(X_t)_{t \in \mathbb{Z}}$ vérifiant :*

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z}.$$

- *Si les racines du polynôme $\Phi(B)$ sont **toutes de module différent de 1**, il existe une **unique solution stationnaire** $(X_t)_{t \in \mathbb{Z}}$. Elle s'écrit alors sous la forme d'un filtrage linéaire du bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$:*

$$X_t = \Psi(B)\varepsilon_t = \sum_{i \in \mathbb{Z}} \psi_i \varepsilon_{t-i}, \quad \forall t \in \mathbb{Z}.$$

- *Si les racines du polynômes sont **toutes à l'extérieur du disque unité**, alors l'écriture moyenne mobile du processus $AR(p)$ ne considère que les valeurs passées du bruit blanc :*

$$X_t = \Psi(B)\varepsilon_t = \sum_{i \in \mathbb{N}} \psi_i \varepsilon_{t-i}, \quad \forall t \in \mathbb{Z}.$$

Preuve

Pour s'en convaincre, il suffit de factoriser le polynôme $\Phi(B)$ de degré p en fonction de ses p racines et d'appliquer p fois le résultat d'inversibilité vu pour le polynôme $I - \lambda B$.

Comme vu dans le cas du processus AR(1), il n'existe pas qu'une seule représentation d'un processus AR(p). En revanche une seule de ces représentations est canonique.

Proposition 5.6. *Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc et $\Phi(B)$ un polynôme en B de degré p exactement dont toutes les racines $(z_i = 1/\lambda_i)_{i=1, \dots, p}$ (non néc. distinctes) sont de module différent de 1.*

Soit $(X_t)_{t \in \mathbb{Z}}$ l'unique processus stationnaire $AR(p)$ vérifiant

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z}.$$

- Le processus $(X_t)_{t \in \mathbb{Z}}$ satisfait plusieurs représentations $AR(p)$ différentes : soit $\tilde{\Phi}(B)$ un nouveau polynôme en B ayant pour racines celles de $\Phi(B)$ ou leurs inverses, le processus $(X_t)_{t \in \mathbb{Z}}$ admet aussi la représentation :

$$\tilde{\Phi}(B)X_t = \eta_t, \text{ pour tout } t \in \mathbb{Z}, \text{ où } (\eta_t)_{t \in \mathbb{Z}} \text{ un bruit blanc.}$$

- Toutes les représentations ont même ordre.
- Une seule représentation est canonique :

$$\tilde{\Phi}(B)X_t = \eta_t, \quad \forall t \in \mathbb{Z},$$

où

$$\tilde{\Phi}(B) = \prod_{i: |z_i| > 1} (I - \lambda_i B) \prod_{i: |z_i| < 1} \left(I - \frac{1}{\lambda_i} B \right).$$

Preuve

On montre seulement le dernier point. La démonstration est assez proche de ce que l'on a fait pour le processus $AR(1)$ et utilise donc les densités spectrales. De l'équation

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z},$$

on tire

$$f_X(\omega) |\Phi(e^{-i\omega})|^2 = f_\varepsilon(\omega).$$

L'équation

$$\tilde{\Phi}(B)X_t = \eta_t, \text{ pour tout } t \in \mathbb{Z},$$

donne quant à elle

$$f_\eta(\omega) = f_X(\omega) |\tilde{\Phi}(e^{-i\omega})|^2.$$

En supposant toutes les racines de $\Phi(z)$ réelles, on peut refaire le même type de raisonnement que celui fait pour le $AR(1)$ et obtenir

$$\begin{aligned} f_\eta(\omega) &= \frac{|\tilde{\Phi}(e^{-i\omega})|^2}{|\Phi(e^{-i\omega})|^2} f_\varepsilon(\omega) = \frac{\sigma^2 \left| \prod_{j: |z_j| > 1} (1 - \lambda_j e^{-i\omega}) \prod_{j: |z_j| < 1} \left(1 - \frac{1}{\lambda_j} e^{-i\omega}\right) \right|^2}{2\pi \left| \prod_{j=1}^p (1 - \lambda_j e^{-i\omega}) \right|^2} \\ &= \frac{\sigma^2 \left| \prod_{j: |z_j| < 1} \left(1 - \frac{1}{\lambda_j} e^{-i\omega}\right) \right|^2}{2\pi \left| \prod_{j: |z_j| < 1} (1 - \lambda_j e^{-i\omega}) \right|^2} = \frac{\sigma^2}{2\pi} \frac{1}{\prod_{j: |z_j| < 1} |\lambda_j|^2}. \end{aligned}$$

Que se passe-t-il si une racine $z = 1/\lambda$ n'est pas réelle ? Le polynôme $\Phi(z)$ étant à coefficient réel, son conjugué \bar{z} est également racine. On peut alors écrire en ne considérant

dans le ratio précédent que les termes correspondants à ces deux racines conjuguées :

$$\begin{aligned} \frac{\left| \left(1 - \frac{1}{\lambda} e^{-i\omega}\right) \left(1 - \frac{1}{\bar{\lambda}} e^{-i\omega}\right) \right|^2}{\left| (1 - \lambda e^{-i\omega}) (1 - \bar{\lambda} e^{-i\omega}) \right|^2} &= \frac{\left| \left(-\frac{1}{\lambda} e^{-i\omega}\right) \left(-\frac{1}{\bar{\lambda}} e^{-i\omega}\right) (1 - \lambda e^{i\omega}) (1 - \bar{\lambda} e^{i\omega}) \right|^2}{\left| (1 - \lambda e^{-i\omega}) (1 - \bar{\lambda} e^{-i\omega}) \right|^2} \\ &= \frac{1}{|\lambda|^4} \text{ (puisque } \overline{1 - \lambda e^{i\omega}} = 1 - \bar{\lambda} e^{-i\omega} \text{)} \end{aligned}$$

et on obtient le même résultat que dans le cas de racines réelles. On a donc prouvé que l'on avait, avec ce nouveau polynôme $\tilde{\Phi}(B)$ obtenu en remplaçant racines à l'intérieur du disque unité par leur inverse, une nouvelle représentation $AR(p)$ du processus $(X_t)_{t \in \mathbb{Z}}$ puisque $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc. Il reste à prouver que la représentation obtenue est canonique. Le raisonnement suit de près ce que l'on a fait pour le cas particulier du processus $AR(1)$. D'une part on montre l'égalité des histoires des processus $(X_t)_{t \in \mathbb{Z}}$ et $(\eta_t)_{t \in \mathbb{Z}}$, i.e.

$$-\infty^t(X) = -\infty^t(\eta).$$

On dit dans ce cas que le processus est **régulier**. C'est bien le cas ici puisque l'équation

$$\tilde{\Phi}(B)X_t = \eta_t, \text{ pour tout } t \in \mathbb{Z},$$

où $\tilde{\Phi}(B)$ est un polynôme avec puissance positives de B , assure la première inclusion : $-\infty^t(\eta) \subset -\infty^t(X)$. Mais comme $\tilde{\Phi}(B)$ possède par construction toutes ses racines hors du disque unité, il est inversible d'inverse ne faisant intervenir que les puissances positives de B :

$$X_t = \sum_{i \in \mathbb{N}} \tilde{\psi}_i \eta_{t-i}.$$

On en déduit l'inclusion inverse : $-\infty^t(X) \subset -\infty^t(\eta)$ et donc l'égalité. La représentation canonique est alors bien acquise puisque, en tant que bruit blanc, le processus $(\eta_t)_{t \in \mathbb{Z}}$ est évidemment orthogonal à son passé, donc ici également orthogonal au passé de $(X_t)_{t \in \mathbb{Z}}$.

Proposition 5.7. *Si la représentation du processus $AR(p)$*

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z},$$

est canonique, le bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ utilisé dans sa représentation est aussi le processus des innovations.

Preuve

Par définition le processus des innovations est donné par

$$\begin{aligned} X_{t+1} - \hat{X}_t(1) &= X_{t+1} - \text{EL}(\varphi_1 X_t + \dots + \varphi_p X_{t+1-p} + \varepsilon_{t+1} | -\infty^t(X)) \\ &= X_{t+1} - \varphi_1 X_t - \dots - \varphi_p X_{t+1-p} + 0 \text{ (puisque } \varepsilon_{t+1} \perp -\infty^t(X)) \\ &= \varepsilon_{t+1}, \end{aligned}$$

ce qui est bien le résultat annoncé.

Proposition 5.8. *Tout processus AR(p) de représentation canonique*

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z},$$

admet la représentation MA(∞) :

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}.$$

Preuve

Nous avons déjà obtenu les éléments de cette proposition. La seule nouveauté, que l'on aurait pu noter plus tôt, est l'égalité à 1 du premier coefficient ψ_0 dans l'inverse du polynôme $\Phi(z)$. Ceci se voit facilement par identification du terme constant.

5.2.4 Liaisons temporelles

Soit un processus $(X_t)_{t \in \mathbb{Z}}$ de représentation AR(p) canonique :

$$\Phi(B)X_t = \varepsilon_t, \quad \forall t \in \mathbb{Z}.$$

- Variance du processus $(X_t)_{t \in \mathbb{Z}}$:

$$\begin{aligned} \gamma_X(0) &= \text{Var}(X_t) = \text{Cov} \left(\sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, X_t \right) \\ &= \sum_{i=1}^p \varphi_i \text{Cov}(X_{t-i}, X_t) + \text{Cov}(\varepsilon_t, X_t) \\ &= \sum_{i=1}^p \varphi_i \gamma_X(i) + \text{Cov} \left(\varepsilon_t, \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \right) \\ &= \sum_{i=1}^p \varphi_i \underbrace{\gamma_X(i)}_{\gamma_X(0)\rho_X(i)} + \text{Var}(\varepsilon_t) \text{ (car canonique)}. \end{aligned}$$

où la dernière égalité est obtenue par orthogonalité du bruit blanc par rapport à l'histoire du processus $AR(p)$ dans sa représentation canonique. On a alors

$$\gamma_X(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho_X(i)}.$$

- Autocovariances : pour $h > 0$,

$$\begin{aligned} \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) &= \text{Cov} \left(X_t, \sum_{i=1}^p \varphi_i X_{t+h-i} + \varepsilon_{t+h} \right) \\ &= \sum_{i=1}^p \varphi_i \text{Cov}(X_t, X_{t+h-i}) + \text{Cov}(X_t, \varepsilon_{t+h}) \\ &= \sum_{i=1}^p \varphi_i \gamma_X(h-i) \text{ car } \varepsilon_{t+h} \perp \mathcal{H}_{-\infty}^{t+h-1}(X). \end{aligned}$$

- Autocorrélations : elles satisfont l'équation linéaire récurrente d'ordre p

$$\rho_X(h) = \sum_{i=1}^p \varphi_i \rho_X(h-i), \quad \forall h \in \mathbb{N}^* \quad (5.1)$$

Les autocorrélations vérifient donc l'équation matricielle dite de **Yule-Walker** :

$$\begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(p) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_X(p-1) \\ \rho_X(1) & 1 & \ddots & \rho_X(p-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho_X(p-1) & \cdots & \rho_X(1) & 1 \end{pmatrix}}_{R_X(p)} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}$$

et

$$\gamma_X(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho_X(i)}.$$

Proposition 5.9 (Autocorrélations simples d'un $AR(p)$). *Les autocorrélations simples d'un processus $AR(p)$ décroissent, de manière exponentielle ou sinusoidale amortie, vers 0.*

Preuve

L'équation Equation 5.1 vérifiée par les autocorrélations d'un AR(p) est une équation de récurrence linéaire de polynôme caractéristique

$$r^p - \varphi_1 r^{p-1} - \dots - \varphi_{p-1} r - \varphi_p = r^p \Phi\left(\frac{1}{r}\right).$$

Les racines (non nulles car φ_p est non nul) de ce polynôme sont les inverses de celles du polynôme $\Phi(z)$ définissant l'AR(p). La solution de l'équation Equation 5.1 dans \mathbb{C} est donc une combinaison linéaire de solutions de la forme

$$(a_{i,0} + a_{i,1}h + \dots + a_{i,k-1}h^{k-1})\lambda_i^h,$$

où k est l'ordre de multiplicité de la racine $z_i = 1/\lambda_i$ de $\Phi(z)$. Or, puisqu'on considère la représentation canonique, les racines z_i de $\Phi(z)$ sont de module à l'extérieur du disque unité. Ainsi, si toutes les solutions sont réelles alors les autocorrélations décroissent effectivement de manière exponentielle vers 0. Si une des racines $z_i = 1/\lambda_i$ de $\Phi(z)$ n'est pas réelle, on sait alors que son conjugué l'est aussi et que les solutions réelles correspondant à ces deux racines conjuguées s'écrivent sous la forme

$$|\lambda_i|^h \left[(a_{i,0} + a_{i,1}h + \dots + a_{i,k-1}h^{k-1}) \cos(\omega_i h) + (b_{i,0} + b_{i,1}h + \dots + b_{i,k-1}h^{k-1}) \sin(\omega_i h) \right],$$

où ω_i est le module de λ_i . Elles sont donc bien sinusoidales amorties.

Exemple 5.1. Pour un AR(1) d'équation canonique $X_t = \varphi X_{t-1} + \varepsilon_t$, on a $\rho_X(h) = \varphi^{|h|}$. On a bien la décroissance exponentielle des autocorrélations et on constate que le coefficient φ d'un AR(1) est en fait la première autocovariance.

Exemple 5.2. Soit le processus AR(2) défini par $X_t = 0.7X_{t-1} - 0.1X_{t-2} + \varepsilon_t$. Les racines du polynôme $\Phi(z) = 0.1z^2 - 0.7z + 1$ sont 2 et 5. La Figure 5.1 montre la décroissance exponentielle des autocorrélations empiriques de ce processus.

Soit le processus AR(2) défini par $X_t = \frac{3}{4}X_{t-1} - \frac{3}{16}X_{t-2} + \varepsilon_t$. Les racines du polynôme $\Phi(z) = \frac{3}{16}z^2 - \frac{3}{4}z + 1$ sont $2 \pm i2\sqrt{3}/3$. La Figure 5.2 montre les autocorrélations empiriques de ce processus.

Proposition 5.10. Autocorrélations partielles d'un AR(p)

Les autocorrélations partielles d'un processus AR(p) sont nulles à partir du rang $p + 1$:

- $r_X(p) = \varphi_p \neq 0$
- $r_X(k) = 0, \forall k > p$

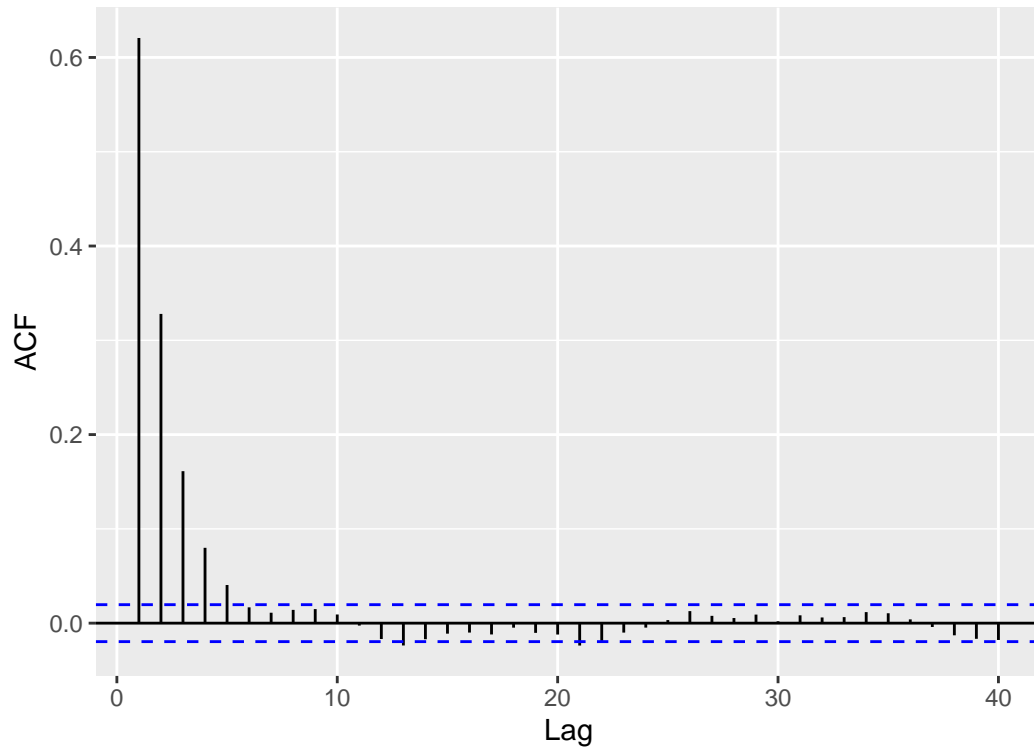


Figure 5.1: ACF empirique du processus AR(2) $X_t = 0.7X_{t-1} - 0.1X_{t-2} + \varepsilon_t$

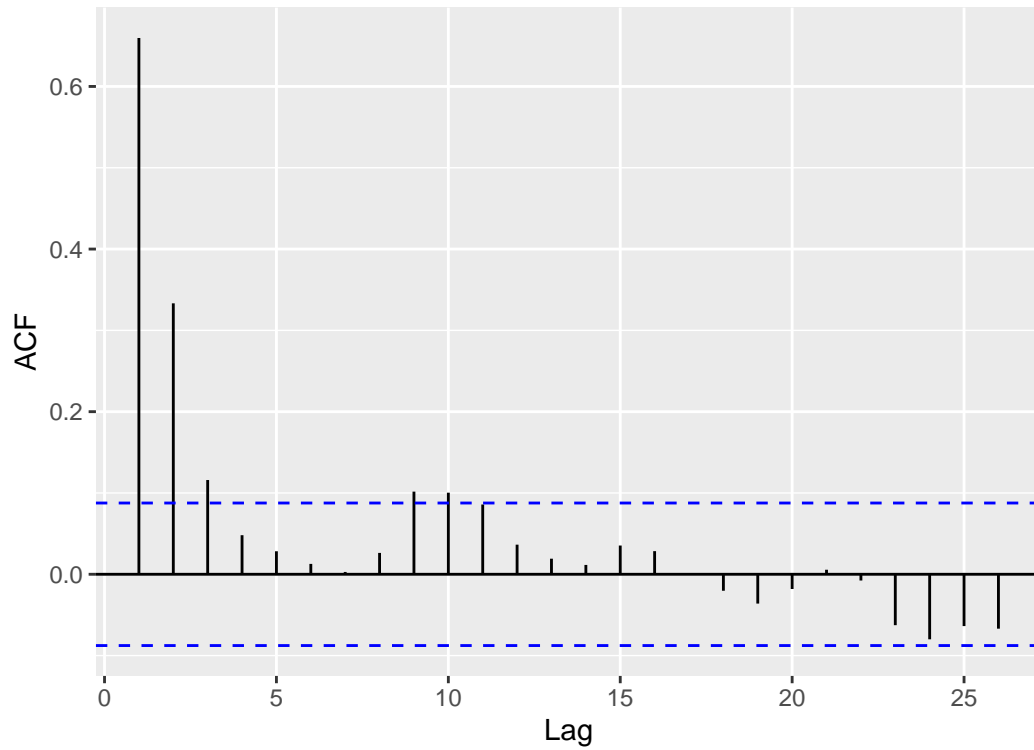


Figure 5.2: ACF empirique du processus AR(2) $X_t = \frac{3}{4}X_{t-1} - \frac{3}{16}X_{t-2} + \varepsilon_t$

Preuve

Il suffit d'écrire la définition de l'autocorrélation partielle. On sait que le coefficient $r_X(k)$ est donné par le coefficient $\alpha_k(k)$ dans la projection de X_t sur $t - k^{t-1}$, i.e. :

$$\text{EL}(X_t|t - k^{t-1}) = \alpha_1(k)X_{t-1} + \dots + \alpha_k(k)X_{t-k}.$$

Ainsi, pour $k = p$, on peut écrire

$$\begin{aligned}\text{EL}(X_t|t - p^{t-1}) &= \text{EL}(\varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t, |t - p^{t-1}) \\ &= \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p},\end{aligned}$$

dont on tire $r_X(p) = \varphi_p \neq 0$. Pour $k > p$, on a :

$$\begin{aligned}\text{EL}(X_t|t - k^{t-1}) &= \text{EL}(\varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t, |t - k^{t-1}) \\ &= \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} \\ &= \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} \\ &\quad + 0 \times X_{t-p-1} + \dots + 0 \times X_{t-k}\end{aligned}$$

et donc $r_X(k) = 0$.

Remarques

- La propriété $r_X(p) = \varphi_p$ n'est vraie que dans le cas d'une représentation canonique.
- On peut montrer que la propriété d'autocorrélations partielles nulles à partir d'un certain rang $p + 1$ est caractéristique d'un $\text{AR}(p)$.

Exemple 5.3. Pour illustrer la propriété que les autocorrélations partielles d'un processus $\text{AR}(p)$ sont nulles à partir du rang $p+1$, les autocorrélations partielles empiriques sont tracées en Figure 5.3 pour le processus $X_t = 0.7X_{t-1} - 0.1X_{t-2} + \varepsilon_t$ (à gauche - $\text{AR}(2)$) et le processus $X_t = 0.7X_{t-1} - 0.1X_{t-2} + 0.2X_{t-3} + \varepsilon_t$ (à droite - $\text{AR}(3)$).

5.3 Processus MA

5.3.1 Définition

Définition 5.3. On dit qu'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation en **moyenne mobile d'ordre q** (noté $\text{MA}(q)$) s'il vérifie l'équation

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

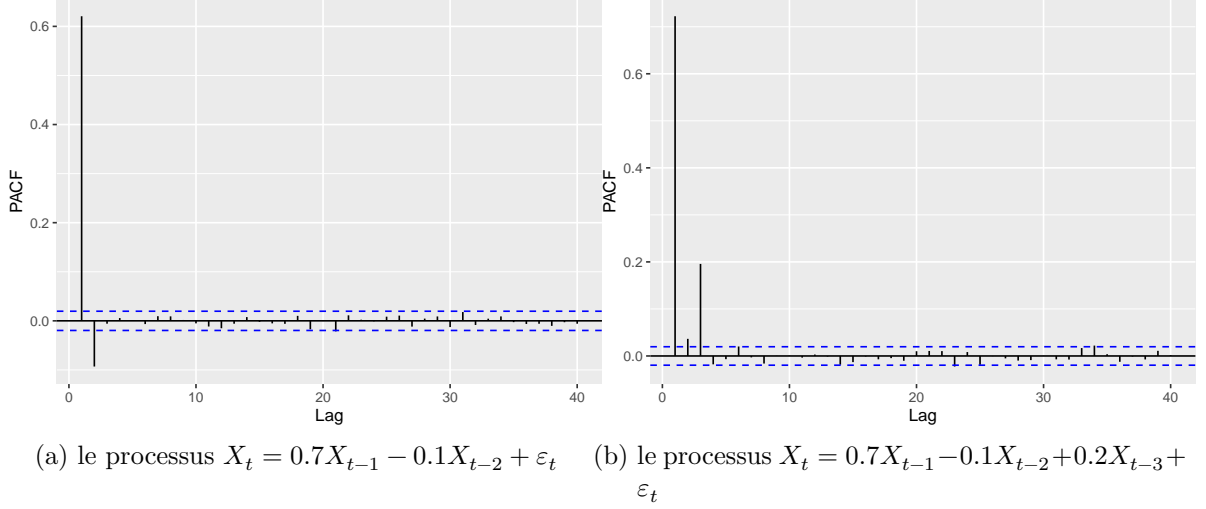


Figure 5.3: Autocorrélations partielles empiriques pour ...

où $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$ et $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$, $\theta_q \neq 0$.

Réécriture : $X_t = \Theta(B)\varepsilon_t$, $\forall t \in \mathbb{Z}$ avec le polynôme en B $\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$.

Cette représentation est dite **canonique** si les racines du polynôme $\Theta(z)$ sont toutes à l'extérieur du disque unité.

Remarque

Contrairement à un AR, un MA est entièrement spécifié. A bruit blanc et paramètres $\theta_1, \dots, \theta_q$ fixés, il ne correspond qu'un seul processus MA(q) qui est le filtrage linéaire de $(\varepsilon_t)_{t \in \mathbb{Z}}$ par le filtre

$$\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

5.3.2 Propriétés

Proposition 5.11. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc et $\Theta(B)$ un polynôme en B de degré q exactement dont toutes les racines (non néc. distinctes) sont de module différent de 1.

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) vérifiant $X_t = \Theta(B)\varepsilon_t$, $\forall t \in \mathbb{Z}$.

- Le processus $(X_t)_{t \in \mathbb{Z}}$ satisfait plusieurs représentations MA(q) différentes : soit $\tilde{\Theta}(B)$ un nouveau polynôme en B ayant pour racines celles de $\Theta(B)$ ou leurs inverses, le processus $(X_t)_{t \in \mathbb{Z}}$ admet aussi la représentation :

$$X_t = \tilde{\Theta}(B)\eta_t, \quad \forall t \in \mathbb{Z}, \quad \text{où } (\eta_t)_{t \in \mathbb{Z}} \text{ un bruit blanc.}$$

- Une seule représentation est canonique. Elle est obtenue en prenant le polynôme avec toutes les racines à l'extérieur du disque unité.
- Un processus MA est stationnaire **quelque soit sa représentation** car filtrage linéaire d'un bruit blanc.
- Un processus MA est toujours centré.

Proposition 5.12. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) de représentation canonique

$$X_t = \Theta(B)\varepsilon_t, \quad \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$$

Alors le processus $(X_t)_{t \in \mathbb{Z}}$ possède une représentation AR(∞) :

$$\varepsilon_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i} \text{ avec } \pi_0 = 1$$

Preuve

Par hypothèse la représentation est canonique donc les racines du polynôme $\Theta(z)$ sont toutes situées à l'extérieur du disque unité. Ainsi l'opérateur $\Theta(B)$ est inversible d'inverse

$$\Theta^{-1}(B) = \sum_{i=0}^{+\infty} \pi_i B^i = I + \sum_{i=1}^{+\infty} \pi_i B^i,$$

la dernière égalité étant obtenue par identification des termes constants. On a donc bien

$$\varepsilon_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i}.$$

Proposition 5.13. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) de représentation canonique

$$X_t = \Theta(B)\varepsilon_t, \quad \forall t \in \mathbb{Z} \text{ avec } (\varepsilon_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$$

Alors le processus des innovations correspond au bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ de sa représentation canonique.

Preuve

A partir de la Proposition 5.12, on a que

$$\varepsilon_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i}.$$

Les écritures $MA(q)$ et $AR(\infty)$ du processus $(X_t)_{t \in \mathbb{Z}}$ montrent que $-\infty^t(X) = -\infty^t(\varepsilon)$. On peut ainsi écrire

$$\begin{aligned} \text{EL}(X_t | -\infty^{t-1}(X)) &= \text{EL}(X_t | -\infty^{t-1}(\varepsilon)) \\ &= \text{EL}(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} | -\infty^{t-1}(\varepsilon)) \\ &= \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

par orthogonilité du bruit blanc par rapport à son passé. On en déduit pour l'innovation

$$X_t - \text{EL}(X_t | -\infty^{t-1}(X)) = X_t - (\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}) = \varepsilon_t,$$

ce qui achève la démonstration.

5.3.3 Liaisons temporelles

On considère un processus $MA(q)$

$$X_t = \Theta(B)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

avec $\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ et $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$.

- Variance du processus $(X_t)_{t \in \mathbb{Z}}$:
Par la propriété d'orthogonalité d'un bruit blanc, on a

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}\left(\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}\right) \\ &= \sigma^2 \left(1 + \sum_{j=1}^q \theta_j^2\right). \end{aligned}$$

- Fonction d'autocovariance de $(X_t)_{t \in \mathbb{Z}}$:

$$\begin{aligned} \gamma_X(h) &= \text{Cov}(X_t, X_{t+h}) \\ &= \text{Cov}\left(\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}\right). \end{aligned}$$

A partir de cette expression, on peut immédiatement remarquer que $\gamma_X(h) = 0$ pour $h > q$.

Pour $h = 1$, $\text{Cov}(X_t, X_{t+1}) = \mathbb{E} \left[(\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j})(\varepsilon_{t+1} + \sum_{i=1}^q \theta_i \varepsilon_{t+1-i}) \right] = \sigma^2(\theta_1 + \theta_1\theta_2 + \dots + \theta_{q-1}\theta_q)$.

Pour $h = 2$, $\text{Cov}(X_t, X_{t+2}) = \mathbb{E} \left[(\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j})(\varepsilon_{t+2} + \sum_{j=1}^q \theta_j \varepsilon_{t+2-j}) \right] = \sigma^2(\theta_2 + \theta_1\theta_3 + \dots + \theta_{q-2}\theta_q)$.

On en déduit pour tout h dans \mathbb{N}^* :

$$\gamma_X(h) = \sigma^2(\theta_h + \theta_1\theta_{1+h} + \dots + \theta_{q-h}\theta_q) = \sigma^2 \left(\theta_h + \sum_{i=1}^{q-h} \theta_i \theta_{i+h} \right).$$

- Fonction d'autocorrélation de $(X_t)_{t \in \mathbb{Z}}$:

Proposition 5.14 (Autocorrélations simples d'un $\text{MA}(q)$).

- Les autocorrélations simples d'un processus $\text{MA}(q)$ sont nulles à partir du rang $q + 1$:

$$\rho_X(q) \neq 0 \text{ et } \rho_X(h) = 0, \forall h > q.$$

- La propriété d'autocorrélations simples nulles à partir d'un certain rang $q + 1$ est caractéristique d'un $\text{MA}(q)$

On considère le processus $\text{MA}(2)$ défini par $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$. La Figure 5.4 représente les autocorrélations empiriques de ce processus. On a bien des autocorrélations proche de 0 à partir de $h > 2$.

La Figure 5.5 représente les autocorrélations empiriques du processus défini par $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2} - 0.2\varepsilon_{t-3}$. Les autocorrélations empiriques sont bien proche de 0 à partir de $h > 3$.

Proposition 5.15 (Autocorrélations partielles d'un $\text{MA}(q)$). *Les autocorrélations partielles d'un processus $\text{MA}(q)$ sont solutions d'une équation linéaire récurrente d'ordre q . Elles décroissent, de manière exponentielle ou sinusoïdale amortie, vers 0.*

Exemple 5.4. On reprend le processus $\text{MA}(2)$ défini par $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$. Les racines du polynôme $\Theta(z)$ associé sont 2 et 5. La Figure 5.6 représente les autocorrélations partielles empiriques.

La Figure 5.7 représente les autocorrélations partielles empirique du processus $\text{MA}(2)$ défini par $\text{MA}(2) X_t = \varepsilon_t + 0.4\varepsilon_{t-1} - 0.45\varepsilon_{t-2}$, où les racines du polynôme $\Theta(z)$ associé sont $2 \pm i2\sqrt{3}/3$.

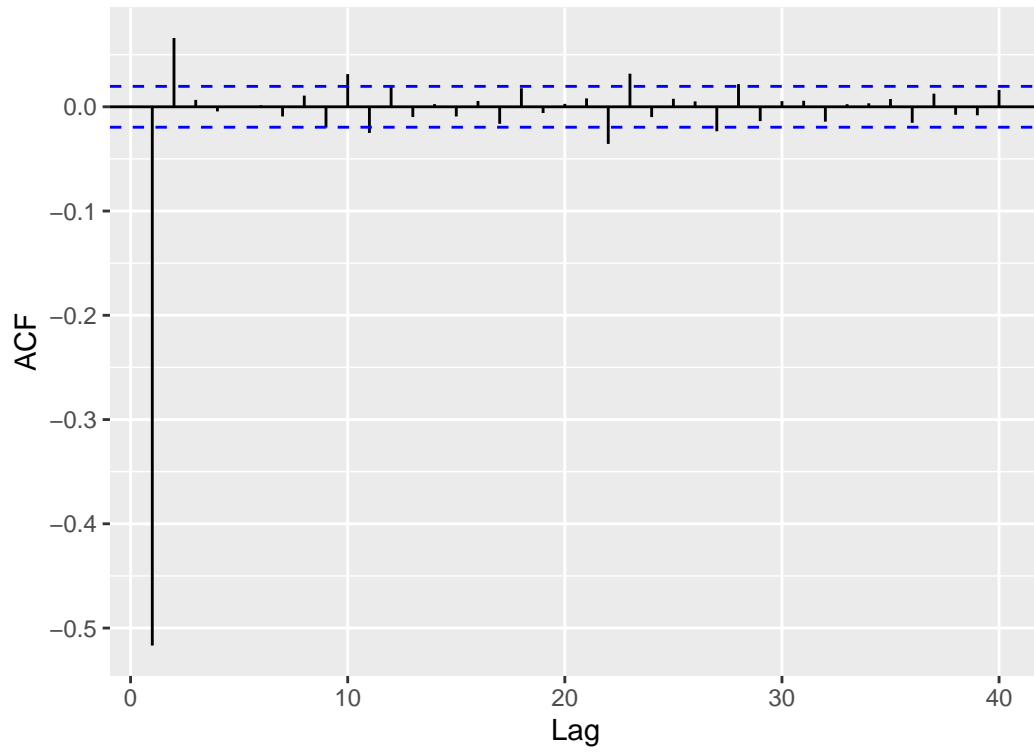


Figure 5.4: ACF empirique du processus MA(2) $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$

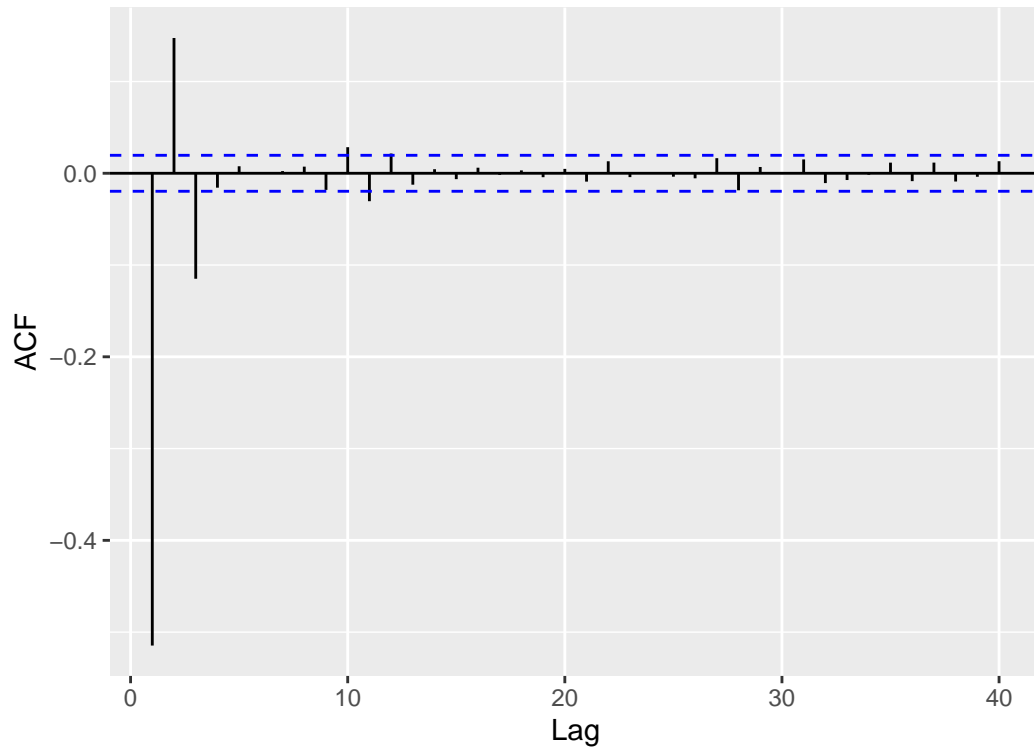


Figure 5.5: ACF empirique du processus MA(3) $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2} - 0.2\varepsilon_{t-3}$

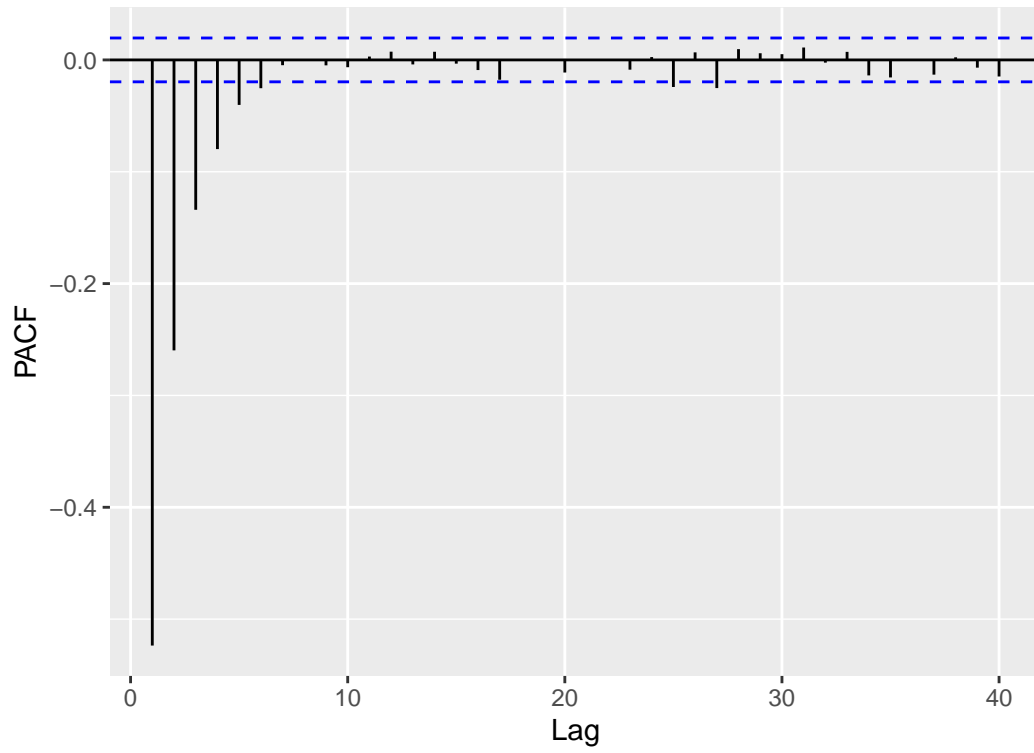


Figure 5.6: pACF empirique du processus MA(2) $X_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$

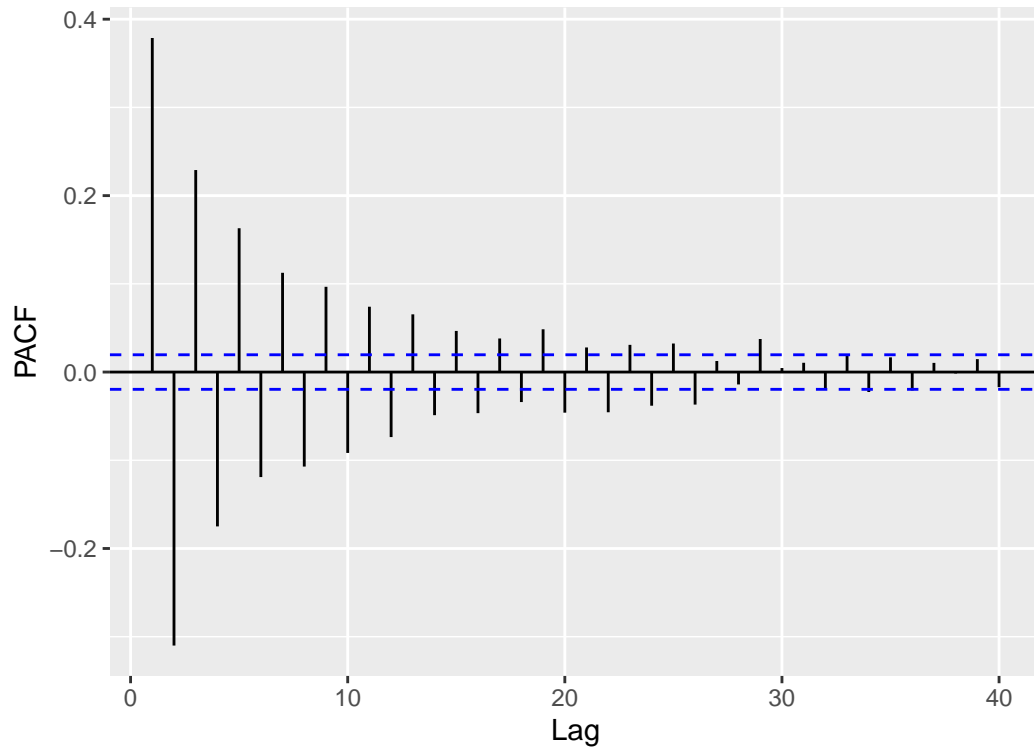


Figure 5.7: pACF empirique du processus MA(2) $X_t = \varepsilon_t + 0.4\varepsilon_{t-1} - 0.45\varepsilon_{t-2}$

5.4 Processus ARMA

Nous allons maintenant étudier des processus stationnaires avec une partie AR et une partie MA. On les appelle processus ARMA pour **AutoRegressive Moving Average**. Ils sont très importants en pratique car on peut montrer que tout processus stationnaire peut être approché par un processus ARMA.

5.4.1 Définition

Définition 5.4. On dit qu'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation **ARMA** (p, q) s'il vérifie l'équation

$$X_t = \sum_{j=1}^p \varphi_j X_{t-j} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

avec

- $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, \sigma^2)$
- $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p, \varphi_p \neq 0$
- $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q, \theta_q \neq 0$

Réécriture : En utilisant les polynômes en B

$$\begin{aligned} \Phi(B) &= I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \\ \text{et } \Theta(B) &= I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \end{aligned}$$

un processus **ARMA** (p, q) vérifie l'équation :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t, \quad \forall t \in \mathbb{Z}.$$

Définition 5.5. La représentation d'un processus ARMA (p, q) $\Phi(B)X_t = \Theta(B)\varepsilon_t$ est dite

- **minimale** si les polynômes $\Phi(z)$ et $\Theta(z)$ n'ont pas de racine commune
- **causale** si le polynôme $\Phi(z)$ a toutes ses racines à l'extérieur du disque unité
- **inversible** si le polynôme $\Theta(z)$ a toutes ses racines à l'extérieur du disque unité
- **canonique** si elle est causale et inversible.

Remarques

Si la représentation n'est pas minimale ($\Phi(z)$ et $\Theta(z)$ ont une ou des racine(s) commune(s)), alors

- Soit aucune de ces racines communes n'est sur le cercle unité. Dans ce cas, le

processus $(X_t)_{t \in \mathbb{Z}}$ admet aussi la représentation

$$\tilde{\Phi}(B)X_t = \tilde{\Theta}(B)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

où les nouveaux polynômes sont obtenus à partir des précédents en enlevant les racines communes.

- Si au moins une des racines communes est sur le cercle unité, alors il peut y avoir plus d'un unique processus stationnaire vérifiant l'équation. **Dans la suite, on ne considèrera que des représentations minimales.**

5.4.2 Ecritures $\text{MA}(\infty)$ et $\text{AR}(\infty)$

Proposition 5.16 (Ecriture $\text{MA}(\infty)$). *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $\text{ARMA}(p, q)$ de **représentation minimale et causale**. Il admet alors la représentation $\text{MA}(\infty)$*

$$X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i},$$

où les coefficients $(\psi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\psi_i - \sum_{j=1}^p \varphi_j \psi_{i-j} = \theta_i, \quad \forall i \in \mathbb{N},$$

avec $\psi_i = 0$ pour $i < 0$, $\psi_0 = 1$, $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$.

Preuve

La représentation étant causale, les racines du polynôme $\Phi(z)$ sont toutes à l'extérieur du disque unité. L'inverse de $\Phi(z)$ s'écrit alors comme une série de puissances positives de B . Multipliée par $\Theta(B)$, on garde une série en puissances positives de B et on a donc l'écriture :

$$X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}, \quad \text{pour tout } t \in \mathbb{Z}.$$

Grâce à cette écriture $\text{MA}(\infty)$ de $(X_t)_{t \in \mathbb{Z}}$, on peut réécrire l'équation

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

sous la forme

$$\begin{aligned} \left(\varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i} \right) - \varphi_1 \left(\varepsilon_{t-1} + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-1-i} \right) - \cdots - \varphi_p \left(\varepsilon_{t-p} + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-p-i} \right) \\ = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}. \end{aligned}$$

En identifiant les coefficients devant chaque ε_t , on obtient bien la formule annoncée.

Proposition 5.17 (Ecriture $AR(\infty)$). Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ de **représentation minimale et inversible**. Il admet alors la représentation $AR(\infty)$

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i},$$

où les coefficients $(\pi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\pi_i + \sum_{j=1}^q \theta_j \pi_{i-j} = -\varphi_i, \quad \forall i \in \mathbb{N},$$

avec $\pi_i = 0$ pour $i < 0$, $\varphi_0 = -1$ et $\varphi_i = 0$ pour $i > p$.

Preuve

La représentation étant inversible, les racines du polynôme $\Theta(z)$ sont toutes à l'extérieur du disque unité. L'inverse de $\Theta(z)$ s'écrit alors comme une série de puissances positives de B . Multipliée par $\Phi(B)$, on garde une série en puissances positives de B et on a donc l'écriture :

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i}, \quad \text{pour tout } t \in \mathbb{Z}.$$

Grâce à cette écriture $AR(\infty)$, on peut réécrire l'équation

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

sous la forme

$$\begin{aligned} & X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p} \\ &= \left(X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i} \right) + \theta_1 \left(X_{t-1} + \sum_{i=1}^{+\infty} \pi_i X_{t-i-1} \right) + \cdots + \theta_q \left(X_{t-q} + \sum_{i=1}^{+\infty} \pi_i X_{t-i-q} \right) \end{aligned}$$

En identifiant les coefficients devant chaque X_t , on obtient bien la formule annoncée.

5.4.3 Liaisons temporelles

On considère un processus $(X_t)_{t \in \mathbb{Z}}$ qui est un $ARMA(p, q)$ d'équation

$$X_t - \sum_{j=1}^p \varphi_j X_{t-j} = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}.$$

- Fonction d'autocorrélation à partir de l'écriture ARMA(p, q) : On considère $X_{t+h} - \sum_{j=1}^p \varphi_j X_{t+h-j} = \varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}$ et on prend la covariance avec X_t ce qui donne

$$\begin{aligned} \gamma_X(h) - \sum_{j=1}^p \varphi_j \gamma_X(h-j) &= \text{Cov} \left(\varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}, X_t \right) \\ &= \text{Cov} \left(\varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}, \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i} \right). \end{aligned}$$

On a aussi que pour $h > q$, $\gamma_X(h) - \sum_{j=1}^p \varphi_j \gamma_X(h-j) = 0$. Pour $0 \leq h \leq q$, on obtient

$$\gamma_X(h) - \sum_{j=1}^p \varphi_j \gamma_X(h-j) = \sigma^2 \left(\sum_{i=0}^{q-h} \theta_{h+i} \psi_i \right) \text{ avec } \psi_0 = 1$$

- Fonction d'autocorrélation avec l'écriture MA(∞) :

En utilisant l'écriture MA(∞) du processus $(X_t)_{t \in \mathbb{Z}}$ on obtient cette fois-ci

$$\gamma_X(h) = \text{Cov} \left(\varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}, \varepsilon_{t+h} + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t+h-i} \right) = \sigma^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+h}.$$

Remarques

- La première expression obtenue sous forme de relation de récurrence, permet de montrer que les autocorrélations simples décroissent de manière exponentielle ou sinusoidale amortie vers 0 avec h . On peut montrer le même genre de résultat pour les autocorrélations partielles.
- On constate, qu'à la différence des cas particuliers des processus AR(p) ou MA(q), il n'existe pas de caractérisation aisée pour les modèles ARMA(p, q). Les autocorrélations simples ou partielles ne s'annulent pas à partir d'un certain rang.

5.5 Vers les processus ARIMA/ SARIMA

5.5.1 ARIMA

Dans la pratique, les processus sont rarement stationnaires. Les modèles **ARIMA** (**Au**-**toRegressive Integrated Moving Average**) sont une extension des processus ARMA aux

processus non stationnaires. Ils sont basés sur l'idée générale suivante essentiellement conçue pour les processus non stationnaires à tendance polynomiale : on peut différencier suffisamment le processus initial afin d'obtenir un processus sans tendance et appliquer un modèle ARMA sur le processus différencié.

Definition 5.6. On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $[\mathbf{ARIMA}(p, d, q)]$ $(p, d, q \in \mathbb{N})$ s'il vérifie l'équation :

$$\Phi(B)(I - B)^d X_t = c + \Theta(B)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

où

$$\begin{aligned} c &\in \mathbb{R} \\ \Phi(B) &= I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \quad (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p, \quad \varphi_p \neq 0 \\ \Theta(B) &= I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (\theta_1, \dots, \theta_q) \in \mathbb{R}^q, \quad \theta_q \neq 0 \\ (\varepsilon_t)_{t \in \mathbb{Z}} &\sim \text{WN}(0, \sigma^2) \end{aligned}$$

Remarques

- On peut montrer que, dans le cas d'un processus $\text{ARIMA}(p, d, q)$, le processus $(I - B)^d X_t$ est asymptotiquement (au sens quand $t \rightarrow +\infty$) un processus $\text{ARMA}(p, q)$.
- On peut aussi établir des représentations $\text{AR}(\infty)$ et $\text{MA}(\infty)$ pour les processus ARIMA .

Par rapport aux caractéristiques évoquées précédemment (voir [?@sec-liaisonsAR](#) et [?@sec-liaisonsMA](#)) sur les autocorrélations et autocorrélations partielles pour les processus AR et MA (voir [?@sec-liaisonsAR](#) et [?@sec-liaisonsMA](#)), on a le résultat suivant.

Proposition 5.18.

- Si la série temporelle suit un $\text{ARIMA}(p, d, 0)$ alors l'ACF et le PACF sur les **données différenciées** vérifient
 - l'ACF décroît exponentiellement ou sinusoidal
 - Un pic significatif au lag p sur le PACF et pas après le lag p
- Si la série temporelle suit un $\text{ARIMA}(0, d, q)$ alors l'ACF et le PACF sur les **données différenciées** vérifient :
 - le PACF est décroissant exponentiellement ou sinusoidal
 - Un pic significatif au lag q pour l'ACF et pas après le lag q

Exemple 5.5. Pour illustrer la Proposition 5.18, la Figure 5.8 montre les autocorrélations et autocorrélations partielles empiriques pour les mesures différenciées une fois du processus

$$(I - 0.1B - 0.2B^2 - 0.6B^3)(I - B)X_t = \varepsilon_t, (\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$$

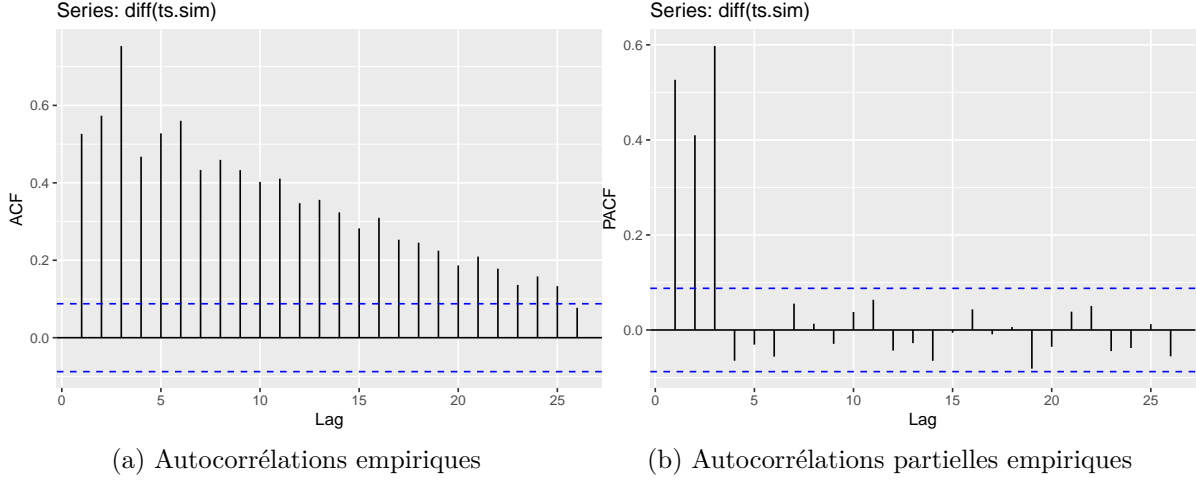


Figure 5.8: Résultats pour le processus ARIMA(3,1,0) $(I - 0.1B - 0.2B^2 - 0.6B^3)(I - B)X_t = \varepsilon_t, (\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$

Pour illustrer la Proposition 5.18, la Figure 5.9 montre les autocorrélations et autocorrélations partielles empiriques pour les mesures différenciées une fois du processus

$$(I - B)X_t = (I - 0.6B - 0.6B^2 - 0.2B^3)\varepsilon_t, (\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$$

5.5.2 Sélection de modèle

Quand on étudie une série temporelle, on peut être amené à hésiter entre plusieurs modèles pour la modéliser au mieux. Supposons que l'on considère une collection de modèles ARIMA $\mathcal{M} = \{\text{mod}_1, \dots, \text{mod}_K\}$. On est dans le cadre classique d'un problème de sélection de modèles. On traite donc ce problème par les étapes suivantes :

- Ajustement de chaque modèle mod_k : Pour chaque modèle fixé, on estime les paramètres (par maximum de vraisemblance le plus souvent)
- On minimise un critère pénalisé pour choisir le meilleur modèle. Par exemple,

- le critère AIC : $\text{AIC}(\text{mod}_k) = -2 \log. \text{ vrais.} + 2(p + q + 1 + \mathbb{1}_{c \neq 0})$
- le critère AIC corrigé : $\text{AICc}(\text{mod}_k) = \text{AIC}(\text{mod}_k) + \frac{2(p+q+1+\mathbb{1}_{c \neq 0})(p+q+2+\mathbb{1}_{c \neq 0})}{n-p-q-2-\mathbb{1}_{c \neq 0}}$

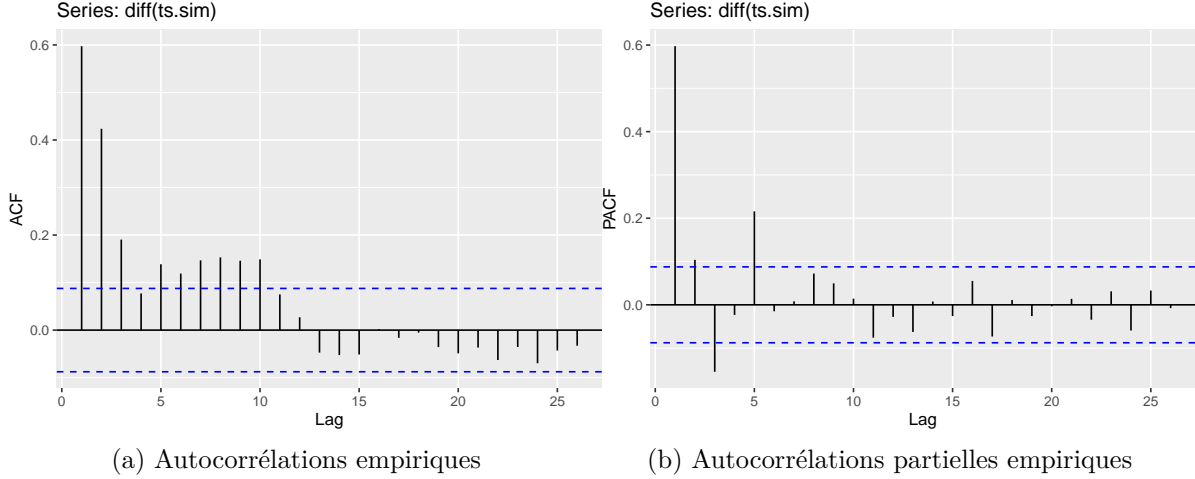


Figure 5.9: Résultats pour le processus ARIMA(0,1,3) $(I - B)X_t = (I - 0.6B - 0.6B^2 - 0.2B^3)\varepsilon_t$, $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{WN}(0, 1)$

– le critère BIC : $\text{BIC}(\text{mod}_k) = -2 \log. \text{ vrais.} + (p + q + 1 + \mathbb{1}_{c \neq 0}) \log(n)$

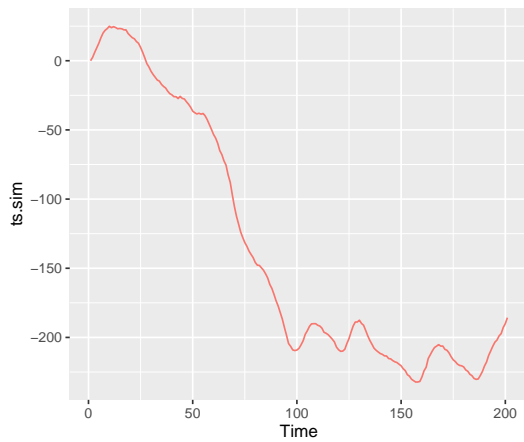
Exemple 5.6. Pour illustrer la sélection de modèle, on a ici simulées une série temporelle selon le modèle ARIMA(1,1,2) suivant

$$(I - 0.8B)(I - B)X_t = (I - 0.3B + 0.6B^2)\varepsilon_t, \quad (\varepsilon_t)_t \sim \text{WN}(0, 1.5)$$

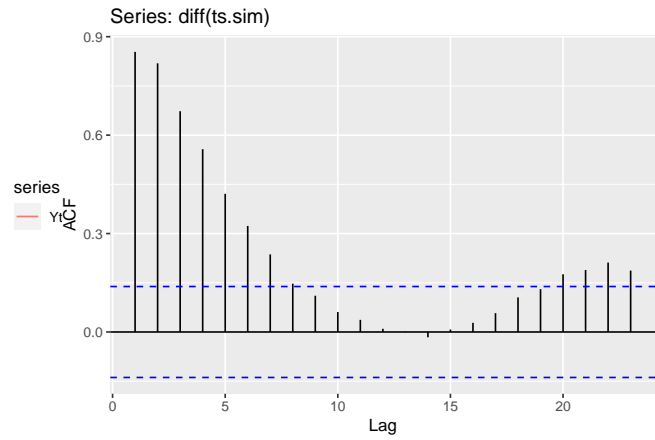
En Figure 5.10, on trace la série temporelle étudiée $(X_t)_{t \in \mathbb{Z}}$ ainsi que les autocorrélations et autocorrélations partielles empiriques de $((I - B)X_t)_{t \in \mathbb{Z}}$.

On considère la collection de modèles suivante $\mathcal{M} = \{\text{ARIMA}(p, 1, q), p \in \{1, 2, 3, 4\}, q \in \{0, 1, 2\}\}$. On calcule les différents critères évoqués précédemment. Au vu des résultats ci-dessous, les trois critères sélectionnent un modèle ARIMA(1,1,2).

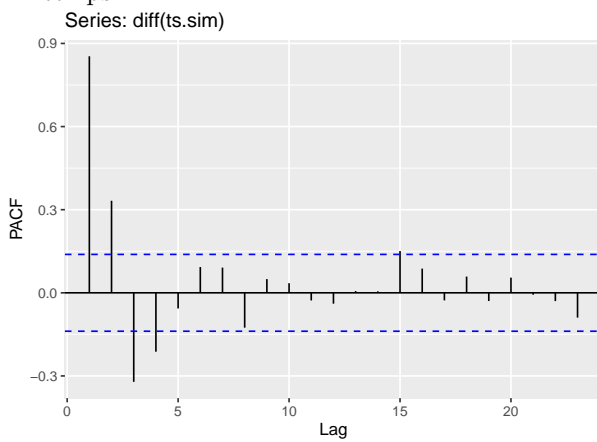
	mod	AIC	AICc	BIC
1	ARIMA(1,1,0)	695.95	696.01	702.55
2	ARIMA(2,1,0)	667.99	668.11	677.88
3	ARIMA(3,1,0)	648.94	649.15	662.13
4	ARIMA(4,1,0)	637.21	637.52	653.70
5	ARIMA(1,1,1)	682.29	682.41	692.19
6	ARIMA(2,1,1)	661.06	661.27	674.26
7	ARIMA(3,1,1)	642.09	642.40	658.58
8	ARIMA(4,1,1)	639.04	639.48	658.83
9	ARIMA(1,1,2)	628.36	628.57	641.56
10	ARIMA(2,1,2)	630.27	630.58	646.76



(a) Série temporelle observée sur les 200 premiers temps



(b) Autocorrélations empiriques



(c) Autocorrélations partielles empiriques

Figure 5.10: Résultats pour une série temporelle simulées selon un modèle ARIMA(1,1,2).

11 ARIMA(3,1,2) 632.25 632.69 652.04
 12 ARIMA(4,1,2) 634.14 634.72 657.22

5.5.3 SARIMA

Les modèles SARIMA, pour Seasonal ARIMA, sont une extension des modèles ARIMA à des séries temporelles présentant une saisonnalité.

Definition 5.7. On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation **SARIMA** $(p, d, q)(P, D, Q)[s]$ s'il vérifie l'équation :

$$(I - B^s)^D \Phi_P(B^s) \Phi_p(B) (I - B)^d X_t = c + \Theta_q(B) \Theta_Q(B^s) \varepsilon_t, \quad \forall t \in \mathbb{Z}$$

où

$$\begin{aligned} c &\in \mathbb{R} \\ \Phi_p(B) &= I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \quad (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p, \quad \varphi_p \neq 0 \\ \Phi_P(B^s) &= I - \phi_1 B^s - \phi_2 (B^s)^2 - \dots - \phi_P (B^s)^P, \quad (\phi_1, \dots, \phi_P) \in \mathbb{R}^P, \quad \phi_P \neq 0 \\ \Theta_q(B) &= I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (\theta_1, \dots, \theta_q) \in \mathbb{R}^q, \quad \theta_q \neq 0 \\ \Theta_Q(B^s) &= I + \vartheta_1 (B^s) + \vartheta_2 (B^s)^2 + \dots + \vartheta_Q (B^s)^Q, \quad (\vartheta_1, \dots, \vartheta_Q) \in \mathbb{R}^Q, \quad \vartheta_Q \neq 0 \\ (\varepsilon_t)_{t \in \mathbb{Z}} &\sim \text{WN}(0, \sigma^2) \end{aligned}$$

En pratique, voici comment on peut procéder pour “identifier” les différents ordres :

- Choix de d , D et s :
 - Allure de la série pour détecter la non-stationnarité et la saisonnalité s
 - Différencier la série successivement jusqu'à stationnarité ($d = 1, 2$ et $D = 0, 1$ suffisent souvent)
- Choix de p , q (P et Q idem sur les lags $k \times s$) : A partir de ACF et pACF sur la série différenciée
 - Cassure de l'ACF au lag q -> MA(q)
 - Cassure pACF au lag p -> AR(p)
 - Aucune cassure -> ARMA
 - Décroissance très lente ou inexistante, il faut peut être encore différencier (mais avec parcimonie)

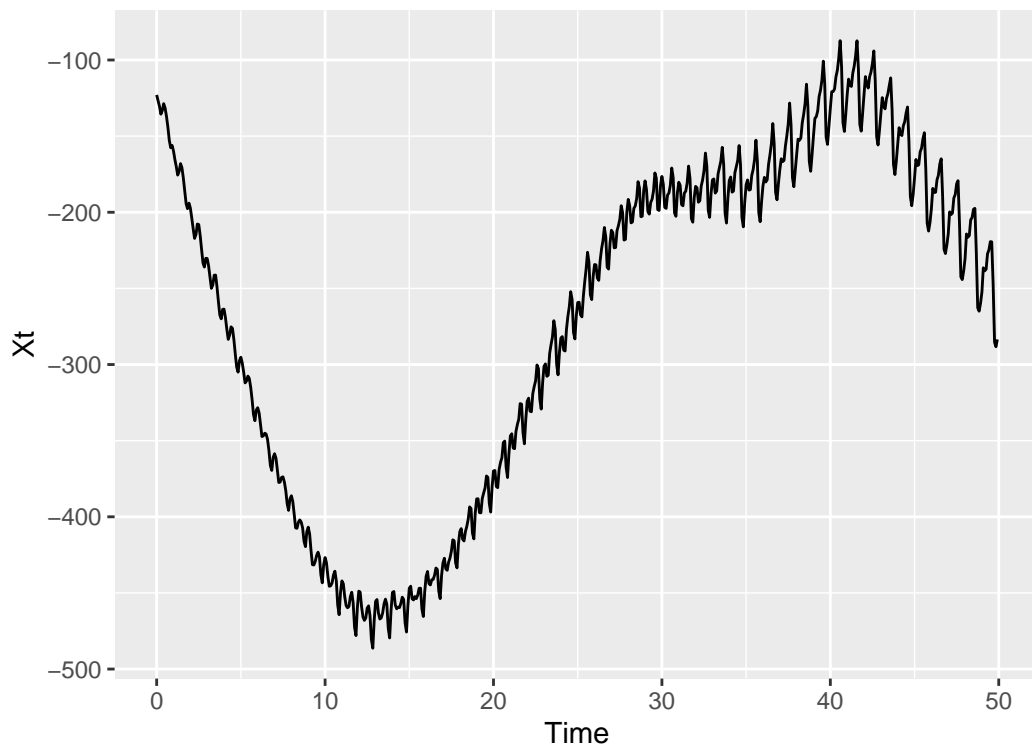


Figure 5.11: Série temporelle simulée selon un modèle SARIMA(0,1,1)(0,1,1)[12]

Exemple 5.7. Dans cet exemple, on simule une série temporelle selon un modèle SARIMA(0,1,1)(0,1,1)[12]

$$(I - B^{12})(I - B)X_t = (I + 0.8B)(I + 0.4B^{12})\varepsilon_t$$

La série est représentée en Figure 5.11.

La Figure 5.12 permet de partir sur une saisonnalité par mois pour la série temporelle ($s = 12$).

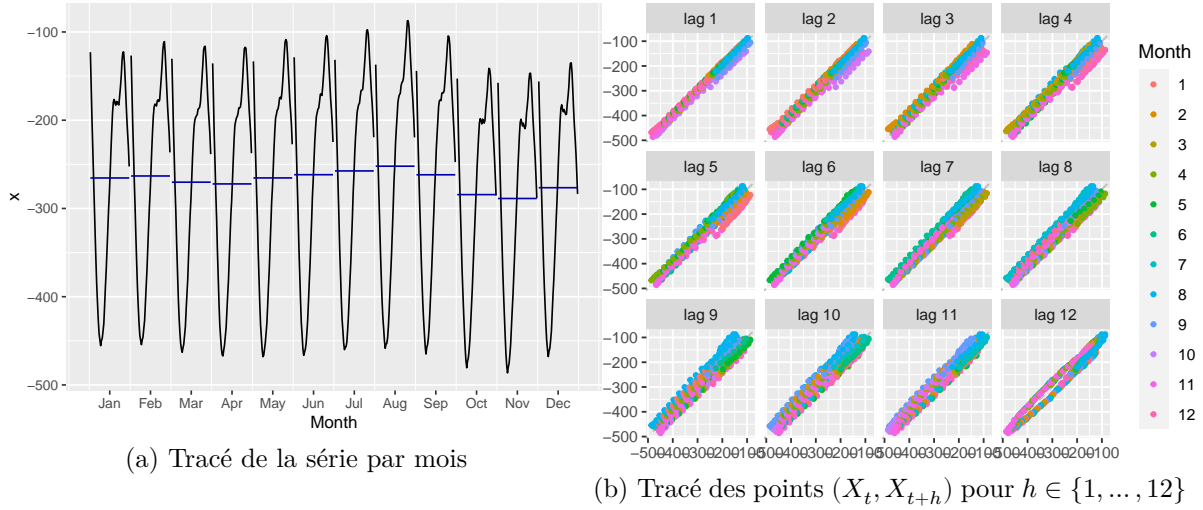


Figure 5.12: Pour l'étude de la saisonnalité.

On trace ensuite les autocorrélations et autocorrélations empiriques pour la série (X_t) (voir Figure 5.13)).

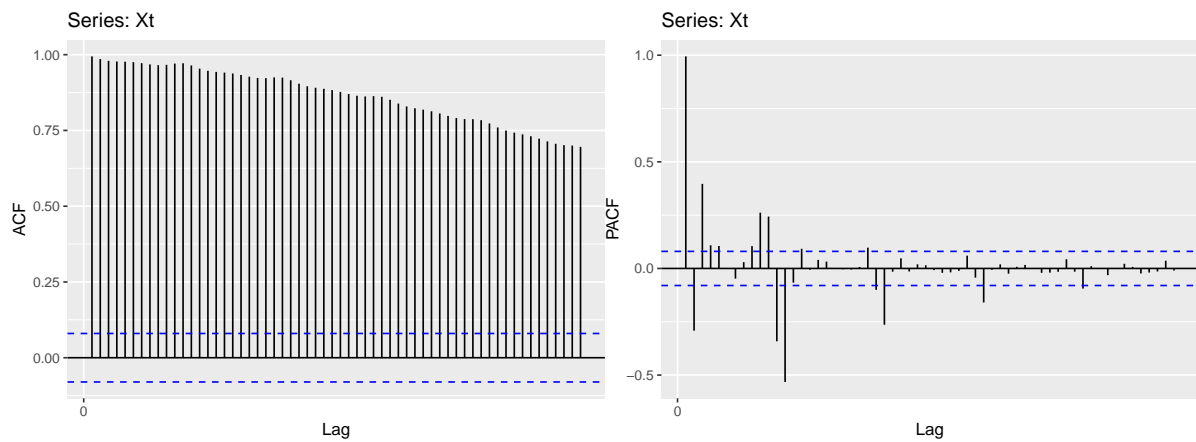
La Figure 5.14 (resp. Figure 5.15) représente la série temporelle différenciée $(I - B)X_t$ (resp. $(I - B^{12})X_t$)

On trace ensuite en Figure 5.16, les autocorrélations et autocorrélations partielles empiriques de la série

$$Y_t = (I - B^{12})(I - B)X_t$$

On décide alors d'ajuster sur la série $(X_t)_{t \in \mathbb{Z}}$ un modèle SARIMA(0,1,1)(0,1,1)[12]. L'estimation des paramètres est donnée ici à l'aide de la fonction `arima()` et on teste ensuite la blancheur des résidus avec la fonction `Box.test()` :

```
res<-arima(Xt,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))
res
```



(a) Autocorrélations empiriques

(b) Autocorrélations partielles empiriques

Figure 5.13: Autocorrélations et autocorrélations partielles de la série $(X_t)_{t \in \mathbb{Z}}$.

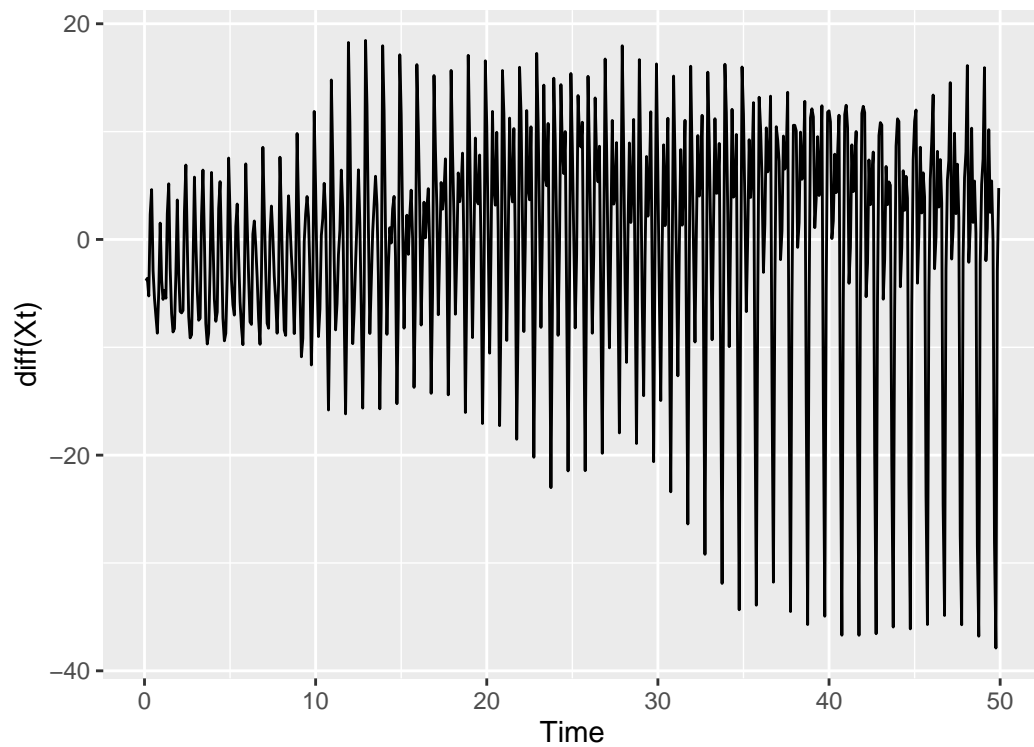


Figure 5.14: Tracé de la série temporelle différenciée $((I - B)X_t)_{t \in \mathbb{Z}}$.

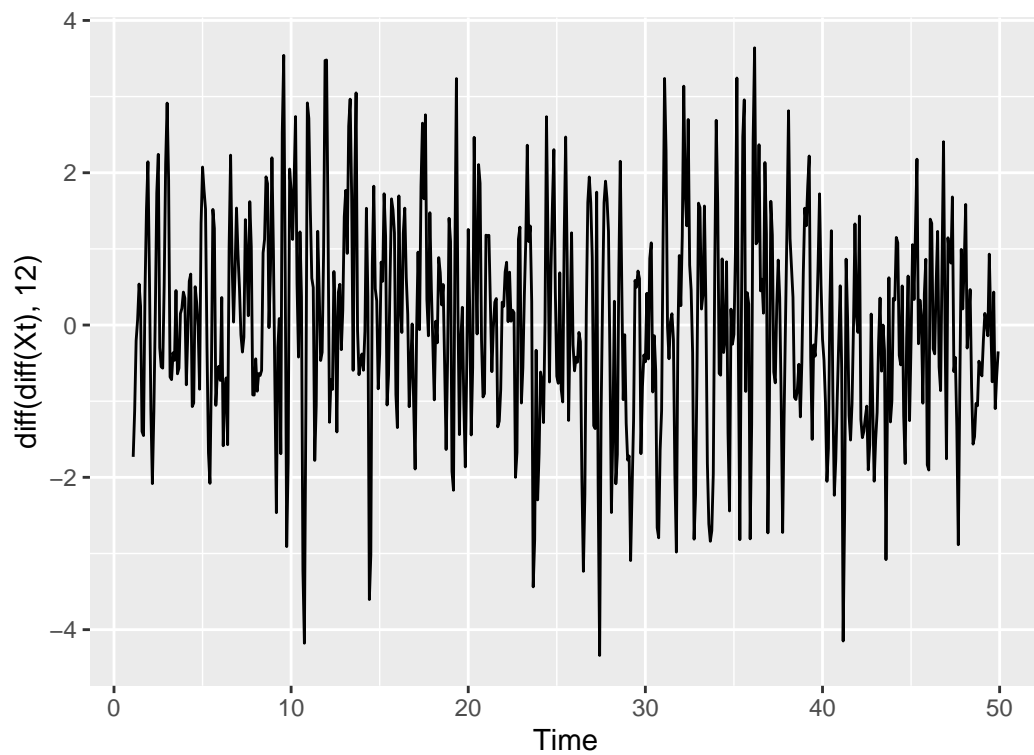
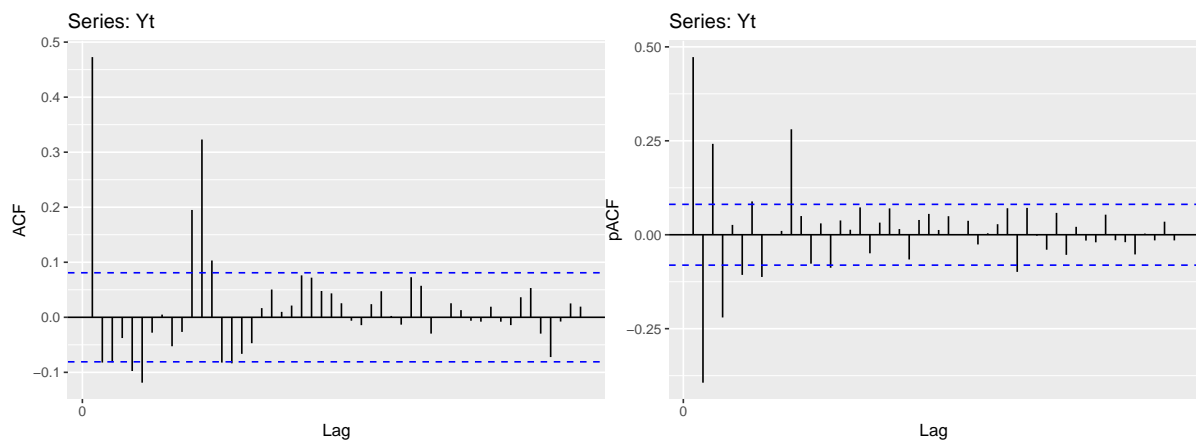


Figure 5.15: Tracé de la série temporelle différenciée $((I - B^{12})X_t)_{t \in \mathbb{Z}}$.



(a) Autocorrélations empiriques

(b) Autocorrélations partielles empiriques

Figure 5.16: Autocorrélations et autocorrélations partielles de la série $Y_t = (I - B^{12})(I - B)X_t$.

```
Call:
arima(x = Xt, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
          ma1      sma1
      0.8236  0.3584
s.e.  0.0220  0.0391

sigma^2 estimated as 0.9851:  log likelihood = -829.86,  aic = 1665.72
```

```
Box.test(res$residuals)
```

Box-Pierce test

```
data:  res$residuals
X-squared = 0.67745, df = 1, p-value = 0.4105
```

Pour finir ce chapitre, on étudie un jeu de données réelles concernant la pollution en oxyde d'azote à Londres.

Exemple 5.8 (Pollution en oxyde d'azote à Londres). La Figure 5.17 représente la mesure de pollution en oxyde d'azote à Londres par heure entre le 1/01/98 et le 21/01/98. On trace les autocorrélations empiriques (voir en Figure 5.18) et les autocorrélations partielles empiriques (voir en Figure 5.19) de cette série temporelle (X_t).

On analyse ensuite les autocorrélations et autocorrélations partielles empiriques de la série différenciée $Y_t = (I - B^{24})X_t$ (voir Figure 5.20). On peut alors considérer des modèles SARIMA avec $0 \leq q \leq 12$, $0 \leq p \leq 2$, $0 \leq Q \leq 2$, $0 \leq P \leq 1$.

On met donc en place une sélection de modèles avec le critère AIC à l'aide de la fonction `auto.arima()`. On retient un modèle SARIMA(0, 1, 1)(0, 0, 1)[24] et on contrôle la blancheur des résidus obtenus.

```
set.seed(1234)
res<-auto.arima(tta,ic=c("aic"),
               max.p=2,max.q=12,
               max.P=1,max.Q=2,seasonal=TRUE)

res
```

```
Series: tta
ARIMA(0,1,1)(0,0,1)[24] with drift
```

```
Coefficients:
```

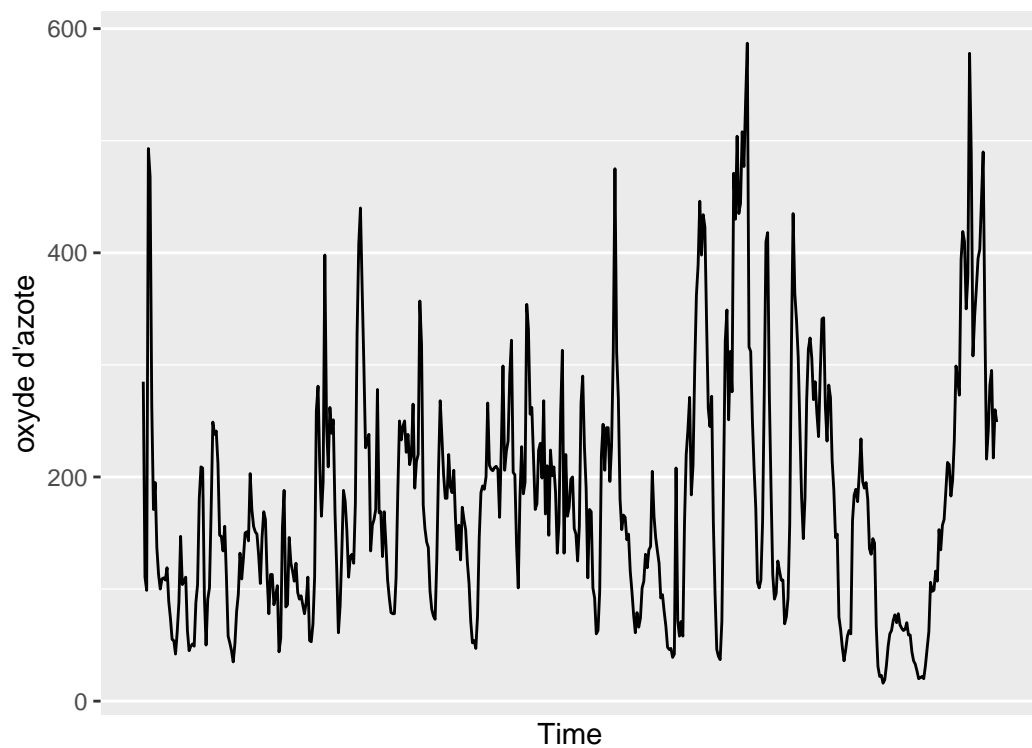



Figure 5.17: Mesure de l'oxyde d'azote dans l'air à Londres par heure entre le 1/01/98 et le 21/01/98.

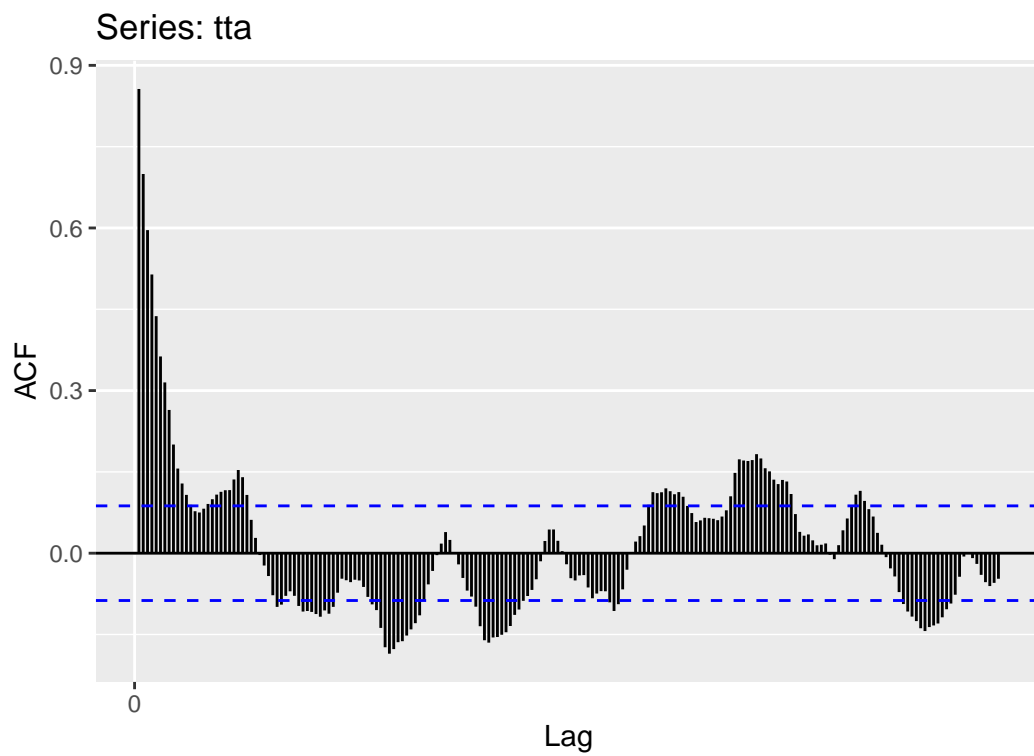


Figure 5.18: Autocorrélations de la série X_t

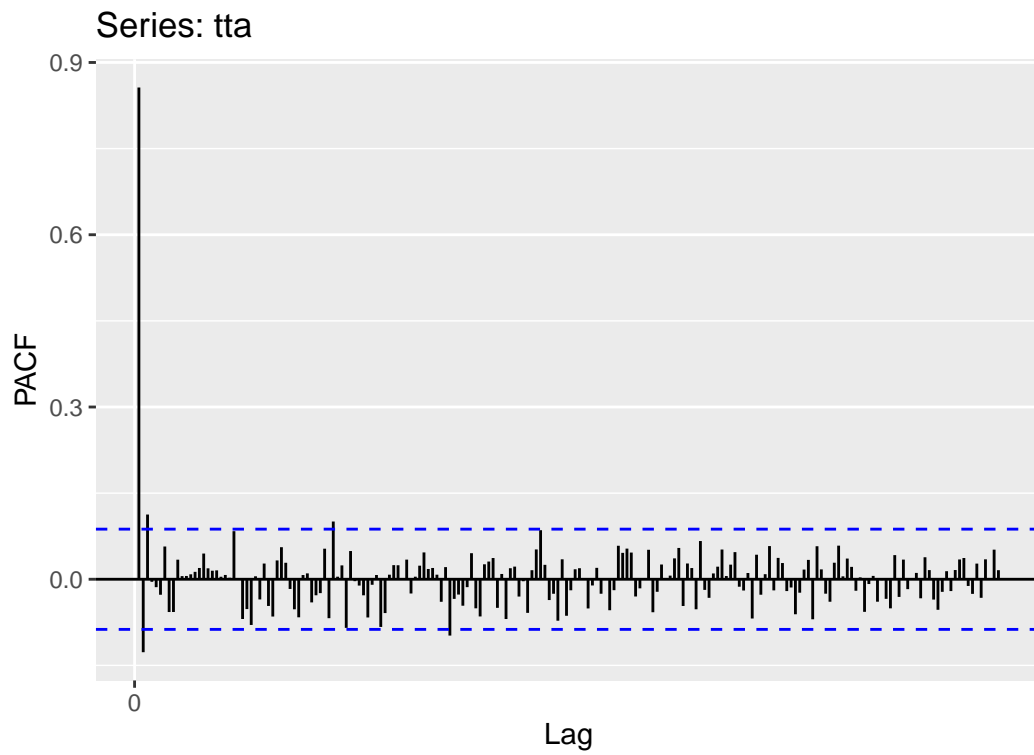


Figure 5.19: Autocorrélations partielles de la série X_t

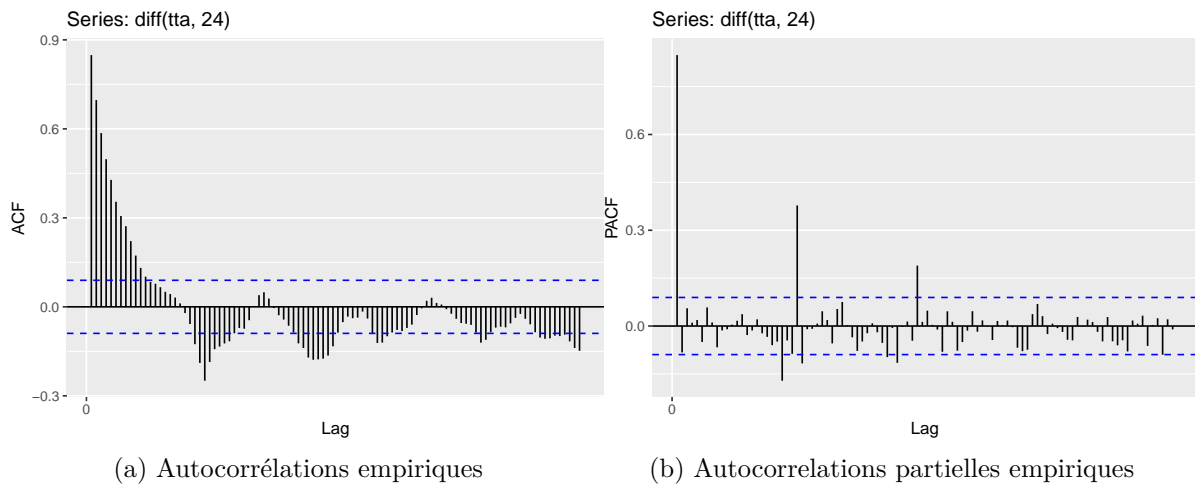


Figure 5.20: Autocorrélations et autocorrélations partielles de la série $Y_t = (I - B^{24})X_t$.

```

          ma1      sma1      drift
0.0784  0.1071  -0.1602
s.e.  0.0549  0.0463   3.0371

sigma^2 = 3305:  log likelihood = -2750.34
AIC=5508.68   AICc=5508.76   BIC=5525.56

```

```
Box.test(res$residuals)
```

Box-Pierce test

```

data:  res$residuals
X-squared = 0.10787, df = 1, p-value = 0.7426

```

On représente en Figure [5.21](#) la série temporelle ajustée par ce modèle par rapport à la série temporelle observée.

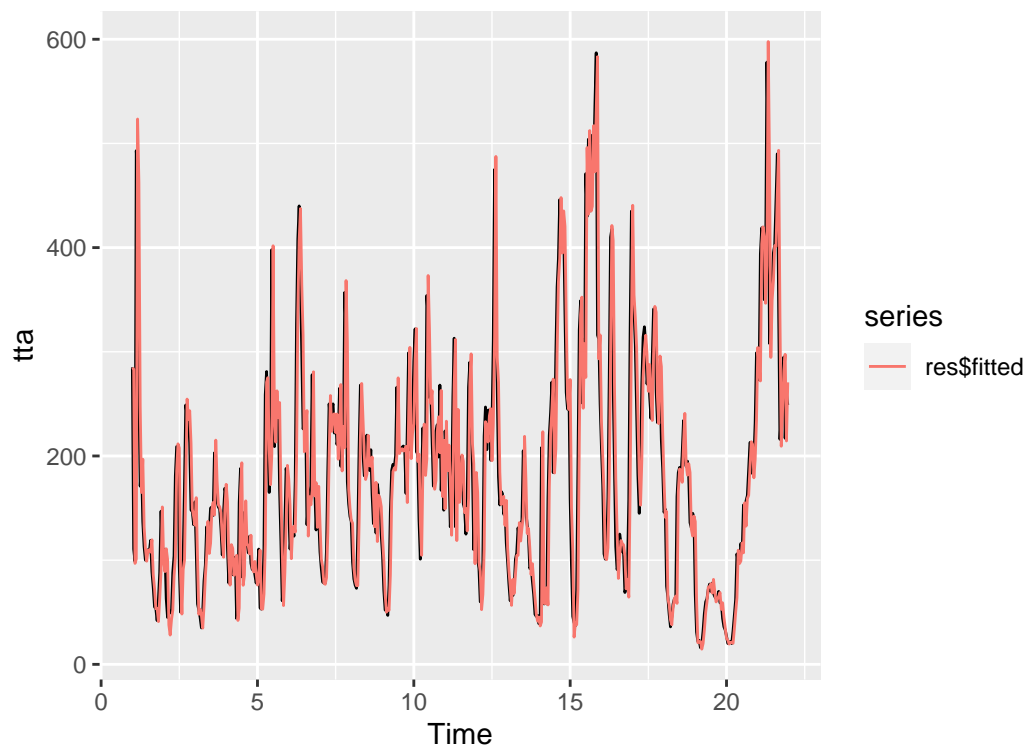


Figure 5.21: Série temporelle estimée avec un modèle SARIMA(0,1,1)(0,0,1)(24) par rapport à la série temporelle observée.

References

- Aragon, Yves. 2016. *Séries Temporelles Avec r*. EDP sciences.
- Brockwell, Peter J, and Richard A Davis. 2002. *Introduction to Time Series and Forecasting*. Springer.
- . 2009. *Time Series: Theory and Methods*. Springer science & business media.
- Dauxois, Jean-Yves. 2020. “Introduction à l’étude Des Séries Temporelles.” Polycopié cours INSA Toulouse.