

Mathématiques et Informatique de la Décision et des Organisations

Introduction aux séries temporelles

Master 1 Mathématiques Appliquées

Notes de cours

Année 2017/2018



Quartier d'affaires de la Défense et bois de Boulogne
Vus du bureau B518-bis de l'Université Paris-Dauphine

DAUPHINE
UNIVERSITÉ PARIS

MIDO
MATHÉMATIQUES ET INFORMATIQUE
DE LA DÉCISION ET DES ORGANISATIONS

Version électronique composée avec L^AT_EX le 22 novembre 2017

-
- Auteurs des notes :
 - Chafaï, Djalil (enseignant, Paris-Dauphine, 2013–2017)
 - Lévy-Leduc, Céline (enseignante, Paris-Dauphine, 2010-2012)
 - Roche, Angelina (enseignante, Paris-Dauphine, 2017–, modifications mineures)
 - Chasseurs de coquilles :
 - Akrim, Anass (étudiant, Paris-Dauphine, 2016–2017)
 - Azouzi, Youssef (enseignant, Tunis-Dauphine, 2015–2016)
 - Benaych-Georges, Florent (enseignant, Paris-Descartes, 2014–2015)
 - Bobo, Grégoire (étudiant, Paris-Dauphine, 2016-2017)
 - Bonniau, Donatien (étudiant, Paris-Dauphine, 2016–2017)
 - Eremeeva, Evgeniia (étudiante, Paris-Dauphine, 2014–2015)
 - Georget, Maxime (étudiant, Paris-Dauphine, 2016-2017)
 - Ivanoff, Stéphane (enseignant, Paris-Dauphine, 2013–2015)
 - Lee, Chun-Hsien (étudiant, Paris-Dauphine, 2015–2016)
 - Macchi, Emmanuel (étudiant, Paris-Dauphine, 2014–2015)
 - Malouche, Dhafer (enseignant, Tunis-Dauphine, 2014–2015)
 - Munier, Robin (étudiant, Paris-Dauphine, 2016–2017)
 - Murcy, Pierre-Germain (étudiant, Paris-Dauphine, 2017–2018)
 - Pagnard, Camille (enseignant, Paris-Dauphine, 2014–)
 - Razafindrabary Iarivonjy, Bisous Mirenty Tsiky (étudiante Paris-Dauphine, 2017–2018)
 - Stern, Mathilde (étudiante, Paris-Dauphine, 2016–2017)
 - Tan, Xiaolu (enseignant, Paris-Dauphine, 2016–2017)

Avant-propos

Le cours de Master 1 d'introduction aux séries temporelles est découpé en 13 séances de 90 minutes. Le prérequis est un niveau de licence en mathématiques, notamment en analyse, probabilités, et statistique. L'évaluation consiste en un examen partiel en novembre et un examen final en janvier, avec pour formule $\max(f, \frac{2}{5}p + \frac{3}{5}f)$. Pour réussir aux examens, il faut d'abord comprendre les notes de cours puis s'entraîner sur les exercices de travaux dirigés, les annales d'examens, et les démonstrations du cours. Ces notes de cours, librement disponibles sur Internet, sont inspirées de celles de Céline Lévy-Leduc datant de l'année universitaire 2011/2012. Le sujet des séries temporelles est plutôt plaisant. Lié au concret, il pose de réelles questions sur les plans méthodologiques et philosophiques¹. Lié à l'abstrait, il donne l'occasion de mettre en œuvre quelques concepts de la théorie du signal, de l'analyse harmonique, des probabilités et statistique, de l'algèbre linéaire, etc. En toile de fond, utilitarisme, mythe du quantitatif, et «bigdata».

Notre Master 1 de mathématiques a pour but de former des ingénieurs, et dans une moindre mesure des chercheurs, adaptés au monde d'aujourd'hui. Les ingénieurs et les chercheurs ne cessent d'apprendre des concepts et des techniques, tout au long de la vie professionnelle². Ces notes de cours fournissent des bases incontournables de la modélisation des séries temporelles, que tout ingénieur ou chercheur mathématicien devrait connaître. Elles sont conçues pour être à la fois accessibles et exigeantes intellectuellement. La formation initiale des ingénieurs et des chercheurs doit mettre l'accent sur les concepts plutôt que sur les techniques. Ceux qui devront ensuite, dans leur vie professionnelle, traiter des données, pourront apprendre les techniques appropriées le moment voulu, en s'appuyant sur leurs connaissances conceptuelles. On trouvera par exemple dans [Ar, C] deux très bons cours orientés vers des techniques concrètes de traitement des données³.

1. À ce sujet, on peut penser par exemple au phénomène du réchauffement climatique.

2. C'est mal connaître le monde de l'ingénieur que de le réduire à celui du technicien.

3. Le niveau en mathématiques des élèves de Master est variable d'un établissement à l'autre.

Table des matières

Table des matières	v
1 Introduction (x2)	1
1.1 Tendance, saisonnalité, bruit, opérateur retard	1
1.2 Processus stationnaires	7
1.3 Suites récurrentes aléatoires et processus ARMA	16
1.4 Autocovariance et matrices de Toeplitz	18
1.5 Compléments	20
2 Filtrage linéaire (x2.5)	25
2.1 Motivation	25
2.2 Théorème de filtrage	26
2.3 Composition et convolution	29
2.4 Compléments	36
3 Processus ARMA (x1.5)	37
3.1 Processus MA et processus AR	38
3.2 Résolution de l'équation ARMA	40
3.3 ARMA causal et inversible	42
3.4 Autocovariance d'un ARMA	47
3.5 Compléments	48
4 Analyse spectrale (x2)	53
4.1 Coefficients de Fourier	53
4.2 Théorème de Herglotz	56
4.3 Mesure spectrale d'un processus stationnaire	57
4.4 Compléments	62
5 Prédiction linéaire (x3)	67
5.1 Prédicteur linéaire optimal	67
5.2 Équations de Yule-Walker	69
5.3 Algorithme de Cholesky	73
5.4 Algorithme de Levinson-Durbin	76
5.5 Algorithme de Gram-Schmidt (ou des innovations)	79
5.6 Compléments	81

6	Estimation (x1)	83
6.1	Estimation de la moyenne	83
6.2	Estimation de l'autocovariance	87
6.3	Application aux tests	87
A	Rappels d'analyse et probabilités	89
A.1	Espaces de suites	89
A.2	Espaces de fonctions	89
A.3	Covariance	91
A.4	Vecteur moyenne et matrice de covariance	91
	Bibliographie	93

Plan du cours. Le chapitre 1 introduit le concept de série temporelle, sa modélisation stochastique, les notions de tendance, de saisonnalité, de processus stationnaire, d'autocovariance et de bruit blanc. Le chapitre 2 est consacré au filtrage linéaire des processus stationnaires, un concept fondamental pour résoudre les équations linéaires du type ARMA. Le chapitre 3 est consacré à l'étude des processus ARMA, qui constituent la classe la plus courante de processus stationnaires. Le chapitre 4 est consacré à l'analyse spectrale des processus stationnaires, éclairant d'une lumière toute spectrale les notions d'autocovariance, de filtrage, et de causalité des processus ARMA. Le chapitre 5 est consacré à la prédiction linéaire pour les processus stationnaires, en utilisant l'autocovariance. Le chapitre 6 est consacré à l'estimation de la moyenne et de l'autocovariance. L'annexe ?? propose des exemples concrets, et l'annexe A des rappels d'analyse et de probabilités. Une bibliographie et un index figurent en fin de document. Les nombres entre parenthèse à la fin de l'intitulé des chapitres indiquent le nombre approximatif de séances de cours.

Introduction (x2)

Mots-clés : série temporelle ; modèle additif ; tendance ; saisonnalité ; stationnarité ; autocovariance ; bruit blanc ; opérateur retard ; opérateur différence.

Une **série temporelle** ou **série chronologique** est une suite d'observations d'un phénomène physique¹ faites au cours du temps : consommation d'électricité, cours du pétrole, population française, rythme cardiaque, relevé d'un sismographe, trafic Internet, ventes de téléphones mobiles, hauteurs des crues du nil, température des océans, concentration en dioxyde de carbone de l'atmosphère, taux de glucose dans le sang, côte de popularité du président, etc. Il s'agit d'une suite finie² de valeurs réelles³, indicées par un temps continu ou discret régulier, typiquement d'un signal échantillonné à une fréquence fixe. Les figures 1.2 et 1.3 donnent des représentations graphiques de quelques séries temporelles, fabriquées grâce au programme pour le logiciel GNU-R donné par la figure 1.1⁴.

L'étude des séries temporelles poursuit plusieurs buts pratiques :

- une meilleure compréhension du phénomène physique représenté par la série ;
- une représentation simplifiée par un modèle stochastique ;
- une prédiction du futur de la série à partir de la partie observée.

1.1 Tendances, saisonnalité, bruit, opérateur retard

Une idée serait de tracer le graphe de la série temporelle $t \mapsto y_t$, puis de déterminer une famille de fonctions qui ont la même allure, puis enfin de déterminer la meilleure fonction en minimisant un critère d'ajustement comme les moindres carrés par exemple, avec pénalisation du critère optimisé par la complexité de la fonction utilisée. Dans cette approche, la famille de fonctions choisie constitue un modèle. On dit à ce propos : *tous les modèles sont faux, certains sont utiles*. L'incorporation de bruit dans cette approche conduit aux modèles analysés dans ce cours. Plus précisément, on s'intéresse à des modèles additifs stochastiques du type

$$x_t = d_t + z_t$$

où $t \mapsto d_t$ est une fonction déterministe et où z_t est un bruit aléatoire. La fonction déterministe dépend typiquement d'un nombre réduit de paramètres qu'on cherche à estimer,

1. Physique au sens large, englobant la nature, par opposition à Mathématique ou Informatique.

2. Les données réelles sont des suites finies. Les modèles mathématiques «vont jusqu'à l'infini».

3. Rien n'empêche d'imaginer des valeurs complexes voire même vectorielles bien entendu.

4. Sous Debian GNU/Linux : `sudo apt-get install r-base`. L'environnement de développement RStudio est disponible pour Debian, cf. <https://www.rstudio.com/products/rstudio/download/>. Plus généralement, on pourra consulter <https://cran.r-project.org/web/views/TimeSeries.html>

```
# Inspiré du livre "Séries temporelles avec R" de Yves Aragon.
# install.packages("datasets") # déjà installé en général !
require(datasets)
#
data(Nile)
plot.ts(Nile,xlab='année',ylab='Nil')
data(uspop)
plot.ts(uspop,xlab='année',ylab='USpop')
#
data("AirPassengers")
plot.ts(AirPassengers,xlab='temps (mensuel, de 1949 à 1961)',ylab='passagers',las=1)
polygon(c(1958,1960,1960,1958),c(200,200,600,600),lty=2)
deb=c(1958,1); fin=c(1960,12)
plot.ts(window(AirPassengers,start=deb,end=fin),xlab='année',ylab='passagers',las=1)
#
plot.ts(diff(AirPassengers),xlab="temps (mensuel)",ylab="log(passagers)")
plot.ts(diff(log(AirPassengers)),xlab="temps (mensuel)",ylab="log(diff(passagers))")
```

FIGURE 1.1 – Code GNU-R pour générer les figures 1.2 et 1.3.

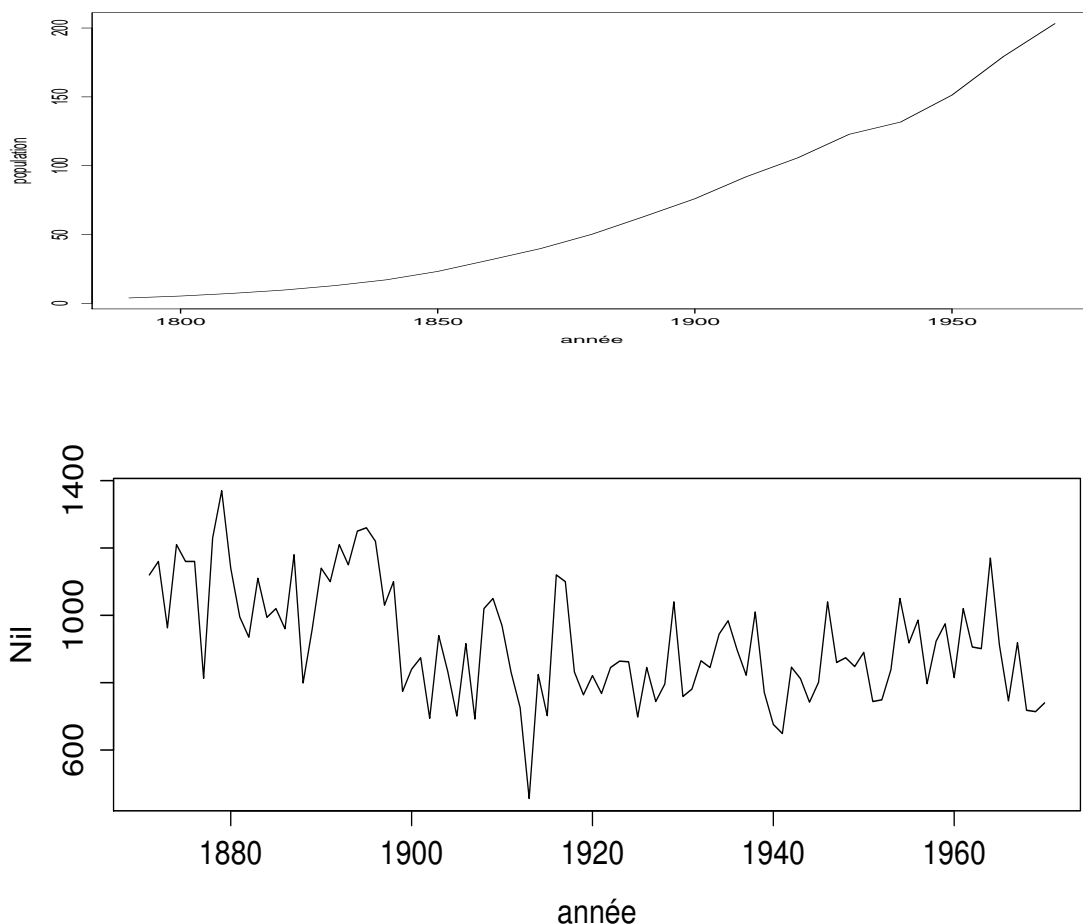
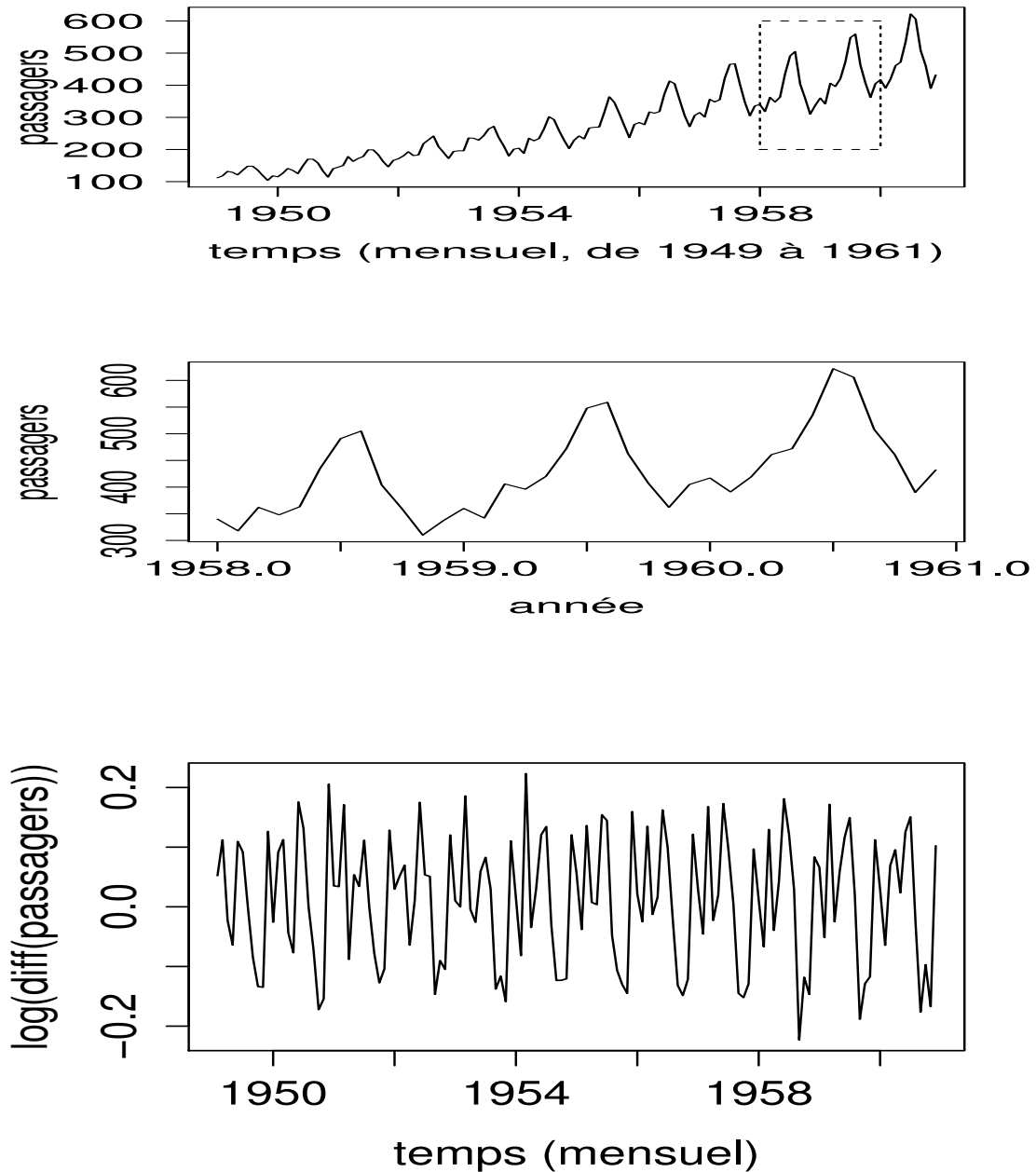


FIGURE 1.2 – Population des États-Unis (haut) et débit du Nil à Assouan (bas).

Ce type de graphique avec le temps en abscisse et la série en ordonnée est appelé **chronogramme**. Ici le temps explique bien l'allure des deux séries.



La série d'origine figure dans le graphique du haut. Elle comporte une saisonnalité marquée avec une variabilité croissante au cours du temps. Le graphique du bas figure la même série après une différenciation, puis une transformation logarithmique qui a pour effet d'en réduire l'hétérogénéité de la variance (**hétéroscédasticité**). Le graphique du milieu est un zoom sur la zone rectangulaire en pointillés du graphique du haut. Il suggère une analyse de la saisonnalité de la série pour comparer les mois par exemple.

FIGURE 1.3 – Nombre de passagers aériens.

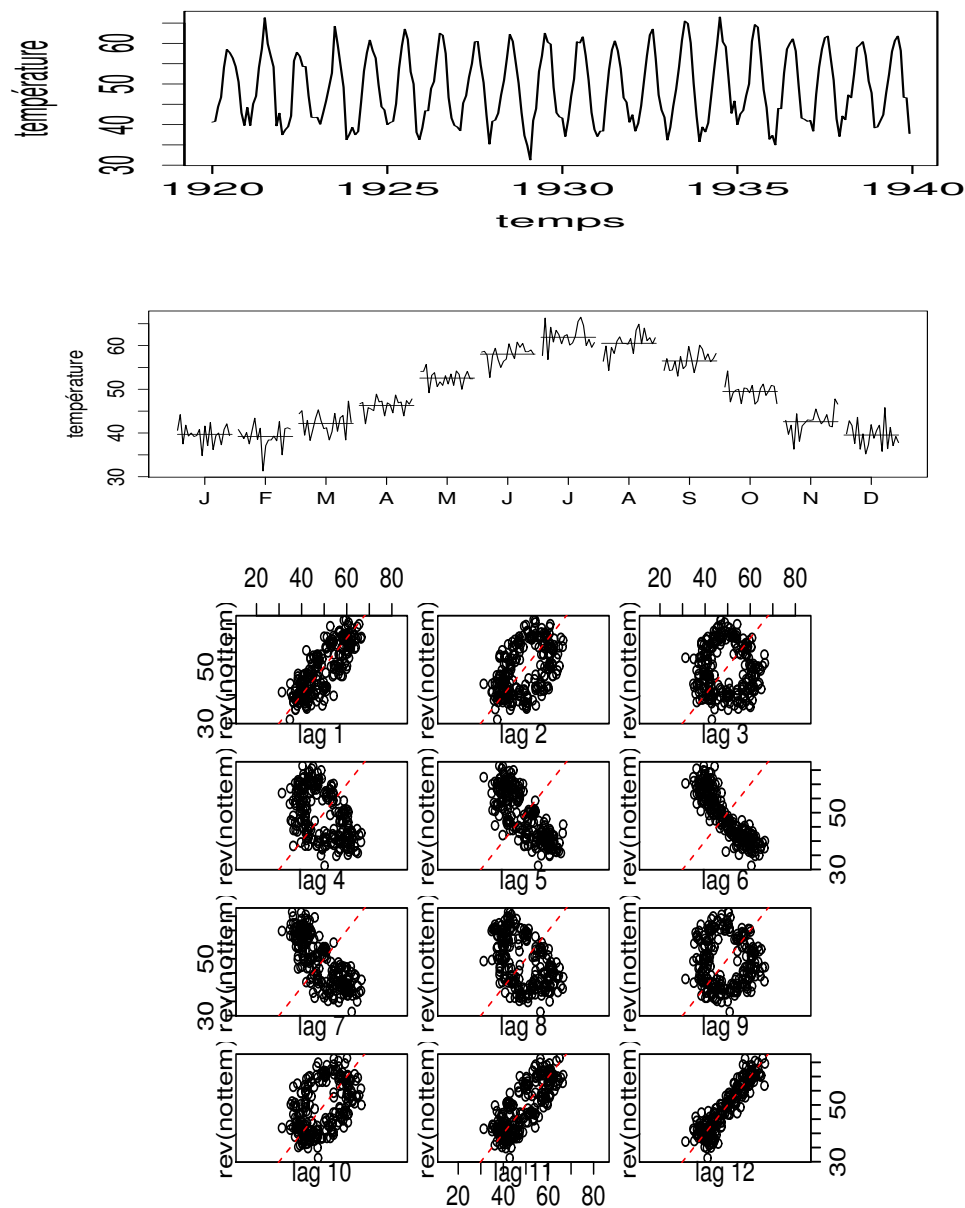


FIGURE 1.4 – Températures mensuelles à Nottingham Castle.

Chronogramme (haut), «month plot» (milieu), et «lag plot» (bas) de la série, réalisés avec le code GNU-R ci-dessous, ou comment révéler la saisonnalité de cette série.

```
# Inspiré du livre "Séries temporelles avec R" de Yves Aragon.
data(nottem)
plot.ts(nottem,xlab="temps",ylab="température")
lag.plot(rev(nottem),12,layout=c(4,3),do.lines=FALSE,diag.col="red",col.main="blue")
monthplot(nottem,ylab="température",main="",cex.main=1)
```

tandis que le bruit est supposé être **stationnaire**, c'est-à-dire que ses caractéristiques statistiques comme son espérance et sa covariance ne varient pas au cours du temps. Dans cette approche stochastique, $(x_t)_{t \in T}$ est modélisée par une trajectoire (ω fixé!) d'un processus stochastique $(X_t)_{t \in T}$.

Une idée naturelle est de rechercher d'abord, au vu des données, la meilleure fonction déterministe d_t telle que le **résidu** $x_t - d_t$ soit raisonnablement stationnaire. On est parfois conduit à transformer les données pour stabiliser la variance, ou encore à les découper en plages temporelles en fonction de ruptures structurelles.

La stationnarité de la partie aléatoire est utile pour effectuer une prévision par translation. Si par exemple on observe Z_1, \dots, Z_t alors on peut estimer la matrice de covariance de ce vecteur aléatoire, c'est-à-dire la structure L^2 du vecteur aléatoire, qui est aussi par stationnarité une estimation de la structure de covariance de Z_{t+1}, \dots, Z_{t+1} , ce qui permet de prédire le futur Z_{t+1} en utilisant une projection par moindres carrés sur les observations Z_2, \dots, Z_t .

Les séries temporelles stationnaires sont rares. Dans la pratique, on tente de s'y ramener en effectuant des transformations qui éliminent par exemple une composante structurelle déterministe. Le modèle additif le plus simple est de la forme

$$X_t = \underbrace{m_t + s_t}_{\text{partie déterministe}} + \underbrace{Z_t}_{\text{partie aléatoire}}$$

où m_t , s_t , et Z_t jouent les rôles suivants :

1. **Tendance.** La fonction $t \mapsto m_t$ est une fonction qui varie lentement, appelée **tendance**. Exemple : tendance polynomiale $m_t = a_0 + a_1 t + \dots + a_d t^d$ (linéaire si $d = 1$). Concrètement, une analyse graphique suggère souvent une valeur de d , ce qui permet ensuite d'estimer a par **moindres carrés** :

$$\hat{a} = \arg \min_a \sum_{t=1}^n (x_t - m_t)^2.$$

Pour $d = 2$ par exemple, et des temps d'observations t_1, \dots, t_n , on pose

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}$$

ce qui donne l'**estimation par moindre carrés**⁵

$$\hat{a} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \arg \min_a \|X - Aa\|_2^2 = (A^\top A)^{-1} A^\top X,$$

avec pour **résidus d'estimation**

$$x - A\hat{a} = x - A(A^\top A)^{-1} A^\top X;$$

5. Revient à minimiser la fonction quadratique convexe $\varphi(a) = \|X - Aa\|_2^2$. On a $\varphi(a+h) - \varphi(a) = 2\langle Aa, Ah \rangle - 2\langle X, Ah \rangle + \|Ah\|_2^2 = \langle \nabla \varphi(a), h \rangle + o(\|h\|)$, d'où $\nabla \varphi(a) = 2A^\top Aa - 2A^\top X$. Or $A^\top A$ est inversible car A est de rang plein ($t_i \neq t_j$ si $i \neq j$), d'où $\nabla \varphi(a) = 0$ ssi $a = (A^\top A)^{-1} A^\top X$.

2. **Saisonnalité.** La fonction $t \mapsto s_t$ est une fonction périodique, appelée **composante saisonnière** (journalière, hebdomadaire, mensuelle, annuelle, etc). Un exemple est donné par une fonction trigonométrique de la forme

$$s_t = a_0 + \sum_{j=1}^k (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

où les a_j et b_j sont inconnus et où les λ_i et λ_j sont des multiples entiers connus de $2\pi/d$. On devine d et les λ_i et λ_j puis on estime les a_i, b_j par moindres carrés. Par exemple, pour $k = 1$ et $\lambda_1 = 2\pi/12$ on utilise

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} = (A^\top A)^{-1} A^\top X \quad \text{avec} \quad A = \begin{pmatrix} 1 & \cos(\lambda_1 t_1) & \sin(\lambda_1 t_1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(\lambda_1 t_n) & \sin(\lambda_1 t_n) \end{pmatrix};$$

3. **Bruit.** Le processus $(Z_t)_{t \in \mathbb{Z}}$ est un bruit qu'on espère **stationnaire**.

Dans la pratique, voici les étapes qu'on peut tenter de suivre :

1. Tracer la série temporelle et rechercher graphiquement la présence d'une tendance et d'une composante saisonnière (la nature des données peut aider) ;
2. Modéliser la tendance et la composante saisonnière (on peut s'aider des opérateurs Δ et Δ_d définis ci-après pour aller vite). Dans les moindres carrés, à valeur égale du critère minimisé, on préfère le modèle le moins complexe ;
3. Modéliser les résidus (en utilisant l'autocovariance définie plus loin).

Ci-dessous, un **opérateur** est une application définie sur l'ensemble \mathcal{X} des processus.

Définition 1.1 : Opérateurs retard et différence

L'**opérateur retard** B décale le processus d'un cran vers le passé :

$$\forall t \in \mathbb{Z}, \quad (BX)_t = X_{t-1}.$$

L'**opérateur différence** Δ est défini par ^a

$$\forall t \in \mathbb{Z}, \quad (\Delta X)_t = X_t - X_{t-1} = (1 - B)X_t.$$

Pour tout entier $d \geq 1$, l'**opérateur différence saisonnier** Δ_d est défini par

$$\forall t \in \mathbb{Z}, \quad (\Delta_d X)_t = X_t - X_{t-d} = ((1 - B^d)X)_t.$$

^a. Ici 1 ou I désigne l'opérateur identité.

Notons que $\Delta = \Delta_1$. L'opérateur différence Δ agit comme une dérivation par rapport au paramètre t . Il permet d'éliminer une tendance linéaire :

$$\Delta(a + bt + Z_t) = b + Z_t - Z_{t-1}.$$

La partie constante a est éliminée, la partie linéaire bt est transformée en la constante b , tandis que la partie aléatoire Z_t subit une transformation linéaire (préserve l'éventuel caractère stationnaire). Plus généralement, l'**opérateur différence d'ordre n**

$$\Delta^n = \underbrace{\Delta \cdots \Delta}_{n\text{-fois}} = (1 - B)^n = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} B^{n-k},$$

(avec la convention $B^0 = 1$) élimine les tendances polynomiales de degré $< n$, par exemple

$$\Delta^2(a + bt + ct^2 + Z_t) = 2c + (Z_t - 2Z_{t-1} + Z_{t-2}).$$

Pour éliminer une composante saisonnière de période d , on utilise l'**opérateur différence saisonnier**

$$\Delta_d = (1 - B^d).$$

Si par exemple s_t est de période d alors $\Delta_d(m_t + s_t + Z_t) = m_t - m_{t-d} + Z_t - Z_{t-d}$. Notons que Δ_d transforme une tendance linéaire at en une constante ad .

Des exemples concrets d'élimination de tendance et de saisonnalité par différenciation sont donnés dans les figures 1.5, 1.6, 1.7.

De manière synthétique, l'approche revient à modéliser la série temporelle observée par un processus X solution du modèle implicite (ou problème inverse)

$$P(B)X = Y$$

où

- $P(B)$ est un polynôme (estimable) de l'opérateur retard B ;
- Y est un processus stationnaire (covariance estimable).

On dit que X s'obtient en intégrant (inverse de la différenciation) Y . Si l'opérateur $P(B)$ ne fait intervenir que Δ et si de plus Y est modélisé par un processus stationnaire **ARMA** (Auto Regressive Moving Average, introduit par la suite), alors on dit que X est un processus **ARIMA** (Auto Regressive Integrated Moving Average). Lorsque $P(B)$ fait également intervenir l'opérateur de différenciation saisonnière Δ_d , alors on dit que X est un processus **SARIMA** (Seasonal ARIMA).

La suite du cours est consacrée à la modélisation des processus stationnaires. Sauf mention explicite du contraire, on considère un temps discret indicé par

$$T = \mathbb{Z}.$$

Les variables aléatoires sont définies sur un même espace de probabilité

$$(\Omega, \mathcal{A}, \mathbb{P}).$$

1.2 Processus stationnaires

Un **processus stochastique** ou **processus** est une famille $(X_t)_{t \in \mathbb{Z}}$ de variables aléatoires à valeurs dans \mathbb{R} . Le temps est en indice et l'aléa entre parenthèses :

$$\begin{aligned} \Omega \times \mathbb{Z} &\rightarrow \mathbb{R} \\ (\omega, t) &\mapsto X_t(\omega) \end{aligned}$$

Pour tout $t \in T$, X_t est une variable aléatoire. Pour tout $\omega \in \Omega$, on dit que la fonction du temps $t \mapsto X_t(\omega)$ est une **trajectoire du processus**.

Définition 1.2 : Processus fortement stationnaire

Le processus $(X_t)_{t \in \mathbb{Z}}$ est **fortement stationnaire** lorsque pour tout $h \in \mathbb{Z}$ et toute suite finie $t_1, \dots, t_n \in \mathbb{Z}$ de longueur $n \geq 1$ quelconque, les vecteurs aléatoires

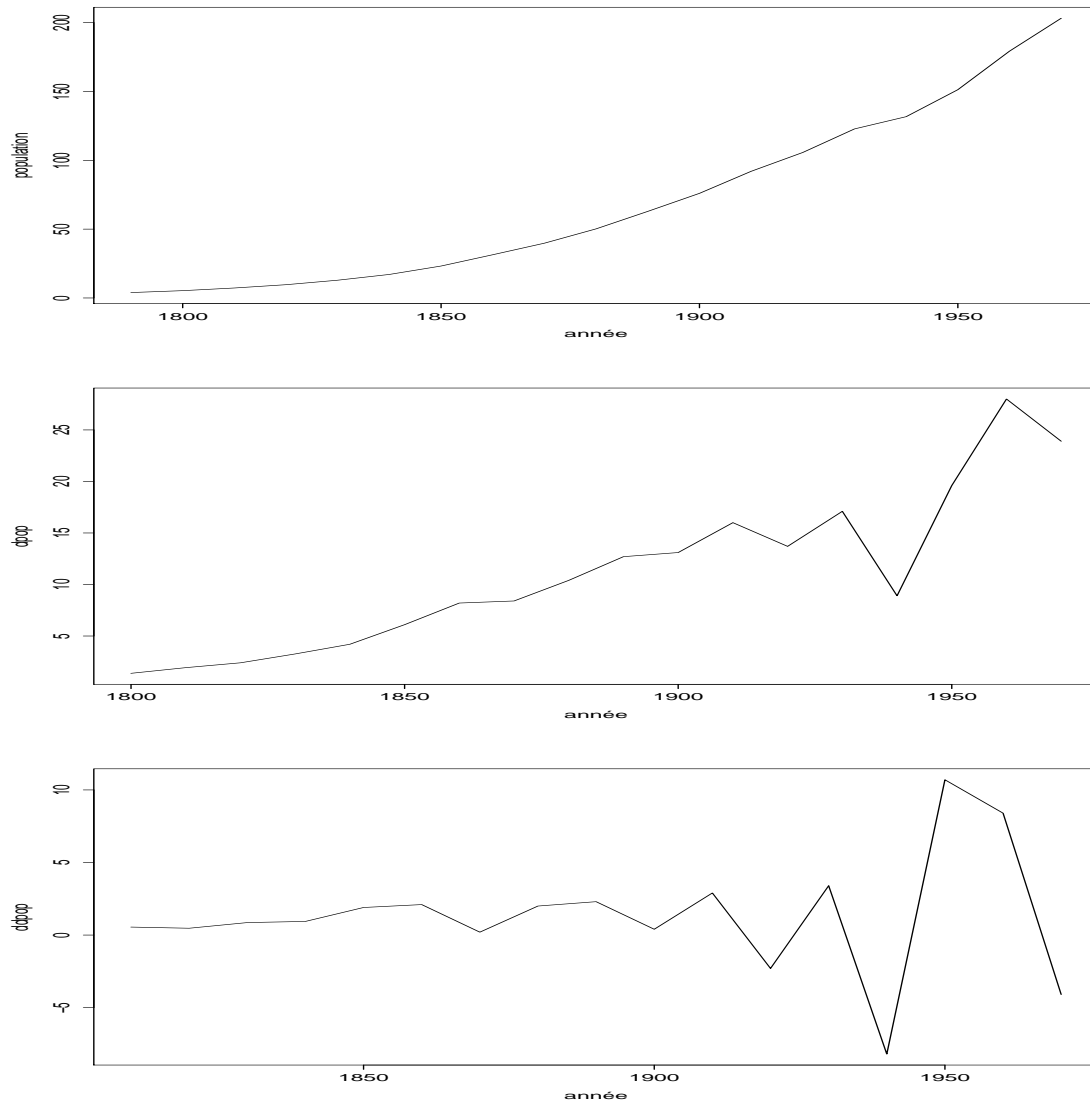


FIGURE 1.5 – Exemple d'élimination de tendance par différenciation.

La série d'origine (haut), après différenciation de degré 1 (milieu), et après différenciation de degré 2 (bas), grâce au code GNU-R ci-dessous.

```
# Inspiré du livre "Séries temporelles avec R" de Yves Aragon.
plot.ts(uspop,xlab='année',ylab='population')
plot.ts(diff(uspop),xlab='année',ylab='dpop')
plot.ts(diff(uspop, differences = 2),xlab='année',ylab='ddpop')
```

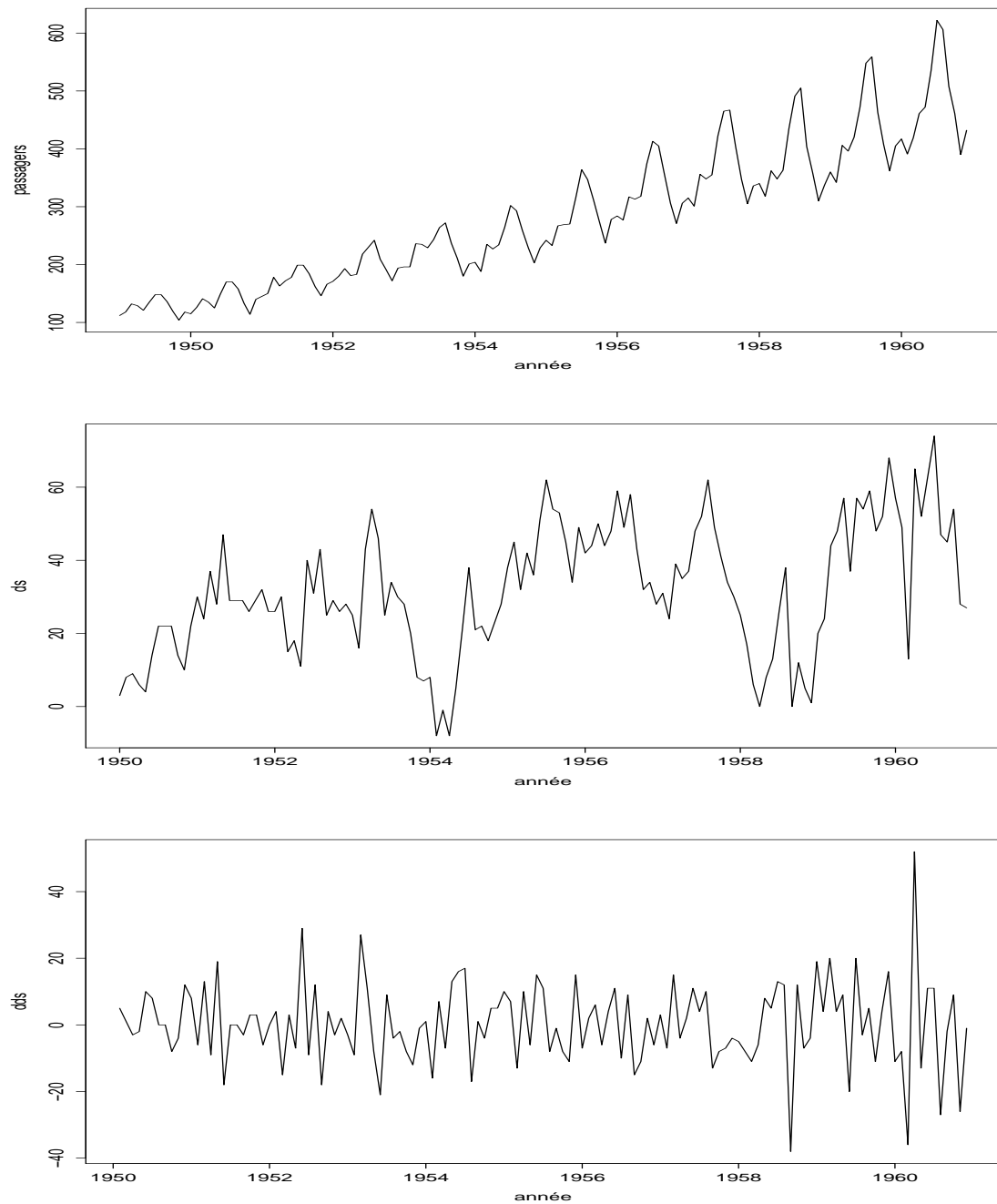


FIGURE 1.6 – Exemple d'élimination de saisonnalité et de tendance par différentiation.

La série d'origine (haut) est celle du nombre de passagers aériens, ce qui conduit à effectuer une différentiation saisonnière Δ_{12} (milieu), puis une différentiation supplémentaire de degré 1 pour éliminer la tendance (bas). Le code GNU-R est donné ci-dessous :

```
# Inspiré du livre "Séries temporelles avec R" de Yves Aragon.
plot.ts(AirPassengers,xlab='année',ylab='passagers')
plot.ts(diff(AirPassengers,lag=12),xlab='année',ylab='ds')
plot.ts(diff(diff(AirPassengers,lag=12)),xlab='année',ylab='dds')
```

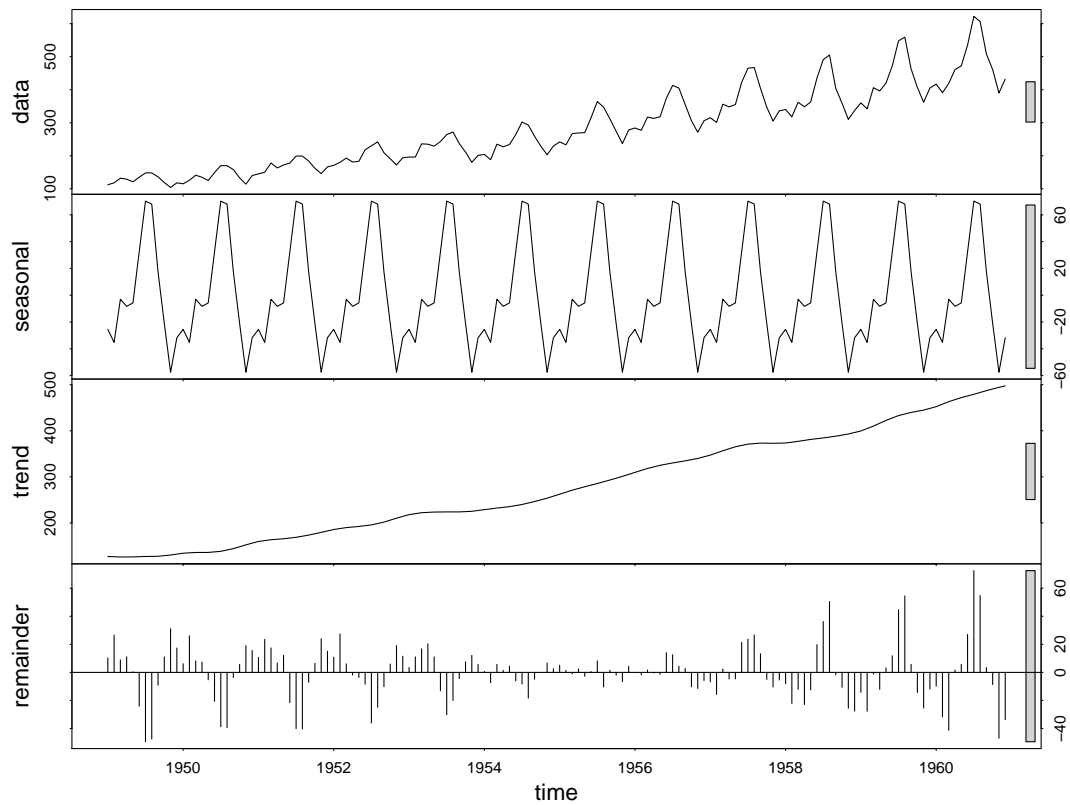


FIGURE 1.7 – Estimation de la tendance et de la saisonnalité par moindre carrés.
Série : nombre de passagers aériens. Code GNU-R :

```
# Inspiré du livre "Séries temporelles avec R" de Yves Aragon.
# Plusieurs fonctions permettent un ajustement (entre autres choses) :
#   lsfit : Find the Least Squares Fit
#   lm    : Fitting Linear Models
#   arima : ARIMA Modelling of Time Series
#   stl   : Seasonal Decomposition of Time Series by "Loess"
# Les fonctions lsfit, lm, et arima nécessitent de se fixer un modèle.
# La fonction stl en revanche intègre un traitement automatique.
plot(stl(AirPassengers,s.window="periodic"))
```


$(X_{t_1}, \dots, X_{t_n})$ et $(X_{t_1+h}, \dots, X_{t_n+h})$ ont même loi.

La stationnarité forte est une propriété d'invariance en loi du processus par translation temporelle. Le concept n'est pas creux : si les variables aléatoires $(X_t)_{t \in T}$ sont i.i.d. alors $(X_t)_{t \in T}$ est fortement stationnaire (la réciproque est fausse).

Définition 1.3 : Second ordre : moyenne et autocovariance

Le processus $(X_t)_{t \in \mathbb{Z}}$ est un **processus du second ordre** lorsque $\mathbb{E}(|X_t|^2) < \infty$ pour tout $t \in \mathbb{Z}$. Pour un tel processus, on définit la **fonction moyenne** $\mu_X : \mathbb{Z} \rightarrow \mathbb{R}$ et la **fonction d'autocovariance** $\gamma_X : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$ par

$$\mu_X(t) = \mathbb{E}(X_t) \quad \text{et} \quad \gamma_X(s, t) = \text{Cov}(X_s, X_t).$$

Rappelons que si L^p désigne l'espace de Lebesgue des variables aléatoires possédant un moment d'ordre p fini, alors $L^2 \subset L^1$. Pour tous $s, t \in \mathbb{Z}$, $X_t \in L^2 \subset L^1$ et

$$\begin{aligned} \gamma_t(s, t) &= \text{Cov}(X_s, X_t) \\ &= \mathbb{E}((X_s - \mathbb{E}(X_s))(X_t - \mathbb{E}(X_t))) \\ &= \mathbb{E}(X_s X_t) - \mathbb{E}(X_s)\mathbb{E}(X_t) \\ &= \mathbb{E}(X_s X_t) - \mu_X(s)\mu_X(t). \end{aligned}$$

Par analogie avec les vecteurs aléatoires finis usuels de \mathbb{R}^n , le vecteur (infini) aléatoire $(X_t)_{t \in \mathbb{Z}}$ a pour vecteur (infini) moyenne μ_X et pour matrice (infinie) de covariance $\gamma_X = (\gamma_X(s, t))_{s, t \in \mathbb{Z}}$. De ce point de vue, il n'est pas étonnant que γ_X soit symétrique à diagonale réelle ≥ 0 (la diagonale donne les variances des composantes) :

$$\gamma_X(t, t) = \text{Var}(X_t) = \mathbb{E}(X_t^2) - \mathbb{E}(X_t)^2.$$

La dépendance à longue portée au sens L^2 (corrélation) est mesurée par $\gamma_X(t, t+h)$ quand h est grand (ne dépend pas de t si le processus est stationnaire).

Définition 1.4 : Stationnarité

On dit qu'un processus du second ordre $(X_t)_{t \in \mathbb{Z}}$ est **faiblement stationnaire** ou tout simplement **stationnaire** lorsque pour tout $h \in \mathbb{Z}$ et toute suite finie $t_1, \dots, t_n \in \mathbb{Z}$ de longueur $n \geq 1$ quelconque, les vecteurs aléatoires

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{et} \quad (X_{t_1+h}, \dots, X_{t_n+h}).$$

ont même **espérance** et **matrice de covariance**. Cela revient à dire, de manière équivalente, que μ_X est constante et γ_X est invariante par translation :

$$\forall s, t, h \in \mathbb{Z}, \quad \mu_X(t) = \mu_X(t+h) \quad \text{et} \quad \gamma_X(s, t) = \gamma_X(s+h, t+h).$$

Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire alors $\gamma_X(s, t) = \gamma_X(0, t-s)$ pour tous $s, t \in \mathbb{Z}$. Ainsi, la covariance $\gamma_X(s, t) = \text{Cov}(X_s, X_t)$ ne dépend que de l'écart temporel $t-s$. Dans ce cas, la fonction à deux variables (matrice infinie) γ_X peut être remplacée par une fonction à une seule variable (vecteur infini), notée également γ_X :

Définition 1.5 : Autocovariance

Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, sa **fonction d'autocovariance** est définie par

$$\begin{aligned} \mathbb{Z} &\rightarrow \mathbb{R} \\ h &\mapsto \gamma_X(h) = \gamma_X(0, h) = \text{Cov}(X_t, X_{t+h}) \quad (\forall t \in \mathbb{Z}) \end{aligned}$$

de sorte que $\gamma_X(s, t) = \gamma_X(t - s)$ pour tous $s, t \in \mathbb{Z}$.

Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, alors non seulement sa fonction moyenne $t \mapsto \mu_X(t)$ est constante, mais sa fonction variance $t \mapsto \sigma_t^2$ aussi car

$$\sigma_t^2 = \text{Var}(X_t) = \gamma_X(t, t) = \gamma_X(0)$$

ne dépend plus de t . Idem pour $t \mapsto \mathbb{E}(X_t^2) = \sigma_t^2 + \mu_X(t)^2 = \gamma_X(0)^2 + \mu_X^2$.

Définition 1.6 : Fonction d'autocorrélation

Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, sa **fonction d'autocorrélation** est définie par

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

Le termes sont bien choisis :

Théorème 1.7 : Stationnarité faible et forte

Un processus du second ordre fortement stationnaire est toujours stationnaire. La réciproque est fautive : il existe des processus stationnaires qui ne sont pas fortement stationnaires.

Démonstration. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus du second ordre fortement stationnaire alors pour tous $s, t, h \in \mathbb{Z}$, les vecteurs aléatoires (X_s, X_t) et (X_{s+h}, X_{t+h}) ont la même loi, et donc la même moyenne et la même matrice de covariance, ce qui entraîne la constance de μ_X et l'invariance par translation de γ_X , d'où la stationnarité de X .

Construisons un contre-exemple pour la réciproque. Soit $(X_t)_{t \in \mathbb{Z}}$ des v.a.r. indépendantes avec X_t de loi \mathcal{L} si t est pair, et de loi $\mathcal{L}' \neq \mathcal{L}$ si t est impair, avec \mathcal{L} et \mathcal{L}' de même moyenne et variance, par exemple $\mathcal{L} = \text{Exp}(1)$ et $\mathcal{L}' = \mathcal{N}(1, 1)$. On a alors $\gamma_X(s, t) = 0$ si $s \neq t$ par indépendance tandis que $\gamma_X(s, s) = 1$, ce qui fait que le processus est stationnaire, mais pas fortement stationnaire car les X_t n'ont pas même loi! \square

1→2

La moyenne et l'autocovariance constituent les deux premiers moments du processus, et les deux premiers moments ne suffisent pas à caractériser pas la loi. Au delà de la stationnarité, si X et Y sont deux processus du second ordre tels que $\mu_X = \mu_Y$ et $\gamma_X = \gamma_Y$ alors X et Y ne sont pas forcément de même loi (sauf pour les processus gaussiens!).

Définition 1.8 : Processus gaussiens

On dit que $(X_t)_{t \in \mathbb{Z}}$ est un **processus gaussien** lorsque $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien pour toute suite finie t_1, \dots, t_n dans \mathbb{Z} de longueur $n \geq 1$ quelconque. Un processus gaussien est toujours du second ordre.

Pour un processus gaussien, la stationnarité forte est équivalente à la stationnarité (faible) car la loi d'un vecteur gaussien est caractérisée par sa moyenne et sa covariance.

Définition 1.9 : Bruit blanc

Un processus stationnaire $(Z_t)_{t \in \mathbb{Z}}$ de moyenne μ et d'autocovariance γ_Z est un...

- **bruit blanc faible** ou **bruit blanc** si $\text{Cov}(Z_s, Z_t) = 0$ pour tous $s \neq t$;
- **bruit blanc moyennement fort** si Z_s et Z_t sont indépendantes pour tous $s \neq t$, c'est-à-dire que $(Z_t)_{t \in \mathbb{Z}}$ sont deux à deux indépendantes;
- **bruit blanc fort** lorsque les variables $(Z_t)_{t \in \mathbb{Z}}$ sont indépendantes;
- **bruit blanc très fort** lorsque les variables $(Z_t)_{t \in \mathbb{Z}}$ sont i.i.d.

Dans tous les cas $\gamma_Z(h) = \sigma^2 \mathbf{1}_{h=0}$ et on note $\text{BB}(\mu, \sigma^2)$. Dans ces notes de cours, lorsque la moyenne du BB n'est pas précisée, elle vaut zéro par convention.

Le terme «bruit blanc» tire son sens de l'analyse spectrale abordée plus loin.

Exemple 1.10 : Bruit blanc gaussien

Si $X = (X_t)_{t \in \mathbb{Z}}$ sont i.i.d. de loi $\mathcal{N}(m, \sigma^2)$ alors X est un $\text{BB}(m, \sigma^2)$ fort gaussien. La figure 1.8 propose une simulation d'un bruit blanc gaussien.

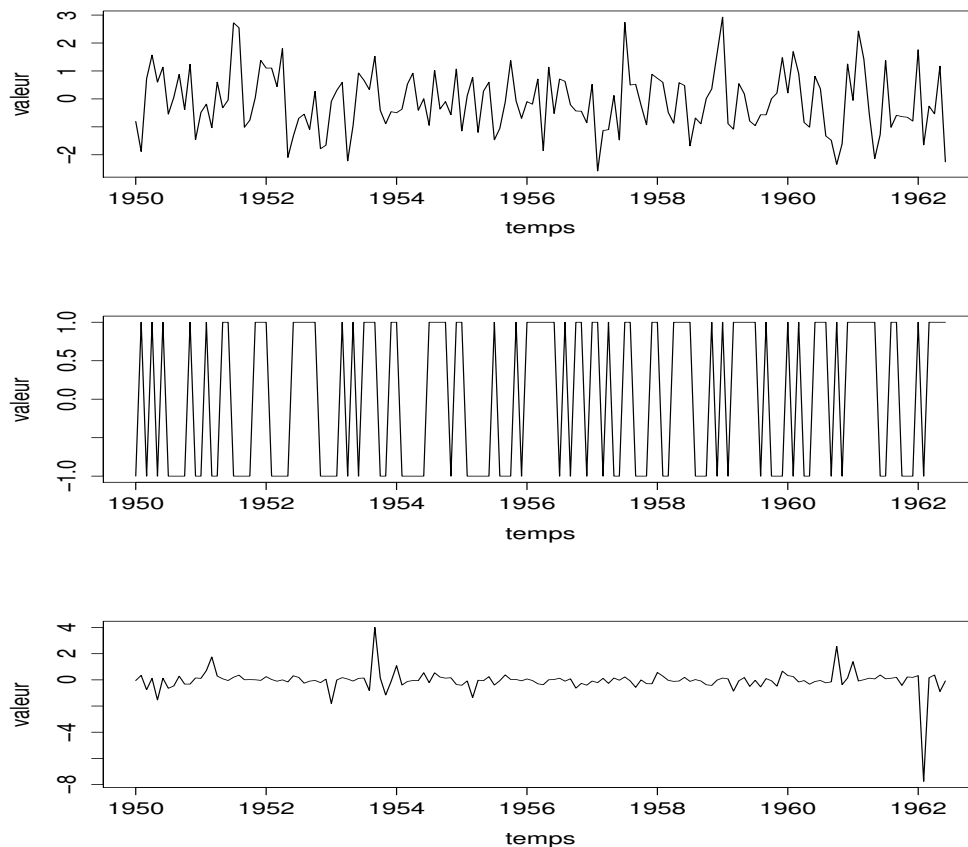
Exemple 1.11 : Élémentaire ?

L'exemple le plus élémentaire de processus stationnaire est donné par $X_t = A$ pour tout $t \in \mathbb{Z}$, où A est une v.a.r. de carré intégrable. Les trajectoires du processus sont des constantes. On a $\mu_X(t) = \mathbb{E}(A)$ et $\gamma_X(s, t) = \text{Var}(A)$ pour tous $s, t \in \mathbb{Z}$.

Exemple 1.12 : Processus harmonique

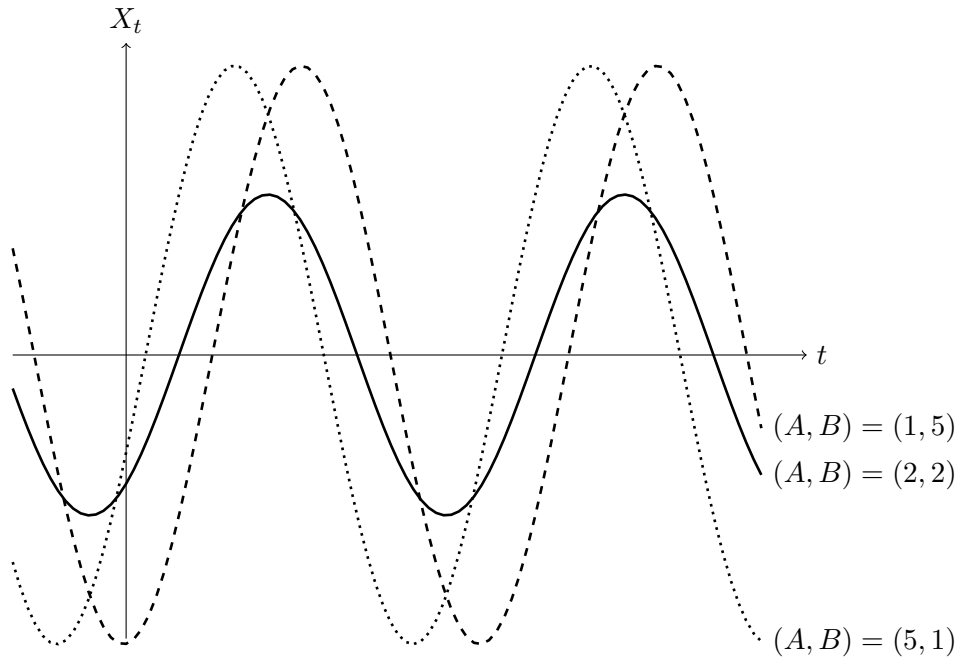
Soient A, B deux v.a.r. non corrélées de moyenne 0 et de variance σ^2 , et $\theta \in [-\pi, \pi]$ une constante. On considère le **processus harmonique** défini par

$$\forall t \in \mathbb{Z}, X_t = A \cos(\theta t) + B \sin(\theta t).$$

FIGURE 1.8 – Simulation d'un $BB(0,1)$ (loi de Gauss, Rademacher, Student).

Code GNU-R utilisé pour la simulation :

```
# Création d'une série mensuelle, à partir de janvier 1950
bbg = ts(rnorm(150,0,1),start=c(1950,1),frequency=12) # Gauss
plot.ts(bbg,xlab='temps',ylab='valeur')
bbb = ts(2*rbinom(150,1,1/2)-1,start=c(1950,1),frequency=12) # Rademacher
plot.ts(bbb,xlab='temps',ylab='valeur')
n = 2.1;
bbt = ts(rt(150,n,0)*sqrt(1-2/n),start=c(1950,1),frequency=12) # Student
plot.ts(bbt,xlab='temps',ylab='valeur')
```

FIGURE 1.9 – Trois trajectoires d'un processus harmonique de fréquence $\theta = 1$.

Le processus est centré : $\mathbb{E}(X_t) = \cos(\theta t)\mathbb{E}(A) + \sin(\theta t)\mathbb{E}(B) = 0$. D'autre part

$$\begin{aligned}\mathbb{E}(X_t X_{t+h}) &= \cos(\theta t) \cos(\theta(t+h))\mathbb{E}(A^2) + \sin(\theta t) \sin(\theta(t+h))\mathbb{E}(B^2) \\ &\quad + (\dots) \underbrace{\mathbb{E}(AB)}_{=0} \\ &= \sigma^2 \Re(e^{i\theta t} e^{-i\theta(t+h)}) = \sigma^2 \Re(e^{-i\theta h}) = \sigma^2 \cos(\theta h),\end{aligned}$$

qui ne dépend pas de t , donc X est stationnaire, d'autocovariance $\gamma_X(h) = \sigma^2 \cos(\theta h)$. Pourtant, les trajectoires du processus sont régulières, et ne ressemblent pas à l'idée qu'on se fait d'un processus stationnaire. Cela vient de la manière de rendre aléatoires les trajectoires, via les amplitudes A et B seulement. Une trajectoire de processus harmonique a la même allure qu'une saisonnalité pure, cf. figure 1.9. Un processus harmonique est une saisonnalité à amplitude aléatoire en quelque sorte. L'exemple 1.20 révèle une propriété déterministe très naturelle du processus harmonique.

Tout processus est une variable aléatoire à valeurs trajectoires. Observer une seule trajectoire du processus revient à observer une seule réalisation de cette variable aléatoire.

Remarque 1.13 : Comment reconnaître une suite stationnaire

Une suite doublement infinie de variables aléatoires est stationnaire quand la moyenne est constante et l'autocovariance est invariante par translation, en particulier la variance est constante. Sauf à se restreindre à un modèle particulier comme les suites récurrentes aléatoires, l'observation d'une seule trajectoire d'un processus ne nous renseigne pas beaucoup sur la structure de son aléa : c'est un échantillon de taille 1 de la loi des trajectoires ! Or la stationnarité est liée à la structure de l'aléa, et

ceci explique pourquoi elle est difficile à appréhender visuellement, comme l'illustre l'exemple 1.12 du processus harmonique, pour lequel les trajectoires sont régulières. Il est malgré tout possible de tester la stationnarité en précisant un modèle. Le test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) de la commande GNU-R `kpss.test` par exemple permet de tester la stationnarité dans le cadre d'un modèle additif simple.

Dans la pratique, la stationnarité des séries temporelles peut s'avérer être une hypothèse raisonnable sur une plage observée, mais restera toujours hypothétique lors d'une prédiction, qui va par définition au delà des plages observées. Concepts reliés : ruptures, catastrophes, «cygnes noirs» (**black swans**), etc.

1.3 Suites récurrentes aléatoires et processus ARMA

Les suites récurrentes aléatoires constituent le modèle de processus le plus répandu.

Définition 1.14 : Suites récurrentes aléatoires d'ordre fini – ARMA(p, q)

Soit $p, q \in \mathbb{N}$ et $T \in \{\mathbb{N}, \mathbb{Z}\}$. On dit qu'un processus $(X_t)_{t \in T}$ est une **suite récurrente aléatoire** d'ordre (p, q) lorsque pour tout $t > p$, la variable aléatoire X_t est une fonction du passé X_{t-1}, \dots, X_{t-p} et de bruits $Z_t, Z_{t-1}, \dots, Z_{t-q}$ où $(Z_t)_{t \in \mathbb{Z}}$ est un BB. C'est le cas par exemple si pour une fonction f fixée on a

$$\forall t > p, X_t = f(X_{t-1}, \dots, X_{t-p}, Z_t, Z_{t-1}, \dots, Z_{t-q}).$$

- On dit que le processus est **markovien** lorsque $(p, q) = (1, 0)$;
- On dit que le processus a une **mémoire longue** lorsque $p = \infty$;
- On dit que le processus est un **ARMA**(p, q) lorsque $T = \mathbb{Z}$ et f est linéaire ; le cas $q = 0$ correspond aux processus **AR**(p) (Auto Regressive) tandis que le cas $p = 0$ correspond aux processus **MA**(q) (Moving Average).

Considérons une équation de récurrence aléatoire d'ordre (p, q) en $X = (X_t)_{t \in T}$. Si $T = \mathbb{N}$ alors X est bien défini par récurrence à partir des valeurs initiales X_0, \dots, X_{p-1} et de Z grâce à l'équation de récurrence. Si en revanche $T = \mathbb{Z}$, alors l'existence d'un tel X n'est pas claire. Le but du cours est d'aborder les questions suivantes :

- Sous quelles conditions les processus ARMA(p, q) existent-ils ?
- Sont-ils stationnaires ? Quelles sont leurs propriétés ?
- Comment les ajuster au mieux aux observations ?
- Comment s'en servir pour effectuer des prévisions ?

Examinons tout d'abord quelques cas particuliers simples.

Exemple 1.15 : Processus à moyenne mobile (MA)

Un processus $(X_t)_{t \in \mathbb{Z}}$ est à **moyenne mobile** (MA : Moving Average) lorsque

$$\forall t \in \mathbb{Z}, X_t = Z_t + \theta Z_{t-1}$$

où $Z = (Z_t)_{t \in \mathbb{Z}}$ est un BB(0, σ^2). On a $\mu(t) = \mathbb{E}(Z_t) + \theta \mathbb{E}(Z_{t-1}) = 0$ et

$$\gamma(s, t) = \mathbb{E}(X_s X_t) = \begin{cases} 0 & \text{si } |s - t| > 1; \\ \theta \sigma^2 & \text{si } |s - t| = 1; \\ (1 + \theta^2) \sigma^2 & \text{si } s = t. \end{cases}$$

En effet, comme $\mu_Z = 0$ et $\gamma_Z(h) = \sigma^2 \mathbf{1}_{h=0}$, il vient, pour tous $t, h \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{E}(X_t X_{t+h}) &= \mathbb{E}((Z_t + \theta Z_{t-1})(Z_{t+h} + \theta Z_{t+h-1})) \\ &= \mathbb{E}(Z_t Z_{t+h}) + \theta^2 \mathbb{E}(Z_{t-1} Z_{t+h-1}) + \theta \mathbb{E}(Z_{t-1} Z_{t+h}) + \theta \mathbb{E}(Z_t Z_{t+h-1}) \\ &= \gamma_Z(h) + \theta^2 \gamma_Z(h) + \theta \gamma_Z(h+1) + \theta \gamma_Z(h-1) \\ &= (1 + \theta^2) \sigma^2 \mathbf{1}_{h=0} + \theta \sigma^2 \mathbf{1}_{h=\pm 1}. \end{aligned}$$

Ce processus du second ordre est donc stationnaire. Si Z est un BB fortement stationnaire, alors X est aussi fortement stationnaire car

$$\begin{aligned} (X_{t_1}, \dots, X_{t_n}) &= (Z_{t_1}, \dots, Z_{t_n}) + \theta(Z_{t_1-1}, \dots, Z_{t_n-1}) \\ &= F(Z_{t_1}, \dots, Z_{t_n}, Z_{t_1-1}, \dots, Z_{t_n-1}) \\ &\stackrel{\text{loi}}{=} F(Z_{t_1+h}, \dots, Z_{t_n+h}, Z_{t_1+h-1}, \dots, Z_{t_n+h-1}) \\ &= (X_{t_1+h}, \dots, X_{t_n+h}). \end{aligned}$$

Plus généralement, pour h et $\theta_0, \dots, \theta_q$ fixés, on peut considérer le processus MA(q)

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}.$$

Proposition 1.16 : Stationnarité et auto-covariance d'un MA(q)

Soit

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

un processus MA(q). Alors $(X_t)_{t \in \mathbb{Z}}$ est stationnaire et $\gamma_X(h) = 0$, $\forall |h| > q$.

Des résultats plus généraux concernant les processus ARMA(p, q) seront présentés au Chapitre 3.

Démonstration. Nous avons bien

$$\mu_X(t) = \sum_{j=0}^q \theta_j \mathbb{E}[Z_{t-j}] = 0$$

constant. Calculons maintenant l'auto-covariance :

$$\begin{aligned}
 \gamma_X(t, t+h) &= \text{Cov}(X_t, X_{t+h}) = \text{Cov}\left(\sum_{j=0}^q \theta_j Z_{t-j}, \sum_{k=0}^q \theta_k Z_{t+h-k}\right) \\
 &= \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \text{Cov}(Z_{t-j}, Z_{t+h-k}) \\
 &= \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \gamma_Z(h-k+j) = \sigma^2 \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \mathbf{1}_{k-j=h},
 \end{aligned}$$

qui ne dépend pas de t . Cela implique que $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, de fonction d'auto-covariance

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \mathbf{1}_{k-j=h}.$$

En particulier, si $j, k \in \{0, \dots, q\}$, nous avons $-q \leq k-j \leq q$, ce qui implique que $\gamma_X(h) = 0$ dès que $h < -q$ ou $h > q$. \square

Exemple 1.17 : Marche aléatoire

On peut étudier les séries temporelles avec une origine des temps, ce qui revient à indexer le temps par l'ensemble \mathbb{N} au lieu de l'ensemble \mathbb{Z} . Considérons par exemple la **marche aléatoire** $X = (X_t)_{t \in \mathbb{N}}$ définie par $X_0 = 0$ et l'équation de récurrence linéaire d'ordre 1 ou AR(1) suivante :

$$X_{t+1} = X_t + Z_{t+1} = Z_1 + \dots + Z_{t+1}$$

où $(Z_n)_{n \geq 1}$ sont des variables aléatoires indépendantes et identiquement distribuées de moyenne 0 et de variance σ^2 . La moyenne vaut $\mu(t) = \mathbb{E}(X_t) = 0$. Le processus X n'est pas stationnaire car comme pour tous $t, h \in \mathbb{N}$,

$$X_t = Z_1 + \dots + Z_t \quad \text{et} \quad X_{t+h} - X_t = Z_{t+1} + \dots + Z_{t+h}$$

sont indépendantes et centrées, on a $\text{Cov}(X_t, X_{t+h} - X_t) = 0$, d'où

$$\begin{aligned}
 \gamma(t, t+h) &= \text{Cov}(X_t, X_{t+h} - X_t + X_t) \\
 &= \text{Cov}(X_t, X_{t+h} - X_t) + \text{Var}(X_t) \\
 &= \text{Var}(X_t) \\
 &= t\sigma^2
 \end{aligned}$$

qui dépend de t . La marche aléatoire est un processus AR(1) non stationnaire.

1.4 Autocovariance et matrices de Toeplitz

Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire alors pour toute suite finie t_1, \dots, t_n dans \mathbb{Z} , la matrice de covariance Γ du vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ vérifie

$$\Gamma_{j,k} = \text{Cov}(X_{t_j}, X_{t_k}) = \gamma(t_k - t_j).$$

De plus, si les t_1, \dots, t_n sont régulièrement espacés alors les coefficients de Γ sont constants le long de chaque diagonale : on dit qu'il s'agit d'une **matrice de Toeplitz**. Plus généralement, le vecteur (infini) aléatoire $(X_t)_{t \in \mathbb{Z}}$ a pour matrice (infinie) de covariance $(\gamma_X(t-s))_{s,t \in \mathbb{Z}}$, dont les coefficients sont constants le long des diagonales.

Dans toute la suite, on réservera la lettre i au nombre complexe $(0, 1)$ et on utilisera plutôt les lettres j, k, h pour indexer les sommes.

$$(\gamma(j-k))_{1 \leq j, k \leq p} = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \ddots & & \vdots \\ \gamma(2) & \gamma(1) & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \gamma(1) & \gamma(2) \\ \vdots & & \ddots & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(p-1) & \dots & \dots & \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix}.$$

Théorème 1.18 : Autocovariance

Si $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ est la fonction d'autocovariance d'un processus stationnaire alors elle est **symétrique de type positif**, c'est-à-dire qu'elle vérifie les propriétés suivantes :

1. $\gamma(h) = \gamma(-h)$ pour tout $h \in \mathbb{Z}$;
2. pour tout $n \geq 1$, et tout $t \in \mathbb{Z}^n$, la matrice $\Gamma = (\gamma(t_j - t_k))_{1 \leq j, k \leq n}$ vérifie :

$$\forall v \in \mathbb{R}^n, \langle v, \Gamma v \rangle = \sum_{j=1}^n \sum_{k=1}^n v_j \gamma(t_j - t_k) v_k \geq 0.$$

De plus $\Gamma_{1,1} = \dots = \Gamma_{n,n} = \gamma(0) \geq 0$ et $|\gamma(h)| \leq \gamma(0)$ pour tout $h \in \mathbb{Z}$. Réciproquement, si $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ est symétrique et de type positif, alors c'est la fonction d'autocovariance d'un processus (fortement) stationnaire gaussien.

Note : la seconde condition signifie que le spectre de la matrice symétrique Γ est ≥ 0 .

Note : on ne perd rien à prendre $t = (1, \dots, n)$, quitte à extraire une sous-matrice.

Preuve (en partie facultative). Le vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ de \mathbb{R}^n a pour matrice de covariance $\Gamma = (\gamma(t_j - t_k))_{1 \leq j, k \leq n}$. En posant

$$V = (X_{t_1} - \mathbb{E}(X_{t_1}), \dots, X_{t_n} - \mathbb{E}(X_{t_n})),$$

on a, en utilisant la linéarité de l'espérance et le fait que $v^\top V$ est une v.a.r. centrée,

$$\sum_{j=1}^n \sum_{k=1}^n v_j \Gamma_{j,k} v_k = v^\top \underbrace{\mathbb{E}(VV^\top)}_{\text{matrice!}} v = \mathbb{E}(|v^\top V|^2) = \text{Var}(v^\top V) \geq 0.$$

Enfin, on a $\gamma(0) = \text{Var}(X_t) \geq 0$ pour tout $t \in \mathbb{Z}$, et pour tout $h \in \mathbb{Z}$,

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t+h}, X_t) = \gamma(-h),$$

tandis que par l'inégalité de Cauchy-Schwarz dans L^2 ,

$$|\gamma(h)| = |\mathbb{E}((X_t - \mathbb{E}(X_t))(X_{t+h} - \mathbb{E}(X_{t+h})))| \leq \sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_{t+h})} = \gamma(0).$$

Cette partie de la preuve est facultative. Il nous reste à construire un processus à fonction d'autocovariance prescrite γ (symétrique de type positif). Munissons $\mathbb{R}^{\mathbb{Z}}$ de la topologie produit, qui est la topologie la moins fine rendant les projections (applications coordonnées) continues, et de la tribu borélienne associée, qui coïncide avec la tribu produit engendrée par les cylindres. Nous allons construire une loi gaussienne sur $\mathbb{R}^{\mathbb{Z}}$ (i.e. les lois marginales de dimension finie sont gaussiennes), invariante par translation, centrée, et de fonction de covariance γ . On considère tout d'abord une suite de v.a.r. i.i.d. $(Z_k)_{k \in \mathbb{Z}}$ de loi gaussienne centrée réduite. La loi de cette suite est une loi gaussienne produit sur $\mathbb{R}^{\mathbb{Z}}$. Pour tout entier $n \geq 0$, on construit, par exemple grâce à la **décomposition de Cholesky**. (théorème 5.6), une matrice $L \in \mathcal{M}_{2n+1, 2n+1}(\mathbb{R})$ telle que

$$LL^{\top} = (\gamma(j-k))_{-n \leq j, k \leq n},$$

de sorte que $Y = L(Z_k)_{-n \leq k \leq n}$ soit un vecteur gaussien de matrice de covariance LL^{\top} . On pose alors $X_k = Y_k$ si $-n \leq k \leq n$ et $X_k = Z_k$ sinon. On peut alors montrer que X converge en loi quand $n \rightarrow \infty$ (fonctions test d'un nombre fini de variables). \square

2→3

1.5 Compléments

Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus du second ordre. Pour tout $t \in \mathbb{Z}$, on note

$$H_{t-1} := \overline{\text{vect}\{X_{t-1}, X_{t-2}, \dots\}}$$

l'adhérence dans L^2 du sous-espace vectoriel $\text{vect}\{X_{t-1}, X_{t-2}, \dots\}$, c'est-à-dire l'ensemble des combinaisons linéaires $\sum_{k=1}^{\infty} \varphi_k X_{t-k}$ convergentes dans L^2 . On note également

$$X_t^* := \text{proj}(X_t, H_{t-1}) = \arg \min_{Y \in H_{t-1}} \|X_t - Y\|_2 \in H_{t-1}$$

la projection orthogonale dans L^2 de $X_t \in L^2$ sur le sous-espace vectoriel fermé $H_{t-1} \subset L^2$. Ces notions hilbertiennes sont centrales dans le chapitre 5 sur la prédiction linéaire. X_t^* est la **prévision optimale** de X_t .

On note également

$$\varepsilon_t = X_t - X_t^*,$$

ε_t est appelé **innovation**.

Définition 1.19 : Processus déterministe

Un processus du second ordre $(X_t)_{t \in \mathbb{Z}}$ est **déterministe** lorsque pour tout $t \in \mathbb{Z}$,

$$X_t \in H_{t-1} := \overline{\text{vect}\{X_{t-1}, X_{t-2}, \dots\}}.$$

Autrement dit $X_t = \text{proj}(X_t, H_{t-1})$ dans L^2 pour tout $t \in \mathbb{Z}$.

Note : si X est gaussien alors $\text{proj}(X_t, H_{t-1}) = \text{proj}(X_t, L^2(\mathcal{F}_{t-1})) = \mathbb{E}(X_t | \mathcal{F}_{t-1})$ où \mathcal{F}_{t-1} est la tribu engendrée par X_{t-1}, X_{t-2}, \dots

Exemple 1.20 : Le processus harmonique stationnaire est déterministe

Montrons que le **processus harmonique** de l'exemple 1.12 est déterministe. Pour tout $t \in \mathbb{Z}$, les formules trigonométriques $2 \cos(a) \cos(b) = \cos(a+b) + \cos(a-b)$ et $2 \cos(a) \sin(b) = \sin(a+b) + \sin(b-a)$ donnent, pour tout $t \in \mathbb{Z}$,

$$X_t = 2 \cos(\theta) X_{t-1} - X_{t-2} \in H_{t-1}, \quad \text{d'où} \quad X_t = \text{proj}(X_t, H_{t-1}).$$

Ceci n'est pas étonnant car les trajectoires du processus harmonique sont des sinusoides dont la seule source d'aléa est l'amplitude, ce qui fait qu'à chaque instant, il est parfaitement possible de prédire le futur de la trajectoire à partir de son passé.

Lemme 1.21 : Processus stationnaire non déterministe

Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. Alors $t \in \mathbb{Z} \mapsto \mathbb{E}((X_t - \text{proj}(X_t, H_{t-1}))^2)$ est constante. Si σ^2 est sa valeur, alors X est déterministe si et seulement si $\sigma^2 = 0$.

Un bruit blanc à variance non nulle est un exemple de processus non déterministe. En effet, si $Z \sim \text{BB}(0, \sigma^2)$ avec $\sigma^2 > 0$ alors $Z_t \perp H_{t-1} := \text{vect}\{Z_{t-1}, Z_{t-2}, \dots\}$ pour tout $t \in \mathbb{Z}$ d'où $\text{proj}(Z_t, H_{t-1}) = 0$ et $\|Z_t - \text{proj}(Z_t, H_{t-1})\|_2^2 = \|Z_t\|_2^2 = \sigma^2 > 0$.

Démonstration du lemme 1.21. Pour tout $t \in \mathbb{Z}$, le sous espace vectoriel H_{t-1} est formé par l'ensemble des séries de la forme

$$\sum_{k=1}^{\infty} \varphi_k X_{t-k}$$

qui convergent dans L^2 . Montrons que l'ensemble Φ des coefficients $(\varphi_k)_{k \geq 1}$ qui garantit la convergence dans L^2 ne dépend pas de t . Pour tous $s \geq r \geq 1$ on a

$$\left\| \sum_{k=s}^r \varphi_k X_{t-k} \right\|_2^2 = \sum_{j,k=s}^r \varphi_j \varphi_k \mathbb{E}[X_{t-k} X_{t-j}] = \sum_{j,k=s}^r \varphi_j \varphi_k (\gamma_X(j-k) + \mu_X^2),$$

quantité qui ne dépend pas de t . Ainsi le critère de Cauchy ne dépend pas de t , ce qui signifie que Φ ne dépend pas de t . À présent on a $\text{proj}(X_t, H_{t-1}) = \sum_{k=1}^{\infty} \varphi_k X_{t-k}$ avec

$$(\varphi_k)_{k \geq 1} = \arg \inf_{\varphi \in \Phi} \sum_{j,k=0}^{\infty} \varphi_j \varphi_k (\gamma_X(j-k) + \mu_X^2)$$

où on a posé $Y = \sum_{k=1}^{\infty} \varphi_k X_{t-k}$ et $\varphi_0 := -1$. Il en découle que

$$\|X_t - \text{proj}(X_t, H_{t-1})\|_2^2 = \inf_{Y \in H_{t-1}} \|X_t - Y\|_2^2 = \inf_{\varphi \in \Phi} \sum_{j,k=0}^{\infty} \varphi_j \varphi_k \gamma_X(j-k),$$

quantité qui ne dépend pas de t . Enfin $\mathbb{E}((X_t - \text{proj}(X_t, H_{t-1}))^2) = 0$ si et seulement si $X_t = \text{proj}(X_t, H_{t-1})$ dans L^2 . \square

La **décomposition de Wold** motive l'étude de processus de la forme $\sum_{k \in \mathbb{Z}} \alpha_k Y_{t-k}$, au moins dans le cas agréable où $\alpha \in \ell^1(\mathbb{Z})$. C'est précisément l'objet du chapitre suivant.

Théorème 1.22 : Décomposition ou représentation de Wold

Si $(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire non déterministe** alors pour tout $t \in \mathbb{Z}$,

$$X_t = D_t + \sum_{k \in \mathbb{Z}} \alpha_k \varepsilon_{t-k}, \quad (1.1)$$

dans L^2 , où $(D_t)_{t \in \mathbb{Z}}$, $(\alpha_k)_{k \in \mathbb{Z}}$, et $(Z_t)_{t \in \mathbb{Z}}$ vérifient les propriétés suivantes :

1. $\alpha \in \ell^2(\mathbb{Z})$ avec $\alpha_0 = 1$ et $\alpha_k = 0$ si $k < 0$;
2. $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$ avec $\sigma^2 > 0$
3. $\varepsilon_t \in H_t$ pour tout $t \in \mathbb{Z}$;
4. $\text{Cov}(\varepsilon_s, D_t) = 0$ pour tous $s, t \in \mathbb{Z}$;
5. $(D_t)_{t \in \mathbb{Z}}$ est un **processus déterministe** ;
6. $D_t \in H_s$ pour tous $s, t \in \mathbb{Z}$.

De plus cette décomposition est **unique**. Enfin, lorsque $\alpha \in \ell^1(\mathbb{Z})$, alors la série dans le membre de droite de (1.1) converge dans L^1 et $\mathbb{E}(X_t) = \mathbb{E}(D_t)$ pour tout $t \in \mathbb{Z}$.

Note : $\ell^1(\mathbb{Z}) \not\subset \ell^2(\mathbb{Z})$!

Démonstration (facultative en première lecture). Le lemme 1.21 appliqué au processus stationnaire non déterministe X donne, pour tout $t \in \mathbb{Z}$,

$$\sigma^2 := \mathbb{E}((X_t - \text{proj}(X_t, H_{t-1}))^2) = \mathbb{E}(\varepsilon_t^2) > 0.$$

Preuve de 2 et 3. Il vient que $\mathbb{E}(\varepsilon_t) = 0$. De plus par définition $\varepsilon_t \in H_t$ et $\varepsilon_t \in H_{t-1}^\perp$, donc $\varepsilon_t \in H_{t-1}^\perp \subset H_{t-2}^\perp \subset \dots$, d'où $\text{Cov}(\varepsilon_s, \varepsilon_t) = \mathbb{E}(\varepsilon_s \varepsilon_t) = 0$ pour tous $s, t \in \mathbb{Z}$ avec $s < t$. Enfin $\text{Var}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2$. Nous avons donc $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$.

Preuve de (1.1). Pour tout $t \in \mathbb{Z}$, on a

$$X_t - \varepsilon_t = X_t^* \in \overline{\text{vect}\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}}$$

et

$$\text{proj}(X_t, \overline{\text{vect}\{\varepsilon_t, \varepsilon_{t-1}, \dots\}}) = \sum_{k \in \mathbb{Z}} \alpha_k \varepsilon_{t-k}$$

où

$$\alpha_k := \begin{cases} 0 & \text{si } k < 0, \\ \frac{1}{\sigma^2} \langle X_t, Z_{t-k} \rangle & \text{si } k \geq 0. \end{cases}$$

Preuve de 1. Les coefficients $(\alpha_k)_{k \in \mathbb{Z}}$ sont indépendants de t par stationnarité et

$$\alpha_0 = \frac{1}{\sigma^2} \langle X_t, Z_t \rangle = \frac{1}{\sigma^2} \langle X_t, X_t - \text{proj}(X_t, H_{t-1}) \rangle = \frac{1}{\sigma^2} \|X_t - \text{proj}(X_t, H_{t-1})\|_2^2 = 1.$$

Preuve de 4. Pour tout $t \in \mathbb{Z}$, posons

$$D_t := X_t - \sum_{k \in \mathbb{Z}} \alpha_k Z_{t-k} = X_t - \sum_{k \in \mathbb{N}} \alpha_k Z_{t-k}.$$

On a $\langle D_t, Z_s \rangle = 0$ pour tous $s, t \in \mathbb{Z}$ avec $s \leq t$, tandis que d'autre part si $s > t$ alors $Z_s \in H_{s-1}^\perp \subset H_t^\perp$ et comme $D_t \in H_t$ il vient $\langle D_t, Z_s \rangle = 0$ pour tous $s < t$.

Preuve de 5 et 6. Comme $D_t \in H_t = H_{t-1} \oplus \text{vect}\{Z_t\}$ et comme $\langle D_t, Z_t \rangle = 0$, il vient $D_t \in H_{t-1} = H_{t-2} \oplus \text{vect}\{Z_{t-1}\}$, etc, d'où $D_t \in \cap_{k \in \mathbb{N}} H_{t-k}$, ce qui donne

$$\overline{\text{vect}\{D_t, D_{t-1}, \dots\}} \subset \cap_{k \in \mathbb{N}} H_{t-k}.$$

Par définition de D_t , $H_t = \overline{\text{vect}\{Z_t, Z_{t-1}, \dots\}} \oplus \overline{\text{vect}\{D_t, D_{t-1}, \dots\}}$. Si $Y \in \cap_{k \in \mathbb{N}} H_{t-k}$ alors $Y \in H_{s-1}$ pour tout $s \in \mathbb{N}$, d'où $\langle Y, Z_s \rangle = 0$ pour tout $s \in \mathbb{N}$, d'où $Y \in \overline{\text{vect}\{D_t, D_{t-1}, \dots\}}$. Ceci signifie que $\cap_{k \in \mathbb{N}} H_{t-k} \subset \overline{\text{vect}\{D_t, D_{t-1}, \dots\}}$ d'où

$$\cap_{k \in \mathbb{N}} H_{t-k} = \overline{\text{vect}\{D_t, D_{t-1}, \dots\}}.$$

Comme cela est valable pour tout $t \in \mathbb{Z}$ et que le membre de gauche ne dépend pas de t , on en déduit que $(D_t)_{t \in \mathbb{Z}}$ est déterministe. De plus $D_t \in H_s$ pour tout $s \in \mathbb{Z}$.

Preuve de l'unicité de la décomposition. Il est possible (exercice!) de déduire de (1.1) que les ingrédients Z , α , et D vérifient les formules indiquées, et il n'y a donc pas le choix.

Cas intégrable. Si $\alpha \in \ell^1(\mathbb{Z})$, alors pour tout $K \subset \mathbb{Z}$ fini,

$$\sum_{k \in K} \|\alpha_k Z_{t-k}\|_1 = \sum_{k \in K} |\alpha_k| \|Z_{t-k}\|_1 \leq \sum_{k \in K} |\alpha_k| \|Z_{t-k}\|_2 = \sigma \sum_{k \in K} |\alpha_k|,$$

et il en découle que la série dans (1.1) vérifie le critère de Cauchy dans l'espace de Banach L^1 . Alternativement la convergence absolue ou normale dans L^1 de la série

$$\sum_{k \in \mathbb{Z}} \|\alpha_k Z_{t-k}\|_1 = \sum_{k \in \mathbb{Z}} |\alpha_k| \|Z_{t-k}\|_1 \leq \sum_{k \in \mathbb{Z}} |\alpha_k| \|Z_{t-k}\|_2 = \sigma \sum_{k \in \mathbb{Z}} |\alpha_k| = \sigma \|\alpha\|_1 < \infty$$

implique automatiquement la convergence de la série dans l'espace de Banach L^1 . \square

La décomposition de Wold est un cas spécial d'une décomposition plus générale en théorie des opérateurs de von Neumann : la décomposition de Wold et von Neumann affirme que toute isométrie linéaire sur un espace de Hilbert est somme directe d'opérateurs de décalages (shift) unilatéraux et d'opérateurs unitaires.

Filtrage linéaire (x2.5)

Mots-clés : filtrage ; processus linéaire ; causalité ; inversibilité ; convolution.

2.1 Motivation

Considérons l'équation autorégressive AR(1) :

$$\forall t \in \mathbb{Z} : X_t = Z_t + \varphi X_{t-1},$$

où $Z = (Z_t)_{t \in \mathbb{Z}}$ est un bruit blanc et où φ est un paramètre réel déterministe (faire un dessin de trajectoire). Pour résoudre cette équation en X , on itère l'équation n fois :

$$X_t = Z_t + \varphi(Z_{t-1} + \varphi X_{t-2}) = \cdots = \sum_{k=0}^n \varphi^k Z_{t-k} + \varphi^{n+1} X_{t-(n+1)}.$$

À présent, si $|\varphi| < 1$ et si une solution X stationnaire existait alors

$$\|\varphi^{n+1} X_{t-(n+1)}\|_2^2 = |\varphi|^{2(n+1)} \|X_{t-(n+1)}\|_2^2 = |\varphi|^{2(n+1)} (\gamma_X(0) + \mu_X^2) \xrightarrow{n \rightarrow \infty} 0,$$

c'est-à-dire que $\lim_{n \rightarrow \infty} \varphi^{n+1} X_{t-(n+1)} = 0$ dans L^2 , ce qui entraînerait que la série aléatoire $\sum_{k=0}^{\infty} \varphi^k Z_{t-k}$ converge dans L^2 vers la solution stationnaire X_t . Quand est-ce que de telles séries aléatoires sont convergentes ? En quel sens ? Ce chapitre a précisément pour objectif d'étudier de telles séries aléatoires, qui apparaissent naturellement dans la résolution d'équation de type AR ou plus généralement ARMA.

Si $(X_t)_{t \in \mathbb{Z}}$ est un processus, et si $(\alpha_k)_{k \in \mathbb{Z}}$ est une suite déterministe à valeurs dans \mathbb{R} , et à support fini : $\text{card}\{k \in \mathbb{Z} : \alpha_k \neq 0\} < \infty$, alors pour tout $t \in \mathbb{Z}$, la variable aléatoire

$$Y_t = \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k}$$

a un sens car la somme est finie. Cette moyenne mobile pondérée constitue une forme de **lissage par convolution**¹ (dessin) appelé **filtrage linéaire**, analogue discret de

$$y(t) = \int_{\mathbb{R}} \alpha(s) x(t-s) ds = (\alpha * x)(t).$$

Le théorème de filtrage ci-après assure que quitte à imposer quelques contraintes de sommabilité, on peut donner un sens à Y_t même si $(\alpha_k)_{k \in \mathbb{Z}}$ n'est pas à support fini.

1. La valeur X_t au temps t est remplacée par la moyenne pondérée $\cdots + \alpha_1 X_{t-1} + \alpha_0 X_t + \alpha_{-1} X_{t+1} + \cdots$.

2.2 Théorème de filtrage

On dit qu'une suite $(x_k)_{k \in \mathbb{Z}}$ indicée par \mathbb{Z} et à valeurs dans \mathbb{R} est **sommable** ssi $\sup_{m,n \geq 0} \sum_{k=-m}^n |x_k| < \infty$. Dans ce cas, les séries $\sum_{k \geq 0} x_k$ et $\sum_{k \leq 0} x_k$ sont absolument convergentes, ce qui permet de définir la somme $\sum_{k \in \mathbb{Z}} x_k$. On note $\ell^1(\mathbb{Z})$ l'espace vectoriel des suites sommables à valeurs dans \mathbb{R} et indicées par \mathbb{Z} . La formule $\|x\|_1 = \sum_{k \in \mathbb{Z}} |x_k|$ définit une norme sur $\ell^1(\mathbb{Z})$ qui en fait un espace de Banach.

Théorème 2.1 : Filtrage linéaire des processus bornés

Soit $\alpha = (\alpha_k)_{k \in \mathbb{Z}} \in \ell^1(\mathbb{Z})$ et soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus, borné dans L^p avec $p \geq 1$, c'est-à-dire que $\sup_{t \in \mathbb{Z}} \mathbb{E}(|X_t|^p) < \infty$. Posons

$$\forall t \in \mathbb{Z}, \forall m, n \in \mathbb{N}, Y_{t,m,n} = \sum_{k=-m}^n \alpha_k X_{t-k}.$$

Alors pour tout $t \in \mathbb{Z}$ la famille $(Y_{t,m,n})_{m,n \geq 1}$ converge presque-sûrement et dans L^p lorsque $m, n \rightarrow \infty$ vers une variable aléatoire $Y_t \in L^p$:

$$Y_{t,m,n} \xrightarrow[p.s.]{m,n \rightarrow \infty} Y_t \in L^p \quad \text{et} \quad \lim_{m,n \rightarrow \infty} \mathbb{E}(|Y_{t,m,n} - Y_t|^p) = 0.$$

De plus, le processus $(Y_t)_{t \in \mathbb{Z}}$ est bien défini p.s. et est borné dans L^p .

Note : la bornitude dans L^p implique la bornitude dans L^1 (inégalité de Hölder).

Démonstration. Pour simplifier les notations, on se contente du cas $m = n$. Comme L^p est un espace de Banach², le théorème de Riesz–Fischer affirme que toute série absolument convergente est convergente (preuve par critère de Cauchy). Or pour tout $t \in \mathbb{Z}$,

$$\sup_{n \geq 1} \sum_{h=-n}^n \|\alpha_h X_{t-h}\|_p = \sup_{n \geq 1} \sum_{h=-n}^n |\alpha_h| \|X_{t-h}\|_p \leq \sup_{t \in \mathbb{Z}} \|X_t\|_p \sum_{h \in \mathbb{Z}} |\alpha_h| < \infty$$

et donc $(Y_{t,n,n})_{n \geq 1} = (\sum_{h=-n}^n \alpha_h X_{t-h})_{n \geq 1}$ converge dans L^p quand $n \rightarrow \infty$.

Il ne reste plus qu'à établir que la convergence a lieu presque sûrement lorsque $p = 1$, et que les deux limites coïncident presque sûrement. Pour cela, si $(X_t)_{t \in \mathbb{Z}}$ est bornée dans L^1 , alors le théorème de convergence monotone (ou de Fubini–Tonelli) donne, dans $[0, +\infty]$,

$$\mathbb{E}\left(\sum_{k \in \mathbb{Z}} |\alpha_k| |X_{t-k}|\right) = \sum_{k \in \mathbb{Z}} \mathbb{E}(|\alpha_k| |X_{t-k}|) \leq \sup_{s \in \mathbb{Z}} \mathbb{E}(|X_s|) \sum_{k \in \mathbb{Z}} |\alpha_k| = \sup_{s \in \mathbb{Z}} \mathbb{E}(|X_s|) \|\alpha\|_1 < \infty.$$

Donc $S := \sum_{k \in \mathbb{Z}} |\alpha_k| |X_{t-k}|$ à valeurs dans $[0, \infty]$ vérifie $\mathbb{E}(S) < \infty$, d'où $S < \infty$ p.s. c'est-à-dire que la série $\sum_{k \in \mathbb{Z}} \alpha_k X_{t-k}$ converge (absolument) p.s. Sa somme est notée Y_t .

La limite p.s. est identique à la limite dans L^1 car les deux types de convergence entraînent la convergence en probabilité. Alternativement, on peut observer que

$$\sup_n |Y_{t,n,n}| \leq \sum_{k \in \mathbb{Z}} |\alpha_k| |X_{t-k}| \in L^1,$$

et comme on sait déjà que $Y_{t,n,n} \rightarrow Y_t$ p.s., le théorème de convergence dominée implique que la limite Y_t est dans L^1 et que la convergence $Y_{t,n,n} \rightarrow Y_t$ a lieu dans L^1 .

2. Espace vectoriel normé complet : toute suite de Cauchy est convergente.

Enfin, si A_t est l'événement presque sûr sur le quel a lieu la convergence presque sûrement pour t , alors l'événement $\cap_{t \in \mathbb{Z}} A_t$ est également presque sûr ce qui assure que $Y = (Y_t)_{t \in \mathbb{Z}}$ est bien défini presque sûrement. Ce processus est borné dans L^p car grâce à la continuité et à l'inégalité triangulaire,

$$\|Y_t\|_p = \left\| \lim_{n \rightarrow \infty} Y_{t,n,n} \right\|_p = \lim_{n \rightarrow \infty} \|Y_{t,n,n}\|_p \leq \sum_{k \in \mathbb{Z}} |\alpha_k| \|X_{t-k}\|_p \leq \|\alpha\|_1 \sup_{t \in \mathbb{Z}} \|X\|_p$$

et cette dernière quantité est finie et indépendante de t . \square

Théorème 2.2 : Filtrage de processus stationnaires

Soit $\alpha \in \ell^1(\mathbb{Z})$ et soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne $\mu_X = \mathbb{E}(X_t)$ et d'autocovariance $\gamma_X(h)$. Alors le processus $Y := F_\alpha(X)$ défini par

$$\forall t \in \mathbb{Z}, \quad (F_\alpha X)_t = \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k}$$

est un processus du second ordre et stationnaire, de moyenne et d'autocovariance

$$\mu_Y = \mu_X \sum_{k \in \mathbb{Z}} \alpha_k \quad \text{et} \quad \gamma_Y(h) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h + j - k).$$

Note : ne pas confondre $\sum_{k \in \mathbb{Z}} \alpha_k$ avec $\sum_{k \in \mathbb{Z}} |\alpha_k| = \|\alpha\|_1$.

Note : la parité de γ_Y provient de celle de γ_X car pour tout $h \in \mathbb{Z}$,

$$\begin{aligned} \gamma_Y(-h) &= \sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(-h + j - k) \\ &= \sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h - j + k) \\ &= \sum_{j,k \in \mathbb{Z}} \alpha_k \alpha_j \gamma_X(h + k - j) \\ &= \gamma_Y(h). \end{aligned}$$

Ceci montre au passage qu'on peut remplacer $j - k$ par $k - j$ dans la formule.

Démonstration. Comme X est stationnaire, il est du second ordre, donc borné dans L^2 car

$$\sup_{t \in \mathbb{Z}} \|X_t\|_2^2 = \sup_{t \in \mathbb{Z}} \mathbb{E}(|X_t|^2) = \gamma_X(0)^2 + \mu_X^2 < \infty.$$

Ceci permet d'utiliser le théorème 2.1 avec $p = 2$, qui garantit que le processus $F_\alpha X$ est bien défini presque-sûrement et dans L^2 . Il s'agit donc d'un processus du second ordre dont nous allons calculer la moyenne et l'autocovariance. Rappelons que le produit scalaire de L^2 est continu : si $\lim_{n \rightarrow \infty} U_n = U$ et $\lim_{n \rightarrow \infty} V_n = V$ dans L^2 alors

$$\lim_{n \rightarrow \infty} \langle U_n, V_n \rangle = \langle \lim_{n \rightarrow \infty} U_n, \lim_{n \rightarrow \infty} V_n \rangle = \langle U, V \rangle.$$

Par conséquent, on a

$$\mathbb{E}((F_\alpha X)_t) = \langle 1, \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k} \rangle$$

$$\begin{aligned}
 &= \langle 1, \lim_{n \rightarrow \infty} \sum_{k \in [-n, n]} \alpha_k X_{t-k} \rangle \\
 &= \lim_{n \rightarrow \infty} \langle 1, \sum_{k \in [-n, n]} \alpha_k X_{t-k} \rangle \\
 &= \lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{k \in [-n, n]} \alpha_k X_{t-k} \right) \\
 &= \lim_{n \rightarrow \infty} \sum_{k \in [-n, n]} \alpha_k \mathbb{E}(X_{t-k}) \\
 &= \lim_{n \rightarrow \infty} \mu_X \sum_{k \in [-n, n]} \alpha_k \\
 &= \mu_X \sum_{k \in \mathbb{Z}} \alpha_k,
 \end{aligned}$$

qui ne dépend pas de t . De même, pour l'autocovariance, on a

$$\begin{aligned}
 \mathbb{E}((F_\alpha X)_s (F_\alpha X)_t) &= \left\langle \sum_{j \in \mathbb{Z}} \alpha_j X_{s-j}, \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k} \right\rangle \\
 &= \left\langle \lim_{n \rightarrow \infty} \sum_{j \in [-n, n]} \alpha_j X_{s-j}, \lim_{n \rightarrow \infty} \sum_{k \in [-n, n]} \alpha_k X_{t-k} \right\rangle \\
 &= \lim_{n \rightarrow \infty} \left\langle \sum_{j \in [-n, n]} \alpha_j X_{s-j}, \sum_{k \in [-n, n]} \alpha_k X_{t-k} \right\rangle \\
 &= \lim_{n \rightarrow \infty} \sum_{j \in [-n, n]} \sum_{k \in [-n, n]} \alpha_j \alpha_k \mathbb{E}(X_{s-j} X_{t-k}) \\
 &= \lim_{n \rightarrow \infty} \sum_{j \in [-n, n]} \sum_{k \in [-n, n]} \alpha_j \alpha_k (\gamma_X(s-j, t-k) + (\mu_X)^2) \\
 &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(t-s+j-k) + (\mu_X \sum_{k \in \mathbb{Z}} \alpha_k)^2,
 \end{aligned}$$

qui ne dépend que de $t-s$. Il en découle que $F_\alpha X$ est stationnaire. Notons que la double série $\sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h+j-k)$ est absolument convergente car

$$\sum_{j,k \in \mathbb{Z}} |\alpha_j \alpha_k \gamma_X(h+j-k)| \leq \gamma_X(0) \sum_{j,k \in \mathbb{Z}} |\alpha_j| |\alpha_k| = \gamma_X(0) \|\alpha\|_1^2 < \infty,$$

où on a utilisé le fait que $|\gamma_X(h)| \leq \gamma_X(0)$ pour tout $h \in \mathbb{Z}$ tiré du théorème 1.18. \square

Exemple 2.3 : Processus linéaires : filtrage d'un BB

Si $(Z_t)_{t \in \mathbb{Z}}$ est un $\text{BB}(0, \sigma^2)$, $\mu \in \mathbb{R}$, et $\alpha \in \ell^1(\mathbb{Z})$, alors le théorème 2.2 dit que $X = \mu + F_\alpha Z$ est un processus stationnaire de moyenne μ et d'autocovariance

$$\gamma_X(h) = \sigma^2 \sum_{j \in \mathbb{Z}} \alpha_j \alpha_{j+h}.$$

C'est l'image d'un BB par une application linéaire : on parle de **processus linéaire**.

Remarque 2.4 : Notation avec opérateur retard

Si $\alpha \in \ell^1(\mathbb{Z})$ alors la fonction

$$f(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k$$

est bien définie pour tout $z \in \mathbb{C}$ tel que $|z| = 1$ car

$$\left| \sum_{-n \leq k \leq n} \alpha_k z^k \right| \leq \sum_{-n \leq k \leq n} |\alpha_k| |z|^k = \sum_{-n \leq k \leq n} |\alpha_k| \xrightarrow{n \rightarrow \infty} \sum_{k \in \mathbb{Z}} |\alpha_k|$$

L'opérateur F_α s'écrit aussi $f(B)$ où B est l'opérateur retard :

$$f(B)X_t = \sum_{k \in \mathbb{Z}} \alpha_k B^k X_t = \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k}.$$

Si α est à support fini : $\text{card}\{k \in \mathbb{Z} : \alpha_k \neq 0\} < \infty$, alors $f(z)$ est bien définie pour tout $z \neq 0$, et prend des valeurs réelles si z et α sont réels. De plus f est un polynôme ssi $\{k \in \mathbb{Z} : \alpha_k \neq 0\}$ est une partie finie de \mathbb{N} , et dans ce cas $f(z)$ est bien définie pour tout $z \in \mathbb{C}$. Réciproquement, si une fonction $f : D \subset \mathbb{C} \rightarrow \mathbb{C}$ est développable en série de puissances de z sous la forme $f(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k$ alors cette série converge absolument lorsque $|z| = 1$ ssi $\alpha \in \ell^1(\mathbb{Z})$, et la notation $F_\alpha = f(B)$ fait sens.

Définition 2.5 : Causalité et inversibilité

Si Z est stationnaire, on dit que le filtre $X = \mu + F_\alpha Z$ de Z est un processus...

- **causal** lorsque $\alpha_k = 0$ pour tout $k < 0$ (X_t ne dépend pas du futur de Z_t). C'est le cas par exemple des processus MA(q), $q \geq 1$, qui vérifient

$$\forall t \in \mathbb{Z}, \quad X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}.$$

Plus généralement, les processus linéaires causaux (Z BB) avec $\alpha_0 = 1$ sont les processus MA(∞) (un processus MA(q) est un MA(q') pour tout $q' \geq q$);

- **inversible** lorsque Z est un processus causal de X , c'est-à-dire qu'il existe un $\beta \in \ell^1(\mathbb{Z})$ tel que $Z = F_\beta(X)$ avec $\beta_k = 0$ pour tout $k < 0$. C'est le cas par exemple des processus AR(p), $p \geq 1$, qui vérifient $Z = F(X)$ car

$$\forall t \in \mathbb{Z}, \quad X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = Z_t.$$

Plus généralement, les processus linéaires inversible (Z BB) avec $\alpha_0 = 1$ sont les processus AR(∞) (un AR(p) est un AR(p') pour tout $p' \geq p$).

2.3 Composition et convolution**Théorème 2.6 : Composition des opérateurs**

Avec les notations du théorème 2.2, si $\alpha, \beta \in \ell^1(\mathbb{Z})$ et si X est stationnaire, alors

$$F_\alpha(F_\beta X) = F_{\alpha * \beta} X \quad \text{où} \quad (\alpha * \beta)_k = \sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j}.$$

La quantité $(\alpha * \beta)_k$ fait sens pour tout $k \in \mathbb{Z}$ car

$$\left| \sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j} \right| \leq \sum_{j \in \mathbb{Z}} |\alpha_j| |\beta_{k-j}| \leq \|\beta\|_1 \sum_{j \in \mathbb{Z}} |\alpha_j| = \|\alpha\|_1 \|\beta\|_1 < \infty.$$

On dit que $\alpha * \beta$ est le **produit de convolution** de α et β . Le même argument montre que $\alpha * \beta$ existe dans $\ell^1(\mathbb{Z})$ dès que $\alpha \in \ell^1(\mathbb{Z})$ et $\beta \in \ell^\infty(\mathbb{Z})$, mais cela n'est pas exploité dans ce cours. Le produit de convolution vérifie les propriétés suivantes :

1. Le produit de convolution est commutatif et associatif :

$$\forall \alpha, \beta, \gamma \in \ell^1(\mathbb{Z}), \quad \alpha * \beta = \beta * \alpha \quad \text{et} \quad (\alpha * \beta) * \gamma = \alpha * (\beta * \gamma);$$

2. Le produit de convolution possède un élément neutre $e = \mathbf{1}_0$:

$$\forall \alpha \in \ell^1(\mathbb{Z}), \quad \alpha * e = \alpha;$$

3. Le produit de convolution est distributif :

$$\forall \alpha, \beta, \gamma \in \ell^1(\mathbb{Z}), \quad \forall \lambda \in \mathbb{R}, \quad (\lambda \alpha + \beta) * \gamma = (\lambda \alpha) * \gamma + \beta * \gamma = \lambda(\alpha * \gamma) + \beta * \gamma;$$

4. Le produit de convolution vérifie, pour tous $\alpha, \beta \in \ell^1(\mathbb{Z})$,

$$\|\alpha * \beta\|_1 = \sum_{k \in \mathbb{Z}} |(\alpha * \beta)_k| \leq \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} |\alpha_j| |\beta_{k-j}| = \sum_{j \in \mathbb{Z}} |\alpha_j| \sum_{k \in \mathbb{Z}} |\beta_{k-j}| = \|\alpha\|_1 \|\beta\|_1.$$

On dit que le produit de convolution $*$ fait de $\ell^1(\mathbb{Z})$ une algèbre de Banach.

Preuve du théorème 2.6. Lorsque α et β sont à support fini, la propriété est évidente car

$$\begin{aligned} (F_\alpha(F_\beta X))_t &= \sum_{j \in \mathbb{Z}} \alpha_j (F_\beta X)_{t-j} \\ &= \sum_{j \in \mathbb{Z}} \alpha_j \sum_{k \in \mathbb{Z}} \beta_k X_{t-j-k} \\ &\stackrel{k'=j+k}{=} \sum_{k' \in \mathbb{Z}} \left(\sum_{j \in \mathbb{Z}} \alpha_j \beta_{k'-j} \right) X_{t-k'} \\ &= (F_{\alpha * \beta} X)_t. \end{aligned}$$

Lorsque α et β sont à support infini, la commutation des sommes est licite grâce au théorème de Fubini–Tonelli (les séries sont presque sûrement absolument convergentes). Alternativement, lorsque α et β sont à support infini, on peut adopter une approche L^2 : on procède par troncature et on exploite l'hypothèse de sommabilité sur α et β . \square

Il est à présent naturel de se poser la question suivante :

- Peut-on défiltrer en utilisant à nouveau un filtre ? Autrement dit, à quelle condition sur $\alpha \in \ell^1(\mathbb{Z})$ existe-t-il $\beta \in \ell^1(\mathbb{Z})$ tel que $\alpha * \beta = e$ où $e = \mathbf{1}_0$ est l'élément neutre de la convolution ?

Il est bien connu en algèbre que si un tel élément β existe, alors il est unique^a. Il s'agit de l'inverse de α pour le produit de convolution $*$ et on le note $\beta = \alpha^{-1}$. Nous

allons voir que le calcul de l'inverse est lié à la série de puissances associée à α .

a. Si $\alpha * \beta = e = \alpha * \beta'$ alors $\beta = \beta * e = \beta * (\alpha * \beta') = (\beta * \alpha) * \beta' = e * \beta' = \beta'$.

Lemme 2.7 : Convolution et séries de puissances

Soit $(\mathcal{F}, +, \times)$ l'algèbre usuelle des fonctions continues définies sur le cercle unité $\{z \in \mathbb{C} : |z| = 1\}$ et à valeurs dans \mathbb{C} . Les propriétés suivantes sont vérifiées :

1. Pour tout $\alpha \in \ell^1(\mathbb{Z})$ et $z \in \mathbb{C}$ tel que $|z| = 1$, on a $(\alpha_k z^k)_{k \in \mathbb{Z}} \in \ell^1_{\mathbb{C}}(\mathbb{Z})$, c'est-à-dire que la série suivante converge absolument sur le cercle unité :

$$P_{\alpha}(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k;$$

2. L'application $\alpha \in \ell^1(\mathbb{Z}) \mapsto P_{\alpha} \in \mathcal{F}$ est un **homomorphisme d'algèbres** :

$$\forall \alpha, \beta, \gamma \in \ell^1(\mathbb{Z}), \forall \lambda \in \mathbb{C}, \quad P_{\alpha * \beta + \lambda \gamma}(z) = P_{\alpha}(z)P_{\beta}(z) + \lambda P_{\gamma}(z)$$

pour tout $z \in \mathbb{C}$ tel que les séries de puissances $P_{\alpha}(z)$, $P_{\beta}(z)$, et $P_{\gamma}(z)$ convergent absolument. En particulier, le produit de convolution dans $\ell^1(\mathbb{Z})$ est transformé en produit standard de fonctions dans \mathcal{F} par l'application $\alpha \mapsto P_{\alpha}$:

$$\forall \alpha, \beta \in \ell^1(\mathbb{Z}), \quad P_{\alpha * \beta} = P_{\alpha}P_{\beta}.$$

3. L'application $\alpha \in \ell^1(\mathbb{Z}) \mapsto P_{\alpha} \in \mathcal{F}$ est injective :

$$\forall \alpha, \beta \in \ell^1(\mathbb{Z}), \quad P_{\alpha} = P_{\beta} \Rightarrow \alpha = \beta;$$

Note : en français, la formule $P_{\alpha * \beta}(z) = P_{\alpha}(z)P_{\beta}(z)$ dit que les coefficients du **produit de Cauchy** de deux séries de puissances sont donnés par le produit de convolution des coefficients des deux séries. Cette liaison entre suites et fonctions est utile.

Note : l'élément neutre est envoyé sur l'élément neutre : $P_e(z) = 1$.

Éléments de preuve. La première propriété est évidente. La seconde propriété n'offre pas non plus de difficulté. Pour $\lambda = 0$ la seconde propriété découle d'un produit de Cauchy de séries de puissances absolument convergentes :

$$\begin{aligned} P_{\alpha}(z)P_{\beta}(z) &= \left(\sum_{k \in \mathbb{Z}} \alpha_k z^k \right) \left(\sum_{k' \in \mathbb{Z}} \beta_{k'} z^{k'} \right) \\ &= \sum_{k, k' \in \mathbb{Z}} \alpha_k \beta_{k'} z^{k+k'} \\ &\stackrel{h=k+k'}{=} \sum_{h \in \mathbb{Z}} \left(\sum_{k \in \mathbb{Z}} \alpha_k \beta_{h-k} \right) z^h \\ &= \sum_{h \in \mathbb{Z}} (\alpha * \beta)_h z^h \\ &= P_{\alpha * \beta}(z). \end{aligned}$$

Établissons la troisième propriété (injectivité). Si $P_{\alpha} = P_{\beta}$ pour $\alpha, \beta \in \ell^1(\mathbb{Z})$ alors d'après la seconde propriété on a $P_{\alpha - \beta} = 0$, c'est-à-dire que pour tout $z \in \mathbb{C}$ avec $|z| = 1$, on

a $\sum_{k \in \mathbb{Z}} \gamma_k z^k = 0$, où $\gamma_k = \alpha_k - \beta_k$. À présent, on note que les éléments de \mathcal{F} sont des fonctions continues et bornées sur le cercle unité, et sont dans l'espace L^2 des fonctions de carré intégrable sur le cercle unité. On rappelle que l'application

$$(f, g) \in L^2 \times L^2 \mapsto \langle f, g \rangle := \frac{1}{2\pi} \int_0^{2\pi} f(\theta) \overline{g(\theta)} d\theta$$

est un produit scalaire, qu'elle fait de L^2 un espace de Hilbert, et que les fonctions trigonométriques $\theta \in [0, 2\pi] \mapsto e^{i\theta k} \in \mathbb{C}$, $k \in \mathbb{Z}$, constituent une base hilbertienne. Le caractère orthonormé provient du fait que pour car pour tous $k, h \in \mathbb{Z}$,

$$\langle \theta \mapsto e^{i\theta k}, \theta \mapsto e^{i\theta h} \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{i\theta(h-k)} d\theta = \mathbf{1}_{h=k}.$$

Aussi, pour tout $h \in \mathbb{Z}$ fixé, en posant $z = e^{i\theta}$, il vient, grâce au théorème de Fubini–Tonelli,

$$0 = \langle \theta \mapsto \sum_{k \in \mathbb{Z}} \gamma_k e^{i\theta k}, \theta \mapsto e^{i\theta h} \rangle = \sum_{k \in \mathbb{Z}} \gamma_k \langle \theta \mapsto e^{i\theta k}, \theta \mapsto e^{i\theta h} \rangle = \sum_{k \in \mathbb{Z}} \gamma_k \mathbf{1}_{h=k} = \gamma_h.$$

□

La condition d'appartenance au cercle unité vient du fait que la série $P_\alpha(z)$ où $\alpha \in \ell^1(\mathbb{Z})$ peut comporter des puissances positives ou négatives arbitrairement grandes de z . La série $P_\alpha(z)$ est un polynôme ssi le support $\{k \in \mathbb{Z} : \alpha_k \neq 0\}$ de α est une partie finie de \mathbb{N} , et dans ce cas la série est une somme finie, qui a un sens pour tout $z \in \mathbb{C}$. Si le support de α contient une infinité de valeurs positives, alors la convergence de $P_\alpha(z)$ n'est pas garantie pour $|z| > 1$. Si le support de α contient des valeurs négatives, alors la convergence de $P_\alpha(z)$ est impossible en $z = 0$. Si le support de α contient une infinité de valeurs négatives alors la convergence de $P_\alpha(z)$ n'est pas garantie pour $|z| < 1$. Seul le cas $|z| = 1$ subsiste quelque soit le support de α , et la convergence (absolue) est alors garantie par $\alpha \in \ell^1(\mathbb{Z})$.

Le théorème suivant fait le lien entre propriétés d'algèbre et d'analyse.

Théorème 2.8 : Inversibilité pour la convolution et séries de puissances

Soit $\alpha \in \ell^1(\mathbb{Z})$ tel que $P_\alpha(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k$ est un polynôme, c'est-à-dire que α est à support fini et ≥ 0 . Les trois propriétés suivantes sont équivalentes :

1. α est inversible pour le produit de convolution dans $\ell^1(\mathbb{Z})$;
2. P_α n'a pas de racine de module 1 ;
3. $z \mapsto 1/P_\alpha(z)$ est développable en série de puissances de z , absolument convergente dans une couronne de \mathbb{C} contenant le cercle unité :

$$\frac{1}{P_\alpha(z)} = \sum_{k \in \mathbb{Z}} \beta_k z^k, \quad \beta \in \ell^1(\mathbb{Z}).$$

Lorsque ces propriétés ont lieu, alors $\alpha^{-1} = \beta$. De plus, si P_α n'a pas de racine de module < 1 alors le support de α^{-1} est ≥ 0 , c'est-à-dire que $\{k \in \mathbb{Z} : \alpha_k^{-1} \neq 0\} \subset \mathbb{N}$.

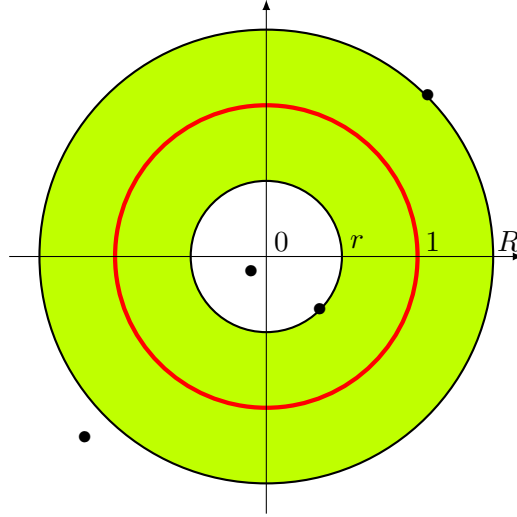


FIGURE 2.1 – Si le polynôme P_α n'a pas de racines de module 1, alors ses racines sont à l'extérieur d'une couronne qui contient le cercle unité.

Démonstration. Montrons que $2 \Rightarrow 3$. Supposons que P_α n'a pas de racine de module 1. Soit z_1, \dots, z_d les racines distinctes de P_α dans \mathbb{C} , de multiplicités respectives m_1, \dots, m_d , de sorte que $P_\alpha(z) = c(z - z_1)^{m_1} \cdots (z - z_d)^{m_d}$ avec $c \in \mathbb{C}$. Si

$$r := \max\{|z| \in \mathbb{C} : P_\alpha(z) = 0, |z| < 1\} = \max\{|z| : z \in \{z_1, \dots, z_d\} \cap \{z \in \mathbb{C} : |z| < 1\}\}$$

et

$$R := \min\{|z| \in \mathbb{C} : P_\alpha(z) = 0, |z| > 1\} = \min\{|z| : z \in \{z_1, \dots, z_d\} \cap \{z \in \mathbb{C} : |z| > 1\}\},$$

alors $r < 1 < R$ et les racines z_1, \dots, z_d de P_α sont à l'extérieur de la couronne (figure 2.3)

$$D(r, R) := \{z \in \mathbb{C} : r < |z| < R\},$$

qui contient le cercle unité $\{z \in \mathbb{C} : |z| = 1\}$. Afin de développer la **fraction rationnelle** $z \mapsto 1/P_\alpha(z)$ en série, on considère sa décomposition en éléments simples

$$\frac{1}{P_\alpha(z)} = \sum_{j=1}^d \sum_{k=1}^{m_j} \frac{c_{j,k}}{(z - z_j)^k},$$

où $c_{j,k} \in \mathbb{C}$ sont des constantes, pour tout $z \notin \{z_1, \dots, z_d\}$. Or pour tout $\alpha \in \mathbb{R}$ et $w \in \mathbb{C}$ tel que $|w| < 1$, une formule de Taylor donne (convergence absolue)

$$(1 + w)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} w^n \quad \text{où} \quad \binom{\alpha}{n} = \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!}.$$

(il s'agit d'une généralisation de la formule du binôme aux puissances quelconques). Donc si $|z_j| < 1$ et $|z| > |z_j|$ alors en posant $w = z_j/z$, on a (convergence absolue)

$$\frac{c_{j,k}}{(z - z_j)^k} = \frac{c_{j,k}}{z^k} \frac{1}{(1 - z_j/z)^k} = \frac{c_{j,k}}{z^k} \sum_{n=0}^{\infty} \binom{-k}{n} (-z_j/z)^n = \sum_{h \in \mathbb{Z}} \beta_{j,k,h} z^h$$

tandis que si $|z_j| > 1$ et $|z| < |z_j|$ en posant $w = z/z_j$, on a (convergence absolue)

$$\frac{c_{j,k}}{(z - z_j)^k} = \frac{c_{j,k}(-1)^k}{z_j^k} \frac{1}{(1 - z/z_j)^k} = \frac{c_{j,k}(-1)^k}{z_j^k} \sum_{n=0}^{\infty} \binom{-k}{n} (-z/z_j)^n = \sum_{h \in \mathbb{Z}} \beta_{j,k,h} z^h.$$

En combinant tout, on obtient enfin le développement en série de $1/P_\alpha(z)$:

$$\forall z \in D(r, R), \quad \frac{1}{P_\alpha(z)} = \sum_{j=1}^d \sum_{k=1}^{m_j} \sum_{h \in \mathbb{Z}} \beta_{j,k,h} z^h = \sum_{h \in \mathbb{Z}} \left(\sum_{j=1}^d \sum_{k=1}^{m_j} \beta_{j,k,h} \right) z^h = \sum_{h \in \mathbb{Z}} \beta_h z^h,$$

et la convergence est absolue, car une somme finie de séries de puissances absolument convergentes est une série de puissance absolument convergente. À présent, en prenant $z = 1 \in D(r, R)$ on a $\beta \in \ell^1_{\mathbb{C}}(\mathbb{Z})$. Enfin, pour établir que β est réel, on observe que comme α est réel, le polynôme P_α est à coefficients réels, donc l'ensemble de ses racines est stable par conjugaison, donc les coefficients $c_{j,k}, \beta_{j,k,h}, \beta_h$ sont en fait tous réels, ce qui assure finalement que β est bien à valeurs réelles : $\beta \in \ell^1(\mathbb{Z})$.

Montrons que $3 \Rightarrow 1$. Pour tout $z \in D(r, R)$, par produit de Cauchy (lemme 2.7),

$$P_e(z) = 1 = P_\alpha(z) \frac{1}{P_\alpha(z)} = P_\alpha(z) P_\beta(z) = P_{\alpha * \beta}(z),$$

ce qui donne $e = \alpha * \beta$ grâce au lemme 2.7. Donc α est inversible d'inverse $\alpha^{-1} = \beta$.

Montrons que $1 \Rightarrow 2$. Comme $\alpha^{-1} \in \ell^1(\mathbb{Z})$, le lemme 2.7 donne

$$P_{\alpha^{-1}}(z) P_\alpha(z) = P_{\alpha^{-1} * \alpha}(z) = P_e(z) = 1$$

pour tout $z \in \mathbb{C}$ tel que $|z| = 1$, et cela interdit à P_α d'avoir une racine de module 1.

Enfin, on observe que seules les racines de P_α de module inférieur à 1 contribuent à la partie de α^{-1} indicée par des indices négatifs. Par conséquent, si P_α n'a pas de racine de module ≤ 1 alors α^{-1} est porté par \mathbb{N} . \square

Exemple 2.9 : Séries géométriques

Examinons trois exemples liés à une série géométrique.

- Cas où $\alpha_0 = 2$, $\alpha_1 = -1$, et $\alpha_k = 0$ si $k \notin \{0, 1\}$. On a alors $P_\alpha(z) = 2 - z$, dont la seule racine 2 est de module > 1 . Le théorème 2.8 s'applique, et pour tout $z \in \mathbb{C}$ tel que $|z| < 2$,

$$\frac{1}{P_\alpha(z)} = \frac{1/2}{1 - z/2} = \frac{1}{2} \left(1 + \frac{z}{2} + \left(\frac{z}{2}\right)^2 + \dots \right).$$

Cela donne $(\alpha^{-1})_k = 2^{-(k+1)} \mathbf{1}_{k \geq 0}$, et on a bien $\alpha^{-1} \in \ell^1(\mathbb{Z})$.

- Cas où $\alpha_0 = 1$, $\alpha_1 = 2$, et $\alpha_k = 0$ si $k \notin \{0, 1\}$. On a alors $P_\alpha(z) = 1 - 2z$, dont la seule racine $1/2$ est de module < 1 . Le théorème 2.8 s'applique, et pour tout $z \in \mathbb{C}$ tel que $|z| > 1/2$ (autrement dit $|2z| > 1$),

$$\frac{1}{P_\alpha(z)} = \frac{1}{1 - 2z} = -\frac{1}{2z} \frac{1}{1 - 1/(2z)} = -\frac{1}{2z} \sum_{k=0}^{\infty} (2z)^{-k} = \sum_{k=1}^{\infty} 2^{-k} z^{-k}.$$

Cela donne $(\alpha^{-1})_k = 2^k \mathbf{1}_{k < 0}$, et on a bien $\alpha^{-1} \in \ell^1(\mathbb{Z})$.

- Cas où $\alpha_0 = 1$ et $\alpha_1 = -1$ et $\alpha_k = 0$ si $k \notin \{0, 1\}$. On a alors $P_\alpha(z) = 1 - z$ dont la seule racine est 1, qui est de module 1. Le théorème 2.8 ne s'applique

pas. Cependant, le développement en série géométrique

$$\frac{1}{P_\alpha(z)} = \frac{1}{1-z} = 1 + z + z^2 + \dots$$

est valable pour tout $z \in \mathbb{C}$ tel que $|z| < 1$, et suggère que $\alpha^{-1} = \mathbf{1}_{\mathbb{N}}$. Cette suite n'appartient pas à $\ell^1(\mathbb{Z})$. Cela suggère au passage que l'inverse d'un élément de $\ell^1(\mathbb{Z})$ peut exister dans $\ell^\infty(\mathbb{Z})$.

Remarque 2.10 : Calcul pratique de l'inverse

L'équation $\alpha * \beta = e$ en β est un système infini d'équations linéaires. La preuve du théorème 2.8 montre qu'on peut déterminer β en déterminant les racines de P_α puis en effectuant des développements en série. Mais cela peut s'avérer laborieux en pratique. Dans le cas où P_α n'a pas de racine de module ≤ 1 , on sait que β est porté par \mathbb{N} , ce qui rend le système d'équations $\alpha * \beta = e$ triangulaire, et donc facile à résoudre par récurrence. En effet, si d est le degré du polynôme P_α alors on obtient une identité entre séries de puissances de z absolument convergentes sur le cercle unité :

$$(\alpha_0 + \alpha_1 z + \dots + \alpha_d z^d)(\beta_0 + \beta_1 z + \dots) = 1.$$

En développant le membre de droite (produit de Cauchy) et en identifiant les coefficients pour chaque puissance grâce au lemme 2.7, on obtient le système triangulaire suivant qu'on peut résoudre de haut en bas :

$$\begin{cases} \alpha_0 \beta_0 & = 1 & (\text{d'où on tire } \beta_0) \\ \alpha_0 \beta_1 + \alpha_1 \beta_0 & = 0 & (\text{d'où on tire } \beta_1) \\ \alpha_0 \beta_2 + \alpha_1 \beta_1 + \alpha_2 \beta_0 & = 0 & (\text{d'où on tire } \beta_2) \\ & \vdots & \downarrow \end{cases}$$

Théorème 2.11 : Identifiabilité

On note \mathcal{S}_2 l'ensemble des processus stationnaires.

1. Si Z est $\text{BB}(0, \sigma^2)$ alors $\alpha \in \ell^1(\mathbb{Z}) \mapsto F_\alpha Z$ est injective :

$$\forall \alpha, \beta \in \ell^1(\mathbb{Z}), \quad F_\alpha Z = F_\beta Z \quad \Rightarrow \quad \alpha = \beta;$$

2. Si $\alpha \in \ell^1(\mathbb{Z})$ est inversible pour $*$ alors $X \in \mathcal{S}_2 \mapsto F_\alpha X$ est injective :

$$\forall X, Y \in \mathcal{S}_2, \quad F_\alpha X = F_\alpha Y \quad \Rightarrow \quad X = Y.$$

Démonstration.

1. Si $F_\alpha Z = F_\beta Z$ alors pour tous $s, t \in \mathbb{Z}$, $\sum_{k \in \mathbb{Z}} Z_s \alpha_k Z_{t-k} = \sum_{k \in \mathbb{Z}} Z_s \beta_k Z_{t-k}$. En prenant l'espérance et en utilisant le fait que $\gamma_Z(h) = \sigma^2 \mathbf{1}_{h=0}$ (car Z est un BB), il vient $\alpha_{t-s} = \beta_{t-s}$, et comme s, t sont quelconques, on a enfin $\alpha = \beta$. Note : on peut invoquer le fait que le produit scalaire dans L^2 est continu pour justifier l'interversion

- entre \mathbb{E} et Σ . Note : ce raisonnement n'est plus valable si on sait seulement que $F_\alpha Z \stackrel{d}{=} F_\beta Z$ où « $\stackrel{d}{=}$ » désigne l'égalité en loi, mais le redevient si $(Z, F_\alpha Z) \stackrel{d}{=} (Z, F_\beta Z)$;
2. Si $F_\alpha X = F_\alpha Y$ avec $\alpha \in \ell^1(\mathbb{Z})$ et $X, Y \in \mathcal{S}_2$, alors, grâce aux théorèmes 2.6 et 2.8,

$$X = F_{\alpha^{-1} * \alpha} X = F_{\alpha^{-1}}(F_\alpha X) = F_{\alpha^{-1}}(F_\alpha Y) = F_{\alpha^{-1} * \alpha} Y = Y.$$

□

2.4 Compléments

Remarque 2.12 : Filtrage faible d'un BB

Soit $(Z_t)_{t \in \mathbb{Z}}$ un BB $(0, \sigma^2)$ et $\alpha \in \ell^2(\mathbb{Z})$ c'est-à-dire que $\sum_{k \in \mathbb{Z}} |\alpha_k|^2 < \infty$. Alors pour tout $K \subset \mathbb{Z}$ fini, le théorème de Pythagore dans l'espace de Hilbert L^2 donne

$$\mathbb{E}(|\sum_{k \in K} \alpha_k Z_{t-k}|^2) = \sum_{j, k \in K} \alpha_j \alpha_k \gamma_Z(j-k) = \sigma^2 \sum_{k \in K} |\alpha_k|^2.$$

Comme $\alpha \in \ell^2(\mathbb{Z})$ le critère de Cauchy dans l'espace complet L^2 indique que la série $\sum_{k \in \mathbb{Z}} \alpha_k Z_{t-k}$ converge dans L^2 . On note $(F_\alpha Z)_t \in L^2$ sa somme, en cohérence avec les filtres habituels. D'autre part, la série est absolument convergente dans L^2 ssi $\alpha \in \ell^1(\mathbb{Z})$ car $\sum_{k \in \mathbb{Z}} \|\alpha_k Z_{t-k}\|_2 = \sum_{k \in \mathbb{Z}} |\alpha_k|$, et dans ce cas la convergence a lieu p.s. d'après le théorème de filtrage! Rappelons à ce sujet que $\ell^1(\mathbb{Z}) \subsetneq \ell^2(\mathbb{Z})$, et que l'inclusion est inversée pour les espaces L^p au sens où $L^2 \subset L^1$ (stricte en général).

Remarque 2.13 : Analyse de Fourier

Le théorème 2.6 est lié à l'associativité du produit de convolution car

$$F_\alpha(F_\beta X) = \alpha * (\beta * X) = (\alpha * \beta) * X = F_{\alpha * \beta} X$$

pour presque tout $\omega \in \Omega$. Le terme **filtrage** vient de la théorie du signal : comme la transformée de Fourier transforme produit de convolution en produit, on a $\widehat{\alpha * x} = \widehat{\alpha} \widehat{x}$ qui n'est rien d'autre qu'un **filtre spectral**^a. L'analyse de Fourier ω par ω des séries temporelles mène au **théorème de Wiener-Khinchine-Kolmogorov**, qui dépasse le cadre de ces notes. Nous feront appel à l'analyse de Fourier dans le chapitre 4 pour étudier la fonction d'autocovariance des processus stationnaires (elle est déterministe).

a. C'est un filtre passe-bande quand $\widehat{\alpha}$ est l'indicatrice d'un intervalle compact.

Remarque 2.14 : Analyse complexe

Il est possible d'étendre le théorème 2.8 d'inversibilité des filtres au delà des suites à support fini en introduisant de la technologie (analyse complexe) comme le développement en série de Laurent des fonctions méromorphes. Bien qu'amusant, cela n'est pas nécessaire pour l'étude, menée plus loin, des processus ARMA(p, q) avec $p, q < \infty$.

Processus ARMA (x1.5)

Mots-clés : processus AR/MA/ARMA ; causalité ; inversibilité.

Les **processus ARMA**(p, q) (Auto Regressive Moving Average) forment une classe de processus stationnaires paramétrés incluant à la fois un mécanisme d'autoregression linéaire (AR) d'ordre p , et de moyenne mobile ou ajustée (MA) d'ordre q .

Définition 3.1 : AR, MA, ARMA

Soient $p, q \in \mathbb{N}$, $\varphi \in \mathbb{R}^p$ et $\theta \in \mathbb{R}^q$ des coefficients fixés, et $(Z_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$. On dit que $(X_t)_{t \in \mathbb{Z}}$ est un **processus ARMA**(p, q), ou ARMA d'ordre (p, q) , lorsqu'il est stationnaire et vérifie l'équation de récurrence linéaire suivante ^a :

$$\forall t \in \mathbb{Z}, X_t = \sum_{k=1}^p \varphi_k X_{t-k} + Z_t + \sum_{k=1}^q \theta_k Z_{t-k}.$$

De plus :

- si $\theta \equiv 0$ ou $q = 0$ alors on dit qu'il s'agit d'un **processus AR**(p) ;
- si $\varphi \equiv 0$ ou $p = 0$ alors on dit qu'il s'agit d'un **processus MA**(q).

^a. Avec la convention $\sum_{k=0}^{-1} = \sum_{\emptyset} = 0$ utile quand $p = 0$ ou $q = 0$.

Les processus MA(q) existent toujours : ce sont les processus linéaires causaux à support fini. En revanche, l'existence des processus AR(p) n'est pas évidente, car le temps est indexé par \mathbb{Z} , ce qui rend délicate l'exploitation de la récurrence. Une idée naturelle consiste à utiliser récursivement l'équation pour produire une suite récurrente aléatoire. Bien qu'il n'y a pas de notion de condition initiale car le temps est indicé par \mathbb{Z} , cette approche conduit à la solution de AR(1) comme nous allons le voir plus loin. L'équation ARMA(p, q) peut aussi être vue comme un système triangulaire doublement infini d'équations indicées par \mathbb{Z} . Nous allons le résoudre dans ce chapitre, en reformulant ARMA(p, q) en terme de filtrage !

L'équation de récurrence ARMA(p, q) s'écrit également

$$\forall t \in \mathbb{Z}, X_t - (\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}) = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

ou encore

$$\Phi(B)X = \Theta(B)Z$$

où B est l'opérateur retard et où Φ et Θ désignent les polynômes

$$\Phi(z) = 1 - (\varphi_1 z + \dots + \varphi_p z^p) \quad \text{et} \quad \Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q.$$

L'écriture avec une **fraction rationnelle** suggère même la solution :

$$X \stackrel{?}{=} \frac{\Theta(B)}{\Phi(B)} Z = \frac{1 + \theta_1 B + \dots + \theta_q B^q}{1 - \varphi_1 B - \dots - \varphi_p B^p} Z \stackrel{?}{=} \sum_{k \in \mathbb{Z}} \beta_k B^k Z = F_\beta(Z).$$

L'équation de récurrence ARMA(p, q) s'écrit aussi avec des filtres

$$F_{\alpha_\varphi} X = F_{\alpha_\theta} Z,$$

où $(\alpha_\varphi)_{k \in \mathbb{Z}}$ et $(\alpha_\theta)_{k \in \mathbb{Z}}$ sont les suites à support fini définies par

$$(\alpha_\varphi)_h = \begin{cases} 1 & \text{si } h = 0 \\ -\varphi_h & \text{si } 1 \leq h \leq p \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad (\alpha_\theta)_h = \begin{cases} 1 & \text{si } h = 0 \\ \theta_h & \text{si } 1 \leq h \leq q \\ 0 & \text{sinon.} \end{cases}$$

Cette écriture avec des filtres suggère aussi la solution :

$$X \stackrel{?}{=} F_{\alpha_\varphi}^{-1} F_{\alpha_\theta} Z = F_{\alpha_\varphi^{-1} * \alpha_\theta} Z.$$

Notons que pour tout $z \in \mathbb{C}$, on a les identités suivantes (les sommes sont finies!)

$$\Phi(z) = P_{\alpha_\varphi}(z) = \sum_{h \in \mathbb{Z}} (\alpha_\varphi)_h z^h \quad \text{et} \quad \Theta(z) = P_{\alpha_\theta}(z) = \sum_{h \in \mathbb{Z}} (\alpha_\theta)_h z^h.$$

3.1 Processus MA et processus AR

Rappelons la définition des processus à moyenne mobile (Moving Average).

Définition 3.2 : Processus MA

Un processus $(X_t)_{t \in \mathbb{Z}}$ est MA(q) pour un entier $q \geq 0$ lorsqu'il est du second ordre, stationnaire, et solution de l'équation de récurrence

$$X_t = Z_t + \sum_{k=1}^q \theta_k Z_{t-k} = (F_{\alpha_\theta} Z)_t = (\Theta(B)Z)_t$$

où $(Z_t)_{t \in \mathbb{Z}}$ est un BB($0, \sigma^2$) et où $\theta \in \mathbb{R}^q$ est un vecteur fixe. On dit que q est l'ordre du processus et que (θ, σ^2) sont ses paramètres.

5→6

Le théorème 2.2 affirme que si $(X_t)_{t \in \mathbb{Z}}$ est MA(q) alors $(X_t)_{t \in \mathbb{Z}}$ est de moyenne nulle et de fonction d'autocovariance $\gamma_X(h) = \sigma^2 \left(\mathbf{1}_{h=0} + \theta_{|h|} \mathbf{1}_{1 \leq |h| \leq q} + \mathbf{1}_{1 \leq |h| < q} \sum_{k=1}^{q-|h|} \theta_k \theta_{k+|h|} \right)$.

Rappelons la définition des processus autoregressifs (Auto Regressive).

Définition 3.3 : Processus AR

Un processus $(X_t)_{t \in \mathbb{Z}}$ est AR(p) pour un entier $p \geq 0$ lorsqu'il est du second ordre,

stationnaire, et solution de l'équation de récurrence

$$X_t = Z_t + \sum_{k=1}^p \varphi_k X_{t-k}$$

en d'autres termes

$$Z_t = (F_{\alpha_\varphi} X)_t = (\Phi(B)X)_t$$

où $(Z_t)_{t \in \mathbb{Z}}$ est un BB $(0, \sigma^2)$ et où $\varphi \in \mathbb{R}^p$ est un vecteur fixe. On dit que p est l'ordre du processus et (φ, σ^2) ses paramètres.

Contrairement aux processus MA, l'existence des processus AR n'est pas évidente, car le temps étant indexé par \mathbb{Z} , il n'y a pas d'origine des temps qui permettrait d'exploiter facilement la récurrence. Le théorème suivant assure l'existence de processus AR(1).

Théorème 3.4 : Existence des processus AR(1)

- Si $|\varphi_1| = 1$ alors l'équation des AR(1) n'a pas de solution stationnaire ;
- Si $|\varphi_1| < 1$ alors la solution est donnée par le processus linéaire causal

$$X_t = \sum_{k=0}^{\infty} \varphi_1^k Z_{t-k},$$

de moyenne 0 et d'autocovariance $\gamma_X(h) = \sigma^2 \varphi_1^{|h|} / (1 - \varphi_1^2)$ pour tout $h \in \mathbb{Z}$. C'est un MA (∞) . C'est de plus l'unique solution stationnaire ;

- Si $|\varphi_1| > 1$ alors la solution est donnée par le processus linéaire non-causal

$$X_t = - \sum_{k=1}^{\infty} \varphi_1^{-k} Z_{t+k},$$

de moyenne 0 et d'autocovariance $\gamma_X(h) = \sigma^2 \varphi_1^{-|h|} / (\varphi_1^2 - 1)$ pour tout $h \in \mathbb{Z}$. C'est là encore l'unique solution stationnaire.

Remarque 3.5 : Infinité de solutions non stationnaires

Si X est solution de AR(1) $X_t = \varphi_1 X_{t-1} + Z_t$, on peut rechercher un processus Y tel que la somme $X + Y$ soit également solution. On a $X_t + Y_t = \varphi_1 (X_{t-1} + Y_{t-1}) + Z_t$ et comme $X_t = \varphi_1 X_{t-1} + Z_t$, on obtient alors $Y_t = \varphi_1 Y_{t-1}$. Ainsi, $Y_t = \varphi_1^t Y_0$ pour tout $t \in \mathbb{Z}$, avec Y_0 quelconque. Cela donne une infinité non dénombrable de solutions, mais ces solutions ne sont pas stationnaires en général (exercice!).

Démonstration. L'équation de récurrence s'écrit $X_t = Z_t + \varphi_1 X_{t-1}$ et donne par récurrence $X_t = \sum_{k=0}^r \varphi_1^k Z_{t-k} + \varphi_1^{r+1} X_{t-(r+1)}$ pour tout $r \geq 0$. Lorsque $|\varphi_1| < 1$ alors $\sum_{k=0}^{\infty} |\varphi_1|^k < \infty$, c'est-à-dire que $\alpha_\varphi \in \ell^1(\mathbb{Z})$, et par le théorème 2.2, $\sum_{k=0}^r \varphi_1^k Z_{t-k}$ converge p.s. et dans L^2 vers $\sum_{k=0}^{\infty} \varphi_1^k Z_{t-k}$, un processus du second ordre stationnaire, qui est un processus linéaire (exemple 2.3), dont on vérifie qu'il est solution de l'équation de récurrence :

$$\varphi_1 X_{t-1} = \varphi_1 \sum_{k=0}^{\infty} \varphi_1^k Z_{(t-1)-k} = \sum_{k=0}^{\infty} \varphi_1^{k+1} Z_{t-(k+1)} = \sum_{k=1}^{\infty} \varphi_1^k Z_{t-k} = X_t - Z_t.$$

Par ailleurs, si $(X_t)_{t \in \mathbb{Z}}$ est solution du second ordre stationnaire alors $\varphi_1^{r+1} X_{t-(r+1)} \rightarrow 0$ dans L^2 quand $r \rightarrow \infty$ car $|\varphi_1| < 1$, et on retrouve le processus linéaire $X = F_\alpha Z$ avec $\alpha_k = \varphi_1^k \mathbf{1}_{k \geq 0}$. L'autocovariance s'obtient en spécialisant celle des filtres du théorème 2.2 et plus précisément celle des processus linéaires de l'exemple 2.3 : pour tout $h \in \mathbb{Z}$ avec $h \geq 0$,

$$\gamma_X(h) = \sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_Z(h+j-k) = \sigma^2 \sum_{j \in \mathbb{Z}} \alpha_j \alpha_{h+j} = \sigma^2 \sum_{j=0}^{\infty} \varphi_1^j \varphi_1^{h+j} = \frac{\sigma^2}{1 - \varphi_1^h}.$$

Lorsque $|\varphi_1| > 1$ on procède de la même manière en utilisant la récurrence renversée $X_t = -\varphi_1^{-1} Z_{t+1} + \varphi_1^{-1} X_{t+1}$ en exploitant le fait que $|\varphi_1^{-1}| < 1$, ce qui conduit à la solution $X = F_\alpha Z$ avec $\alpha_k = \varphi_1^k \mathbf{1}_{k \leq -1}$. Ici aussi, on peut vérifier directement que ce processus $X_t = -\sum_{k=1}^{\infty} \varphi_1^{-k} Z_{t+k}$ est solution de l'équation de récurrence :

$$\varphi_1 X_{t-1} = -\varphi_1 \sum_{k=1}^{\infty} \varphi_1^{-k} Z_{(t-1)+k} = -\sum_{k=1}^{\infty} \varphi_1^{-(k-1)} Z_{t+(k-1)} = -\sum_{k=0}^{\infty} \varphi_1^{-k} Z_{t+k} = X_t - Z_t.$$

De même l'autocovariance s'obtient comme précédemment : pour tout $h \in \mathbb{Z}$ avec $h \leq 0$,

$$\gamma_X(h) = \sigma^2 \sum_{j \in \mathbb{Z}} \alpha_j \alpha_{h+j} = \sigma^2 \sum_{j \leq -1} \varphi_1^j \varphi_1^{h+j} = \sigma^2 \sum_{j=1}^{\infty} \varphi_1^{-j} \varphi_1^{-|h|-j} = \frac{\sigma^2 \varphi_1^{-|h|}}{\varphi_1^2 - 1}.$$

Lorsque $|\varphi_1| = 1$, alors s'il existait une solution stationnaire $(X_t)_{t \in \mathbb{Z}}$ alors

$$\begin{aligned} \mathbb{E}(|X_t - \varphi_1^{r+1} X_{t-(r+1)}|^2) &= \mathbb{E}\left(\left|\sum_{k=0}^r \varphi_1^k Z_{t-k}\right|^2\right) \\ &= \sum_{k=0}^r |\varphi_1|^{2k} \mathbb{E}(|Z_{t-k}|^2) \\ &= (r+1)\sigma^2 \end{aligned}$$

et d'autre part

$$\begin{aligned} \mathbb{E}(|X_t - \varphi_1^{r+1} X_{t-(r+1)}|^2) &= \gamma_X(0) + \gamma_X(0) - 2\gamma_X(t - (r+1) - t) \\ &= 2(\gamma_X(0) - \gamma_X(r+1)) \\ &\leq 4\gamma_X(0), \end{aligned}$$

ce qui n'est pas possible simultanément pour r assez grand. \square

3.2 Résolution de l'équation ARMA

Le théorème suivant généralise considérablement l'étude précédente des AR(1).

Théorème 3.6 : Existence des processus ARMA

Soient Φ et Θ les polynômes associés à l'équation ARMA(p, q).

1. Si Φ n'a pas de racine de module 1 alors ARMA(p, q) possède une unique solution stationnaire, donnée par le processus linéaire $F_\alpha Z$ où $\alpha = \alpha_\theta * \alpha_\varphi^{-1}$ où α_φ^{-1} est l'inverse de α_φ pour $*$ (existe : théorème 2.8) ;
2. Si ARMA(p, q) admet un processus linéaire $F_\alpha Z$ avec $\alpha \in \ell^1(\mathbb{Z})$ comme solution

alors toute racine de module 1 de Φ est également racine de Θ .

Preuve du théorème 3.6.

1. L'équation ARMA(p, q) s'écrit $F_{\alpha_\varphi} X = F_{\alpha_\theta} Z$. Supposons que Φ n'a pas de racine de module 1. Alors le théorème 2.8 assure l'existence de $\alpha_\varphi^{-1} \in \ell^1(\mathbb{Z})$. Donc $F_{\alpha_\varphi^{-1} * \alpha_\theta} Z$ est solution de ARMA(p, q) car le théorème 2.6 donne

$$F_{\alpha_\varphi}(F_{\alpha_\varphi^{-1} * \alpha_\theta} Z) = F_{\alpha_\varphi * (\alpha_\varphi^{-1} * \alpha_\theta)} Z = F_{(\alpha_\varphi * \alpha_\varphi^{-1}) * \alpha_\theta} Z = F_{\alpha_\theta} Z.$$

Le produit de convolution est commutatif, donc $\alpha = \alpha_\varphi^{-1} * \alpha_\theta = \alpha_\theta * \alpha_\varphi^{-1}$. Réciproquement, si X est solution de ARMA(p, q) alors, comme α_φ^{-1} existe,

$$X = F_{\alpha_\varphi^{-1} * \alpha_\varphi} X = F_{\alpha_\varphi^{-1}} F_{\alpha_\varphi} X = F_{\alpha_\varphi^{-1}} F_{\alpha_\theta} Z = F_{\alpha_\varphi^{-1} * \alpha_\theta} Z$$

par le théorème 2.6. On peut aussi utiliser le théorème 2.11 d'identifiabilité.

2. Supposons qu'un processus linéaire $F_\alpha Z$ avec $\alpha \in \ell^1(\mathbb{Z})$ soit solution de ARMA(p, q). Alors par théorème 2.6 de composition des filtres,

$$F_{\alpha_\theta} Z = F_{\alpha_\varphi}(F_\alpha Z) = F_{\alpha * \alpha_\varphi} Z,$$

puis par le théorème 2.11 d'identifiabilité, $\alpha * \alpha_\varphi = \alpha_\theta$. Comme $\alpha \in \ell^1(\mathbb{Z})$, par le lemme 2.7, pour tout $z \in \mathbb{C}$ tel que $|z| = 1$,

$$P_\alpha(z)\Phi(z) = P_\alpha(z)P_{\alpha_\varphi}(z) = P_{\alpha * \alpha_\varphi}(z) = P_{\alpha_\theta}(z) = \Theta(z).$$

Donc toute racine de Φ de module 1 est également racine de Θ . □

Remarque 3.7 : Fraction rationnelle d'un ARMA

La suite $\alpha = \alpha_\theta * \alpha_\varphi^{-1}$ de la solution $F_\alpha Z$ de ARMA(p, q) est liée au développement en série de la **fraction rationnelle** Θ/Φ . En effet, si Φ n'a pas de racine de module 1, alors la preuve du théorème 2.8 dit qu'il existe une couronne $D(r, R) = \{z \in \mathbb{C} : r < |z| < R\}$ contenant le cercle unité, ne contenant aucune racine de Φ , telle que

$$\forall z \in D(r, R), \quad \frac{1}{\Phi(z)} = \frac{1}{P_{\alpha_\varphi}(z)} = \sum_{k \in \mathbb{Z}} (\alpha_\varphi^{-1})_k z^k,$$

et ce développement en série de puissances est absolument convergent. Par conséquent,

en effectuant un produit de Cauchy, on obtient, pour tout $z \in D(r, R)$,

$$\begin{aligned} \frac{\Theta(z)}{\Phi(z)} &= \frac{P_{\alpha_\theta}(z)}{P_{\alpha_\varphi}(z)} = \left(\sum_{k \in \mathbb{Z}} (\alpha_\theta)_k z^k \right) \left(\sum_{k' \in \mathbb{Z}} (\alpha_\varphi^{-1})_{k'} z^{k'} \right) \\ &= \sum_{k, k' \in \mathbb{Z}} (\alpha_\theta)_k (\alpha_\varphi^{-1})_{k'} z^{k+k'} \\ &\stackrel{h=k+k'}{=} \sum_{h \in \mathbb{Z}} \left(\sum_{k \in \mathbb{Z}} (\alpha_\theta)_k (\alpha_\varphi^{-1})_{h-k} \right) z^h \\ &= \sum_{h \in \mathbb{Z}} (\alpha_\theta * \alpha_\varphi^{-1})_h z^h \\ &= \sum_{h \in \mathbb{Z}} \alpha_h z^h. \end{aligned}$$

Cela revient à exploiter la décomposition en éléments simples de Θ/Φ . Deux équations ARMA(p, q) qui admettent une solution stationnaire et qui ont la même **fraction rationnelle** ont la même solution. Cela permet d'éliminer les racines communes de Θ et Φ par simplification, et conduit à une équation ARMA **irréductible**.

Remarque 3.8 : Présence de racines de module 1

Si $\Phi = P_{\alpha_\varphi}$ possède des racines de module 1, alors l'équation ARMA(p, q) ne possède pas en général de solution stationnaire, comme le montre un AR(1) avec $|\varphi_1| = 1$.

3.3 ARMA causal et inversible

D'après la définition 2.5, un filtre $X = F_\alpha Z$ de Z est qualifié de **causal** ssi $\alpha_k = 0$ pour tout $k < 0$. Il est qualifié d'**inversible** ssi Z est un filtre causal de X , c'est-à-dire ssi il existe $\beta \in \ell^1(\mathbb{Z})$ tel que $Z = F_\beta X$ avec $\beta_k = 0$ pour tout $k < 0$.

Théorème 3.9 : Causalité et inversibilité des ARMA

Considérons une équation ARMA(p, q) $F_{\alpha_\varphi} X = F_{\alpha_\theta} Z$ et ses polynômes $\Phi = P_{\alpha_\varphi}$ et $\Theta = P_{\alpha_\theta}$. On suppose que Φ n'a pas de racines de module 1, ce qui assure l'existence d'une solution stationnaire unique $X = F_{\alpha_\varphi^{-1} * \alpha_\theta} Z$ (théorème 3.6). Alors

- la solution X est causale si Φ n'a pas de racine de module ≤ 1 ;
- la solution X est inversible si Θ n'a pas de racine de module ≤ 1 .

Note : comment se souvenir de la condition qui assure la causalité ? La méthode la plus efficace est de refaire la preuve, qui est courte et simple !

Remarque 3.10 : Causalité, inversibilité, et fraction rationnelle

La condition suffisante de causalité donnée par le théorème 3.9 ne constitue pas une condition nécessaire. En effet, comme l'unique solution stationnaire de l'équation

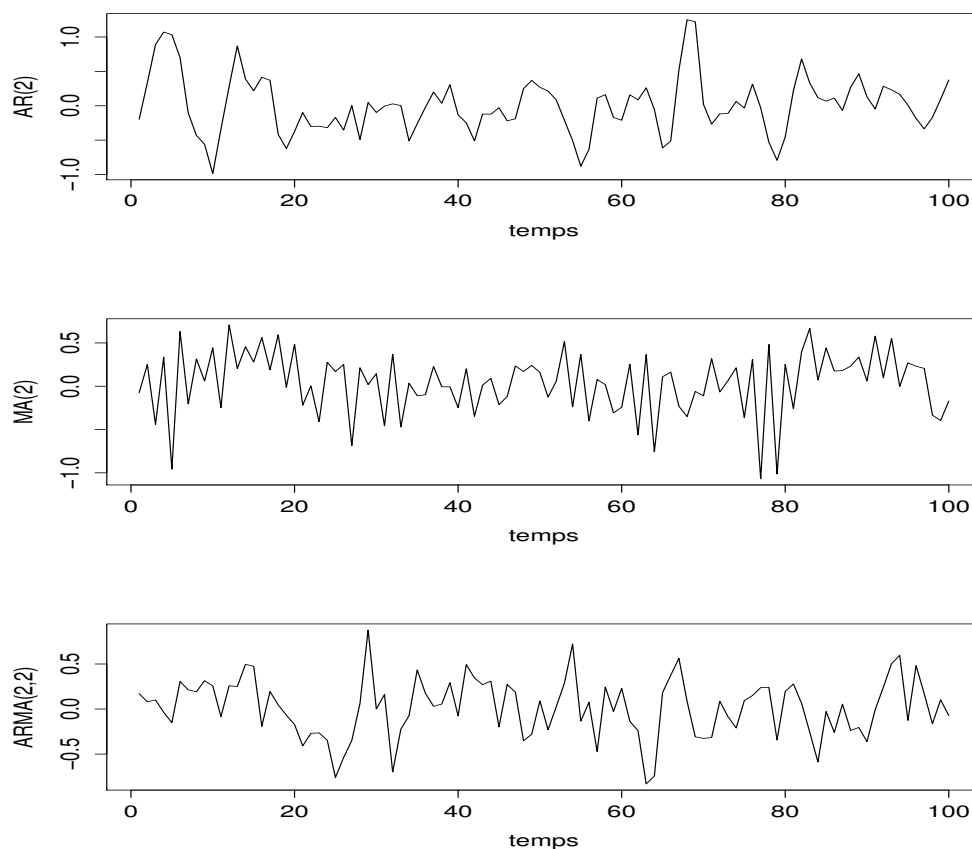


FIGURE 3.1 – Simulation de trajectoires de processus AR(2), MA(2), et ARMA(2,2).

Trajectoires simulées de processus AR, MA, ARMA, avec le code GNU-R ci-dessous.

```
require("stats") # en général déjà installé
serie = arima.sim(n=100,list(ar = c(0.8, -0.4), ma = c(-0.2, 0.2)), sd = 0.3)
plot.ts(serie,xlab="temps",ylab="ARMA(2,2)")
serie = arima.sim(n=100,list(ar = c(0.8, -0.4)), sd = 0.3)
plot.ts(serie,xlab="temps",ylab="AR(2)")
serie = arima.sim(n=100,list(ma = c(-0.2, 0.2)), sd = 0.3)
plot.ts(serie,xlab="temps",ylab="MA(2)")
```

ARMA ne dépend que de la fraction rationnelle Θ/Φ autour du cercle unité, c'est la forme irréductible $\tilde{\Theta}/\tilde{\Phi}$ qui importe, obtenue après simplification des racines communes à Θ et Φ (attentions aux multiplicités). Ainsi si $\tilde{\Phi}$ (respectivement $\tilde{\Theta}$) ne s'annule pas sur le disque unité alors la solution est causale (respectivement inversible).

Démonstration. Le processus X est causal ssi $\alpha_\varphi^{-1} * \alpha_\theta$ est porté par \mathbb{N} . Si Φ n'a pas de racine de module ≤ 1 , alors, d'après le théorème 2.8, α_φ^{-1} est porté par \mathbb{N} , et comme α_θ est porté par \mathbb{N} (Θ est un polynôme), on en déduit que $\alpha_\varphi^{-1} * \alpha_\theta$ est porté par \mathbb{N} .

Le processus X est inversible ssi α_θ est inversible pour $*$ dans $\ell^1(\mathbb{Z})$ et $\alpha_\theta^{-1} * \alpha_\varphi$ est porté par \mathbb{N} . Si Θ n'a pas de racine de module ≤ 1 , alors, d'après le théorème 2.8, α_θ^{-1} existe et est porté par \mathbb{N} , et comme α_φ est porté par \mathbb{N} (Φ est un polynôme), on en déduit que $\alpha_\theta^{-1} * \alpha_\varphi$ est porté par \mathbb{N} . \square

Remarque 3.11 : Résolution pratique des ARMA inversibles

La résolution pratique des ARMA(p, q) inversible peut être menée grâce à la remarque 2.10. En effet, on résout en ξ le système

$$\left(\sum_{k \in \mathbb{Z}} \xi_k z^k \right) (1 - \varphi_1 z - \dots - \varphi_p z^p) = 1 + \theta_1 z + \dots + \theta_q z^q$$

en identifiant les coefficients, ce qui donne dans le cas causal le système triangulaire

$$\begin{aligned} \xi_0 &= \theta_0 = 1 \\ -\xi_0 \varphi_1 + \xi_1 &= \theta_1 \\ -\xi_0 \varphi_2 - \xi_1 \varphi_1 + \xi_2 &= \theta_2 \\ &\vdots \end{aligned}$$

Note : si P_{α_φ} divise P_{α_θ} (c'est-à-dire que si z_j est racine de P_{α_φ} de multiplicité m_j alors elle est aussi racine de P_{α_θ} de multiplicité $\geq m_j$) alors P_ξ est un polynôme. Dans le cas contraire, P_ξ n'est pas un polynôme et contient des puissances de z de degré arbitrairement grand. Exemple : la solution de ARMA(1, 1) quand $|\varphi_1| < 1$ est donnée par $P_\xi(z) = 1 + \sum_{k=1}^{\infty} (\varphi_1 + \theta_1) \varphi_1^k z^k$, qui vérifie bien $(1 - \varphi_1 z) P_\xi(z) = 1 + \theta_1 z$. Ici, P_{α_φ} divise P_{α_θ} ssi $-1/\theta_1 = 1/\varphi_1$, c'est-à-dire ssi $P_\xi(z) = 1$ (et on a $P_{\alpha_\varphi} = P_{\alpha_\theta}$).

Exemple 3.12 : Tiré de l'examen final de janvier 2014

On considère un BB($0, \sigma^2$) Z et l'équation ARMA(1, 1)

$$X_t = Z_t + 2Z_{t-1} + (1/2)X_{t-1}.$$

- **Solution stationnaire, causalité, invertibilité.** Comme $\Phi(z) = 1 - (1/2)z$ a une seule racine, 2, de module > 1 . L'équation admet donc une unique solution stationnaire, causale, de la forme $X = F_\alpha Z$ (processus linéaire) avec $\alpha \in \ell^1(\mathbb{Z})$ et $\alpha_k = 0$ si $k < 0$. D'autre part, $\Theta(z) = 1 + 2z$ admet une unique racine, $-1/2$, de module < 1 , mais cela ne signifie pas que la solution n'est

pas inversible. Pour calculer α , comme $\alpha_k = 0$ pour tout $k < 0$, il suffit de résoudre le système triangulaire

$$(1 - (1/2)z)(\alpha_0 + \alpha_1 z + \dots) = 1 + 2z,$$

c'est-à-dire $\alpha_0 = 1$, $\alpha_1 - (1/2)\alpha_0 = 2$, $\alpha_k - (1/2)\alpha_{k-1} = 0$ pour tout $k \geq 2$, ce qui donne $\alpha_0 = 1$, $\alpha_1 = 2 + 1/2 = 5/2$, $\alpha_k = 5(1/2)^k$ pour tout $k \geq 1$. Alternativement, on peut développer la **fraction rationnelle** de l'équation : comme $|(1/2)z| = 1/2 < 1$ pour $|z| = 1$, on obtient

$$\begin{aligned} \frac{\Theta(z)}{\Phi(z)} &= \frac{1 + 2z}{1 - (1/2)z} \\ &= (1 + 2z) \sum_{k=0}^{\infty} (1/2)^k z^k \\ &= 1 + \sum_{k=1}^{\infty} ((1/2)^k + 2(1/2)^{k-1}) z^k \\ &= 1 + \sum_{k=1}^{\infty} 5(1/2)^k z^k; \end{aligned}$$

Il est agréable de vérifier qu'on a bien $\alpha \in \ell^1(\mathbb{Z})$ (série géométrique convergente).

Il est aussi judicieux de vérifier la formule pour $z = 1$, ce qui donne $6 = 6$;

- **Autocovariance.** Si α est tel que $X = F_\alpha Z$, alors, d'après le théorème de filtrage des processus stationnaires, pour $h \geq 0$,

$$\begin{aligned} \gamma_X(h) &= \sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_Z(h + j - k) \\ &= \sigma^2 \sum_{j=0}^{\infty} \alpha_j \alpha_{h+j} \\ &= \sigma^2 (\alpha_h + \sum_{j=1}^{\infty} 25(1/2)^{2j+h}) \\ &= \sigma^2 (\mathbf{1}_{h=0} + 5(1/2)^{|h|} \mathbf{1}_{|h| \neq 0} + (25/3)(1/2)^{|h|}). \end{aligned}$$

Ainsi, $\gamma_X(h) \rightarrow 0$ exponentiellement vite quand $|h| = |t - s| \rightarrow \infty$.

Exemple 3.13 : Inspiré des examens de septembre et novembre 2015

Cet exemple contient l'exemple 3.12. Soit l'équation ARMA(1,1)

$$X_t - \varphi X_{t-1} = Z_t + \theta Z_{t-1}$$

où φ et θ sont des réels. Sa fraction rationnelle est

$$R(z) := \frac{\Theta(z)}{\varphi(z)} = \frac{1 + \theta z}{1 - \varphi z}.$$

Le polynôme $\varphi(z) = 1 - \varphi z$ a une unique racine $1/\varphi$ de module $1/|\varphi|$. Le polynôme $\Theta(z) = 1 + \theta z$ a une unique racine $-1/\theta$ de module $1/|\theta|$.

— **Condition suffisante d'existence de solution stationnaire.** D'après le théorème 3.6, si φ ne s'annule pas sur le cercle unité (c'est-à-dire que $|\varphi| \neq 1$) alors il existe une unique solution stationnaire X qui est un filtre de Z , donné par $F_\psi Z$ où ψ est le développement en série de puissances de z sur le cercle unité de la fraction rationnelle $z \mapsto R(z)$.

— **Calcul de la solution quand $|\varphi| < 1$.** Si $|z| = 1$ alors $|\varphi z| < 1$ et $1/(1 - \varphi z) = \sum_{k=0}^{\infty} \varphi^k z^k$, d'où

$$\frac{1 + \theta z}{1 - \varphi z} = \sum_{k=0}^{\infty} \varphi^k z^k + \theta z \sum_{k=0}^{\infty} \varphi^k z^k = 1 + \sum_{k=1}^{\infty} \varphi^{k-1} (\varphi + \theta) z^k$$

d'où

$$\psi_k = \mathbf{1}_{k=0} + \varphi^{k-1} (\varphi + \theta) \mathbf{1}_{k>0}.$$

— **Calcul de la solution quand $|\varphi| > 1$.** Si $|z| = 1$ alors $|(\varphi z)^{-1}| < 1$ et

$$\frac{1}{1 - \varphi z} = -\frac{1}{\varphi z} \frac{1}{1 - (\varphi z)^{-1}} = -\frac{1}{\varphi z} \sum_{k=0}^{\infty} \varphi^{-k} z^{-k} = -\sum_{k=1}^{\infty} \varphi^{-k} z^{-k}.$$

et donc

$$\frac{1 + \theta z}{1 - \varphi z} = -\sum_{k=1}^{\infty} \varphi^{-k} z^{-k} - \theta z \sum_{k=1}^{\infty} \varphi^{-k} z^{-k} = -\theta \varphi^{-1} + \sum_{k=1}^{\infty} \varphi^{-k-1} (\varphi + \theta) z^{-k}$$

d'où

$$\psi_k = -\theta \varphi^{-1} \mathbf{1}_{k=0} + \varphi^{k-1} (\varphi + \theta) \mathbf{1}_{k<0}.$$

Dans les deux cas, $F_\varphi^{-1} \circ F_\theta = F_\psi$, pour tout $t \in \mathbb{Z}$,

$$X_t = \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k}$$

(note : $X = Z$ si $\varphi + \theta = 0$), et l'autocovariance vaut, pour tout $h \in \mathbb{Z}$,

$$\gamma_X(h) = \gamma_{F_\psi Z}(h) = \sum_{j,k \in \mathbb{Z}} \psi_j \psi_k \gamma_Z(h + j - k) = \sigma^2 \sum_{j \in \mathbb{Z}} \psi_j \psi_{h+j}.$$

Enfin, le processus X est causal lorsque Φ ne s'annule pas sur le disque unité (c'est-à-dire que $|\varphi| < 1$). Lorsque Θ ne s'annule pas sur le disque unité (c'est-à-dire que $|\theta| < 1$) alors X est inversible.

— **Étude du cas dégénéré $|\varphi| = 1$.** Le théorème 3.6 ne s'applique pas. On procède comme lors de l'étude de l'équation AR(1). Pour tous φ et θ , l'utilisation

réursive de $X_t = \varphi X_{t-1} + Z_t + \theta Z_{t-1}$ donne, pour tout $t \in \mathbb{Z}$ et tout $n \in \mathbb{N}$,

$$\begin{aligned} X_t &= \varphi X_{t-1} + Z_t + \theta Z_{t-1} \\ &= \varphi(\varphi X_{t-2} + Z_{t-1} + \theta Z_{t-2}) + X_{t-1} + Z_t + \theta Z_{t-1} \\ &= \varphi^2 X_{t-2} + Z_t + (\varphi + \theta)Z_{t-1} + \varphi\theta Z_{t-2} \\ &\vdots \\ &= \varphi^n X_{t-n} + Z_t + \sum_{k=1}^{n-1} \varphi^{k-1}(\varphi + \theta)Z_{t-k} + \varphi^{n-1}\theta Z_{t-n}. \end{aligned}$$

On retrouve l'expression de la solution stationnaire quand $|\varphi| < 1$ (une récurrence inversée conduit à la solution du cas $|\varphi| > 1$). D'autre part, si X est solution stationnaire, le calcul de la variance des deux membres de l'équation

$$X_t - \varphi^n X_{t-n} = Z_t + (\varphi + \theta) \sum_{k=1}^{n-1} \varphi^{k-1} Z_{t-k} + \varphi^{n-1} \theta Z_{t-n}$$

donne

$$(1 + \varphi^{2n})\gamma_X(0) - 2\varphi^n \gamma_X(n) = \sigma^2 + \sigma^2(\varphi + \theta)^2 \sum_{k=1}^{n-1} \varphi^{2(k-1)} + \sigma^2 \varphi^{2(n-1)} \theta^2.$$

Or si $|\varphi| = 1$ et $\varphi + \theta \neq 0$ alors on obtient une contradiction lorsque $n \rightarrow \infty$: le membre de gauche reste borné tandis que le membre de droite tend vers l'infini. Il n'y a donc pas de solution stationnaire lorsqu'à la fois $|\varphi| = 1$ et $\varphi + \theta \neq 0$. D'un autre côté si $\varphi + \theta = 0$ alors $X = Z$ est solution stationnaire.

3.4 Autocovariance d'un ARMA

D'après le résultat suivant, la fonction d'autocovariance d'un processus ARMA décroît exponentiellement : ces processus ne sont donc pas de bons modèles pour les **séries temporelles à mémoire longue**, dont l'autocovariance n'a pas ce comportement.

Théorème 3.14 : Autocovariance des ARMA

Si X est la solution stationnaire causale de $\text{ARMA}(p, q)$ fournie par le théorème 3.6 avec $\Phi = P_{\alpha_\varphi}$ sans racines de module ≤ 1 , alors l'autocovariance de X décroît exponentiellement : il existe des constantes $0 \leq \rho < 1$ et $C > 0$ telles que

$$\forall h \in \mathbb{Z}, \quad |\gamma_X(h)| \leq C\rho^{|h|}.$$

Démonstration. On a $X = F_\alpha Z$ avec $\alpha = \alpha_\varphi^{-1} * \alpha_\theta$. Comme P_{α_φ} n'a pas de racines de module ≤ 1 , les coefficients de α_φ^{-1} proviennent de la superposition d'un nombre fini de séries de puissances géométriques portées par \mathbb{N} . Il existe donc des constantes $C > 0$ et $0 \leq \rho < 1$ telles que $|\alpha_k| \leq C\rho^k$ pour tout $k \geq 0$, tandis que $\alpha_k = 0$ si $k < 0$ (car le processus est causal). À présent, l'exemple 2.3 donne, pour tout $h \geq 0$,

$$\gamma_X(h) = \sum_{j,k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_Z(h+j-k) = \sigma^2 \sum_{j \in \mathbb{Z}} \alpha_j \alpha_{j+h} = \sigma^2 \sum_{j=0}^{\infty} \alpha_j \alpha_{j+h}.$$

Il en découle que pour tout $h \in \mathbb{Z}$,

$$|\gamma_X(h)| \leq \sigma^2 \sum_{j=0}^{\infty} C^2 \rho^{2j+|h|} = \frac{\sigma^2 C^2 \rho^{|h|}}{1 - \rho^2} = \mathcal{O}(\rho^{|h|}).$$

□

6→7

3.5 Compléments

Remarque 3.15 : Linéarité des équations ARMA

Considérons l'équation ARMA(p, q) $F_{\alpha_\varphi} X = F_{\alpha_\theta} Z$ en $X = (X_t)_{t \in \mathbb{Z}}$, où $Z = (Z_t)_{t \in \mathbb{Z}}$ est un BB($0, \sigma^2$). Il est possible de concevoir X et Z comme deux variables aléatoires à valeurs dans $\mathbb{R}^{\mathbb{Z}}$, c'est-à-dire comme des vecteurs aléatoires infinis dont les composantes sont indicées par l'ensemble \mathbb{Z} . L'équation ARMA s'écrit alors matriciellement

$$M^{(\varphi)} X = M^{(\theta)} Z$$

où $M^{(\varphi)} = (M_{t,k}^{(\varphi)})_{t,k \in \mathbb{Z}}$ et $M^{(\theta)} = (M_{t,k}^{(\theta)})_{t,k \in \mathbb{Z}}$ sont des matrices infinies, triangulaires inférieures, bande-diagonales (de largeur de bande $p+1$ et $q+1$), de Toeplitz :

$$M_{t,k}^{(\varphi)} = \begin{cases} 1 & \text{si } k = t \\ -\varphi_k & \text{si } k \in \{t-1, \dots, t-p\} \\ 0 & \text{sinon} \end{cases}$$

et

$$M_{t,k}^{(\theta)} = \begin{cases} 1 & \text{si } k = t \\ \theta_k & \text{si } k \in \{t-1, \dots, t-q\} \\ 0 & \text{sinon} \end{cases}$$

(elles seraient inversibles si elles étaient finies!). Autrement dit, pour tout $t \in \mathbb{Z}$,

$$\sum_{k \in \mathbb{Z}} M_{t,k}^{(\varphi)} X_k = \sum_{j \in \mathbb{Z}} M_{t,k}^{(\theta)} Z_k.$$

On veut résoudre cette équation dans l'espace vectoriel \mathcal{S}_2 des processus stationnaires. La linéarité de l'équation fait que si X et X' sont deux solutions alors $X - X'$ est solution de l'équation sans second membre $M^{(\varphi)} X = 0$. En général cette équation sans second membre n'a, dans \mathcal{S}_2 , qu'une seule solution qui est identiquement nulle, d'où l'unicité (si existence), dans \mathcal{S}_2 , de la solution de l'équation avec second membre.

Remarque 3.16 : Filtrage exponentiel et AR(∞)

Soit $\lambda \in \mathbb{R}$ tel que $0 < |\lambda| < 1$ et soit $\varphi_k = -\lambda^k$ pour tout $k \in \mathbb{N}$. Soit Z un BB($0, \sigma^2$).

Recherchons un processus linéaire solution de l'équation AR(∞) suivante

$$X_t = Z_t + \sum_{k=1}^{\infty} \varphi_k X_{t-k} = Z_t - \sum_{k=1}^{\infty} \lambda^k X_{t-k}.$$

Ici $\Theta(z) = 1$ mais $\Phi(z) = 1 + \sum_{k=0}^{\infty} \lambda^k z^k = 1/(1 - \lambda z)$ (pour $|\lambda z| < 1$) n'est pas un polynôme. L'équation $\Phi(z) = 0$ en z n'a pas de solution. De plus, on a la formule $\Theta(z)/\Phi(z) = 1 - \lambda z$ ce qui suggère la solution $X_t = Z_t - \lambda Z_{t-1}$. On vérifie immédiatement que ce processus MA(1) est bien solution de notre équation AR(∞) :

$$Z_t - \sum_{k=1}^{\infty} \lambda^k X_{t-k} = Z_t - \sum_{k=1}^{\infty} \lambda^k Z_{t-k} + \sum_{k=1}^{\infty} \lambda^{k+1} Z_{t-1-k} = Z_t - \lambda Z_{t-1} = X_t.$$

Nous savions qu'un AR(1) causal est un MA(∞). Nous avons le cas dual ici : un AR(∞) causal est un MA(1). Plus généralement, l'analyse des ARMA(∞, ∞) peut être menée en étudiant les suites sommables à support infini au moyen de concepts et d'outils d'analyse complexe comme les fonctions méromorphes et les séries de Laurent.

Il n'y a aucune différence entre d'une part un théorème démontré du cours et d'autre part un exercice corrigé des travaux dirigés. Ici comme ailleurs, il est conseillé de chercher à comprendre plutôt que d'apprendre par cœur.

Analyse spectrale (x2)

Mots-clés : mesure spectrale ; densité spectrale ; représentation isospectrale.

Nous savons (théorème 1.18) que les fonctions d'autocovariance des processus stationnaires sont les fonctions **symétriques de type positif**, et cela provient de la structure Toeplitz de la covariance, elle même conséquence de la stationnarité. Le théorème de Herglotz (théorème 4.6), au centre de ce chapitre, affirme que les fonctions d'autocovariance sont les **coefficients de Fourier** des mesures positives finies paires. Cette bijection apporte un nouveau point de vue très utile sur les processus stationnaires.

Dans ce chapitre, nous effectuons l'analyse spectrale d'un signal, et ce signal est la fonction d'autocovariance, et non pas le processus stationnaire lui-même. Cela permet au passage de comprendre enfin le sens exact des termes «filtrage» et «bruit blanc». Mais nous allons voir que cela permet surtout de construire pour presque tout processus ARMA un processus ARMA causal et inversible de même autocovariance! (théorème 4.15).

4.1 Coefficients de Fourier

Les mesures positives finies sont une généralisation des fonctions positives intégrables. De même, les mesures signées finies sont une généralisation des fonctions intégrables. Plus précisément, si ν_1 et ν_2 sont deux **mesures positives finies** sur $[-\pi, \pi]$, on dit que leur différence $\nu = \nu_1 - \nu_2$ est une **mesure signée finie**. Pour toute fonction «test» mesurable bornée $\varphi : [-\pi, \pi] \rightarrow \mathbb{R}$, on a $\int \varphi d\nu = \int \varphi d\nu_1 - \int \varphi d\nu_2$. On a $L^1(\nu) = L^1(\nu_1) \cap L^1(\nu_2)$.

Pour tout $h \in \mathbb{Z}$, on note $\varphi_h : u \in [-\pi, \pi] \mapsto e^{ihu} \in \mathbb{C}$, où $i := (0, 1)$. Comme $|\varphi_h| = 1$, pour toute mesure signée finie ν sur $[-\pi, \pi]$, on a $\varphi_h \in L^1_{\mathbb{C}}([-\pi, \pi], d\nu)$. Dans ce chapitre, on utilise plusieurs fois le fait trigonométrique suivant : pour tous $j, k \in \mathbb{Z}$,

$$\int_{[-\pi, \pi]} e^{iju} e^{-iku} du = \int_{[-\pi, \pi]} e^{i(j-k)u} du = 2\pi \mathbf{1}_{j=k}.$$

Définition 4.1 : Coefficients de Fourier

Si ν est une mesure signée finie sur $[-\pi, \pi]$ alors la suite $(\widehat{\nu}(h))_{h \in \mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$ de ses **coefficients de Fourier** est définie par

$$\forall h \in \mathbb{Z}, \quad \widehat{\nu}(h) = \int_{[-\pi, \pi]} e^{ihu} \nu(du) = \int \varphi_h d\nu.$$

Si $f \in L^1([-\pi, \pi], du)$ et si ν est la mesure signée finie sur $[-\pi, \pi]$ de densité f , c'est-à-dire que $d\nu(u) = f(u)du$, alors on note $\widehat{f}(h) := \widehat{\nu}(h) = \int_{[-\pi, \pi]} e^{ihu} f(u) du$.

Chaque coefficient de Fourier «teste» ν sur une fonction trigonométrique. Le paramètre h est la «fréquence» de la fonction $u \mapsto \varphi_h(u) = e^{ihu} = \cos(hu) + i \sin(hu)$. Lorsque ν est positive, alors $\widehat{\nu}(0) = \nu([-\pi, \pi]) \geq 0$, ν est une probabilité ssi $\widehat{\nu}(0) = 1$, et

$$\forall h \in \mathbb{Z}, \quad |\widehat{\nu}(h)| \leq \widehat{\nu}(0) = \nu([-\pi, \pi]).$$

Exemple 4.2 : Localisation et principe d'incertitude

- si $\nu = \delta_0$ (localisé) alors $\forall h \in \mathbb{Z}, \widehat{\nu}(h) = 1$ (délocalisé) ;
- si $\nu = \text{Unif}([-\pi, \pi])$ (délocalisé), alors $\forall h \in \mathbb{Z}, \widehat{\nu}(h) = \mathbf{1}_{h=0}$ (localisé).

On voit sur ces deux exemples extrêmes que si ν est localisée dans l'espace alors $\widehat{\nu}$ ne l'est pas, et réciproquement. Plus généralement, au delà de ces deux exemples, on peut établir qu'il est impossible de localiser à la fois ν et $\widehat{\nu}$ au delà d'un certain seuil absolu, phénomène connu sous le nom de principe d'incertitude en analyse harmonique^a.

a. En mécanique quantique, la position et l'impulsion sont liées à des opérateurs associés par une transformation de Fourier, ce qui conduit au principe d'incertitude de Heisenberg.

Le théorème suivant affirme que les coefficients de Fourier caractérisent la mesure.

Théorème 4.3 : Injectivité de la suite des coefficients de Fourier

Si ν_1 et ν_2 sont deux mesures signées finies sur $[-\pi, \pi]$ vérifiant $\widehat{\nu}_1 = \widehat{\nu}_2$ alors $\nu_1 = \nu_2$.

Démonstration. L'intervalle $K = [-\pi, \pi]$ est compact. Soit $\mathcal{C}_{\mathbb{C}}(K)$ l'espace de Banach des fonctions continues $f : K \rightarrow \mathbb{C}$ muni de la norme uniforme $\|f\|_{\infty} = \sup_{t \in [-\pi, \pi]} |f(t)|$. Le théorème de densité de Stone-Weierstrass affirme que $\mathcal{A} \subset \mathcal{C}_{\mathbb{C}}(K)$ est dense lorsque

- \mathcal{A} est une algèbre (stable par combinaisons linéaires finies et multiplication) ;
- \mathcal{A} est stable par conjugaison, et contient les constantes ;
- \mathcal{A} sépare les points : $\forall x \neq y \in K, \exists f \in \mathcal{A} : f(x) \neq f(y)$.

L'algèbre engendrée par les fonctions trigonométriques $\{u \in [-\pi, \pi] \mapsto e^{ihu} : h \in \mathbb{Z}\}$, dont les éléments sont appelés **polynômes trigonométriques**, vérifie ces conditions. Donc si $\widehat{\nu}_1 = \widehat{\nu}_2$ alors ν_1 et ν_2 coïncident sur \mathcal{A} en tant que formes linéaires continues et donc sur $\mathcal{C}_{\mathbb{C}}(K)$ par densité. En effet, pour tout $f \in \mathcal{C}_{\mathbb{C}}(K)$ et tout $\varepsilon > 0$, la densité nous dit qu'il existe $f_{\varepsilon} \in \mathcal{A}$ tel que $\|f - f_{\varepsilon}\|_{\infty} \leq \varepsilon$:

$$\int f d(\nu_1 - \nu_2) = \underbrace{\int (f - f_{\varepsilon}) d(\nu_1 - \nu_2)}_{\mathcal{O}(\varepsilon)} + \underbrace{\int f_{\varepsilon} d(\nu_1 - \nu_2)}_{=0 \text{ car } \widehat{\nu}_1 = \widehat{\nu}_2},$$

où $\int g d(\nu_1 - \nu_2) := \int g d\nu_1 - \int g d\nu_2$. Comme ε est arbitrairement petit, le tour est joué. \square

Intuitivement, on s'attend à un lien entre la régularité de ν et le comportement à l'infini de $\widehat{\nu}$ (hautes fréquences). Dans cet esprit, le théorème suivant affirme qu'une mesure à coefficients de Fourier sommables possède toujours une densité.

Théorème 4.4 : Critère de densité et formule d'inversion

Si ν est une mesure signée finie sur $[-\pi, \pi]$ telle que $\widehat{\nu} \in \ell^1_{\mathbb{C}}(\mathbb{Z})$ alors ν admet une densité $f = \frac{d\nu}{du} \in L^1([-\pi, \pi], du)$ continue donnée pour presque tout $u \in [-\pi, \pi]$ par la série absolument convergente suivante (appelée **formule d'inversion**) :

$$f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-iuh} \widehat{\nu}(h).$$

De plus $f \geq 0$ ssi ν est une mesure positive.

Note : exemple de non sommabilité : si $\nu = \delta_0$ alors $\widehat{\nu} = 1 \notin \ell^1(\mathbb{Z})$.

Démonstration. Comme $\widehat{\nu} \in \ell^1_{\mathbb{C}}(\mathbb{Z})$, la série définissant $f(u)$ est absolument convergente, uniformément bornée par $\|\widehat{\nu}\|_1$, et $f \in L^1([-\pi, \pi], du)$. La fonction f définit une mesure signée $f(u)du$ finie sur $[-\pi, \pi]$. Notons que f est réelle car $\widehat{\nu}$ est **hermitienne** :

$$\forall h \in \mathbb{Z}, \quad \overline{\widehat{\nu}(h)} = \widehat{\nu}(-h).$$

De plus la mesure ν est paire si et seulement si sa densité f est paire, et alors la transformée de Fourier $\widehat{\nu} = \widehat{f}$ est paire et réelle. Enfin, le théorème de Fubini–Tonelli donne

$$\forall h \in \mathbb{Z}, \quad \widehat{f}(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iuh} \sum_{k \in \mathbb{Z}} e^{-iuk} \widehat{\nu}(k) du = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \widehat{\nu}(k) \underbrace{\int_{-\pi}^{\pi} e^{iu(h-k)} du}_{2\pi \mathbf{1}_{h=k}} = \widehat{\nu}(h).$$

Donc $\widehat{f} = \widehat{\nu}$, et donc $f(u)du = d\nu(u)$ par le théorème 4.3 d'injectivité. La somme de la série définissant f est la limite uniforme de fonctions continues sur $[-\pi, \pi]$ (série de fonctions normalement convergente), c'est donc une fonction continue sur $[-\pi, \pi]$. Enfin, la fonction f est ≥ 0 presque partout en tant que densité, et donc partout par continuité. \square

Théorème 4.5 : Paul Lévy – Critère de convergence

Si $(\nu_n)_{n \geq 1}$ est une suite de **mesures de probabilité** sur $[-\pi, \pi]$ telles que

$$\forall h \in \mathbb{Z}, \quad \lim_{n \rightarrow \infty} \widehat{\nu}_n(h) = \varphi(h) \in \mathbb{C},$$

alors il existe une **mesure de probabilité** ν sur $[-\pi, \pi]$ telle que

$$\varphi = \widehat{\nu},$$

et $\lim_{n \rightarrow \infty} \nu_n = \nu$ étroitement : pour toute $g : [-\pi, \pi] \rightarrow \mathbb{R}$ continue et bornée,

$$\lim_{n \rightarrow \infty} \int g d\nu_n = \int g d\nu.$$

Preuve (facultative). On prolonge chaque ν_n en une mesure de probabilité sur \mathbb{R} . Comme $[-\pi, \pi]$ est compact et contient $\text{supp}(\nu_n)$ pour tout $n \geq 1$, la suite $(\nu_n)_{n \geq 1}$ est tendue : pour tout $\varepsilon > 0$, il existe un compact K tel que $\sup_n \nu_n(K^c) \leq \varepsilon$. Il en découle grâce au théorème de Prohorov que $(\nu_n)_{n \geq 1}$ est relativement compacte pour la topologie de la

convergence étroite (qui est métrisable). Ainsi, $(\nu_n)_{n \geq 1}$ converge étroitement ssi elle admet une unique valeur d'adhérence pour la convergence étroite (qui est nécessairement une mesure de probabilité portée par $[-\pi, \pi]$). Or cette unicité découle de l'hypothèse faite sur les transformées de Fourier grâce au théorème 4.3 d'injectivité : les limites des sous-suites de $(\nu_n)_{n \geq 1}$ qui convergent étroitement ont toutes la même transformée de Fourier φ . \square

4.2 Théorème de Herglotz

Théorème 4.6 : Herglotz – Mesure spectrale et densité spectrale

Pour tout $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ les deux propriétés suivantes sont équivalentes :

1. γ est symétrique de type positif :

$$\forall h \in \mathbb{Z}, \quad \gamma(-h) = \gamma(h) \quad \text{et} \quad \forall n \geq 1, \forall v \in \mathbb{R}^n, \quad \sum_{j=1}^n \sum_{k=1}^n v_j v_k \gamma(j-k) \geq 0;$$

2. $(\gamma(h))_{h \in \mathbb{Z}}$ sont les coefficient de Fourier d'une **mesure positive finie** ν sur $[-\pi, \pi]$. En particulier ν est **paire** et $\gamma(0) = \nu([-\pi, \pi])$ car

$$\forall h \in \mathbb{Z}, \quad \gamma(h) = \int_{[-\pi, \pi]} e^{ihu} \nu(du) \in \mathbb{R}.$$

Lorsque ces propriétés ont lieu, la mesure ν est unique, et est appelée **mesure spectrale** (sa densité $f = d\nu/du$, si elle existe, est appelée **densité spectrale**).

Lorsque ces propriétés ont lieu, et si de plus $\gamma \in \ell^1(\mathbb{Z})$, alors la densité spectrale $f = d\nu/du$ existe et est donnée par la série (uniformément convergente en u)

$$\forall u \in [-\pi, \pi], \quad f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ihu},$$

et cette fonction est paire et continue.

Note : la parité de ν fait écho à la nature réelle symétrique de γ

Note : comme f est paire on peut remplacer e^{-ihu} par e^{ihu} dans la formule.

Note : la somme qui définit f est réelle car comme γ est **hermitienne** (car réelle symétrique), on en déduit que pour tout $u \in [-\pi, \pi]$,

$$\overline{\sum_{h \in \mathbb{Z}} \gamma(h) e^{-ihu}} = \sum_{h \in \mathbb{Z}} \gamma(-h) e^{ihu} = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ihu}.$$

Exemple 4.7 : Densité spectrale d'un MA(1)

Si $X_t = Z_t + \theta Z_{t-1}$ est un MA(1) avec $Z \sim \text{BB}(0, \sigma^2)$, alors son autocovariance est donnée par $\gamma_X = \sigma^2(1 + \theta^2)\mathbf{1}_0 + \sigma^2\theta\mathbf{1}_{\pm 1} \in \ell^1(\mathbb{Z})$, et sa densité spectrale est donnée par $f(u) = \frac{1}{2\pi} \sum_{h \in \{0, \pm 1\}} \gamma_X(h) e^{-ihu} = \frac{\sigma^2}{2\pi} (1 + \theta^2 + 2\theta \cos(u))$ pour tout $u \in [-\pi, \pi]$.

Démonstration du théorème 4.6. Montrons que $2 \Rightarrow 1$. Supposons que $\gamma = \hat{\nu}$ pour une mesure positive finie ν sur $[-\pi, \pi]$. Comme $\hat{\nu}$ est hermitienne, dire que $\hat{\nu}$ est réelle est

équivalent à dire que $\widehat{\nu}$ est symétrique. Or comme la fonction γ est par définition réelle et $\widehat{\nu} = \gamma$, on en déduit que γ est symétrique (ce qui implique que ν est paire). Pour établir que γ est de type positif, on observe que pour tous $n \geq 1$ et $v \in \mathbb{R}^n$,

$$\sum_{j=1}^n \sum_{k=1}^n v_j v_k \gamma(j-k) = \int_{-\pi}^{\pi} \sum_{j=1}^n \sum_{k=1}^n v_j e^{iju} v_k e^{-iku} \nu(du) = \int_{-\pi}^{\pi} \left| \sum_{j=1}^n v_j e^{iju} \right|^2 \nu(du) \geq 0.$$

Montrons que $1 \Rightarrow 2$. Soit $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ symétrique de type positif. Dans la définition de positivité, on peut remplacer $v \in \mathbb{R}^n$ par $v \in \mathbb{C}^n$ et $v_j v_k$ par $v_j \overline{v_k}$, car si $v, v' \in \mathbb{R}^n$,

$$\sum_{j,k=1}^n (v_j + i v'_j) \overline{(v_k + i v'_k)} \gamma(j-k) = \underbrace{\sum_{j,k=1}^n (v_j v_k + v'_j v'_k) \gamma(j-k)}_{\geq 0} + i \underbrace{\sum_{j,k=1}^n (v'_j v_k - v_j v'_k) \gamma(j-k)}_{=0}.$$

Recherchons la mesure positive finie ν . On ne peut pas utiliser le théorème 4.4 car la condition $\gamma \in \ell_{\mathbb{C}}^1(\mathbb{Z})$ n'est pas forcément vérifiée. On peut malgré tout s'inspirer du théorème 4.4 : pour tout $n \geq 1$ et $u \in [-\pi, \pi]$, soit,

$$f_n(u) = \frac{1}{2\pi} \sum_{j=1}^n e^{-iju} \frac{1}{n} \sum_{k=1}^n e^{iku} \gamma(j-k) = \frac{1}{2\pi} \sum_{1 \leq j, k \leq n} \frac{1}{n} e^{-i(j-k)u} \gamma(j-k).$$

Comme γ est de type positif, on a $f_n(u) \geq 0$ (prendre $v_j = e^{-iju}$). Comme la quantité sous le signe somme ne dépend que de $\ell = j - k$, on a (dessin !)

$$f_n(u) = \frac{1}{2\pi} \sum_{\ell=-n+1}^{n-1} \frac{\text{card}\{(j, k) : j - k = \ell\}}{n} e^{-i\ell u} \gamma(\ell) = \frac{1}{2\pi} \sum_{\ell=-n+1}^{n-1} \frac{n - |\ell|}{n} e^{-i\ell u} \gamma(\ell).$$

Soit ν_n la mesure positive de densité $f_n = \frac{d\nu_n}{du}$. On a, pour tout $h \in \mathbb{Z}$ avec $|h| \leq n$,

$$\begin{aligned} \widehat{\nu_n}(h) &= \int_{-\pi}^{\pi} e^{ihu} f_n(u) du \\ &= \frac{1}{2\pi} \sum_{\ell=-n+1}^{n-1} \left(1 - \frac{|\ell|}{n}\right) \gamma(\ell) \underbrace{\int_{-\pi}^{\pi} e^{ihu} e^{-i\ell u} du}_{=2\pi \mathbf{1}_{h=\ell}} \\ &= \left(1 - \frac{|h|}{n}\right) \gamma(h). \end{aligned}$$

En posant $\widehat{\nu_n}(h) = 0$ si $|h| > n$, on a $\lim_{n \rightarrow \infty} \widehat{\nu_n} = \gamma$ ponctuellement sur \mathbb{Z} . Notons que $\nu_n([-\pi, \pi]) = \widehat{\nu_n}(0) = \gamma(0)$. Si $\gamma(0) = 0$ alors $\gamma = 0$ et on peut prendre $\nu = 0$. Si $\gamma(0) > 0$, alors on se ramène par dilatation au cas où $\gamma(0) = 1$ et on obtient par le théorème 4.5 de Paul Lévy qu'il existe une mesure de probabilité ν telle que $\gamma = \widehat{\nu}$ et $\nu_n \rightarrow \nu$ étroitement.

L'unicité de ν découle du théorème 4.3, et la densité du théorème 4.4. \square

4.3 Mesure spectrale d'un processus stationnaire

Définition 4.8 : Mesure/densité spectrale d'un processus stationnaire

Si $X = (X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire d'autocovariance γ_X alors la **mesure spectrale** de X , notée ν_X , est la mesure positive de transformée de Fourier γ_X fournie par le théorème 4.6 de Herglotz. De même, si $\gamma_X \in \ell^1(\mathbb{Z})$, alors la **densité spectrale**

de X , notée f_X , est la densité de ν_X fournie par le théorème 4.6 de Herglotz.

Théorème 4.9 : Condition pratique pour être autocovariance

Si $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ est une fonction vérifiant les deux propriétés

- (symétrie) $\forall h \in \mathbb{Z} : \gamma(-h) = \gamma(h)$;
- (sommabilité) $\gamma \in \ell^1(\mathbb{Z})$ c'est-à-dire $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$;

alors γ est la fonction d'autocovariance d'un processus stationnaire ssi

$$\forall u \in [-\pi, \pi], \quad f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ihu} \geq 0$$

De plus, dans ce cas, le processus admet pour densité spectrale f .

Note : comme f est paire on peut remplacer e^{-ihu} par e^{ihu} dans la formule.

Démonstration. Si $\gamma \in \ell^1(\mathbb{Z})$ est l'autocovariance d'un processus stationnaire, alors la seconde partie du théorème 4.6 de Herglotz assure que γ possède une densité spectrale, donnée par $f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-iuh} \gamma(h)$ pour tout $u \in [-\pi, \pi]$, et $f \geq 0$.

Réciproquement, si $f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-iuh} \gamma(h) \geq 0$ pour tout $u \in [-\pi, \pi]$, alors $0 \leq f \in L^1([-\pi, \pi])$ car $\gamma \in \ell^1(\mathbb{Z})$ et f est donc la densité d'une mesure positive finie sur $[-\pi, \pi]$. De plus, pour tout $h \in \mathbb{Z}$, on a

$$\widehat{f}(h) = \int_{-\pi}^{\pi} e^{ihu} f(u) du = \frac{1}{2\pi} \sum_{h' \in \mathbb{Z}} \int_{-\pi}^{\pi} e^{ihu} e^{ih'u} \gamma(h') du = \gamma(h).$$

Donc par le théorème 4.6 de Herglotz, γ est symétrique de type positif, et par le théorème 1.18, γ est l'autocovariance d'un processus stationnaire de densité spectrale f . \square

Exemple 4.10 : Processus harmonique

Considérons à nouveau le **processus harmonique** de l'exemple 1.12, défini par la formule $X_t = A \cos(\theta t) + B \sin(\theta t)$ où A et B sont des variables aléatoires non corrélées de moyenne 0 et de variance σ^2 , et où $\theta \in [-\pi, \pi]$ est une constante. Nous avons calculé que $\gamma_X(h) = \sigma^2 \cos(\theta h)$. En ce qui concerne la mesure spectrale, on a

$$\gamma_X(h) = \sigma^2 \cos(\theta h) = \frac{\sigma^2}{2} (e^{-i\theta h} + e^{i\theta h}) = \int_{[-\pi, \pi]} e^{iuh} \mu(du)$$

pour $\mu = \frac{\sigma^2}{2} (\delta_{-\theta} + \delta_{\theta})$ (mesure de Bernoulli). La présence d'atomes fait que la densité spectrale n'existe pas. Donc $\gamma_X \notin \ell^1(\mathbb{Z})$ car sinon cela contredirait le théorème de Herglotz. Notons d'ailleurs que la série $\sum_{h \in \mathbb{Z}} |\gamma_X(h)|$ diverge car $|\cos(\theta h)| \not\rightarrow 0$ (facile à voir quand $\theta \in \mathbb{Q}\pi$) et donc $\gamma_X \notin \ell^1(\mathbb{Z})$.

Remarque 4.11 : Localisation spectrale et mémoire

Si $\nu_X = \delta_0$ alors $\gamma_X = 1$ et donc X est à mémoire longue, tandis que si $\nu_X = du$ alors $\gamma_X = 2\pi\mathbf{1}_0$ et donc X est à mémoire courte (cf. ci-dessous l'exemple du bruit blanc).

Le théorème suivant montre en particulier que le filtrage a pour effet de modifier les poids affectés aux différentes fréquences, ce qui explique son nom.

Théorème 4.12 : Spectre et filtrage

Si X est stationnaire de mesure spectrale ν_X alors pour tout $\alpha \in \ell^1(\mathbb{Z})$, la mesure spectrale $\nu_{F_\alpha X}$ du processus filtré stationnaire $F_\alpha X$ est absolument continue par rapport à ν_X , de densité $u \in [-\pi, \pi] \mapsto |P_\alpha(e^{-iu})|^2$ où

$$P_\alpha(z) = \sum_{j \in \mathbb{Z}} \alpha_j z^j.$$

En particulier, si X possède une densité spectrale f_X alors $Y = F_\alpha X$ aussi et

$$\forall u \in [-\pi, \pi], \quad f_Y(u) = |P_\alpha(e^{-iu})|^2 f_X(u).$$

Note : comme les densités spectrales sont paires on peut remplacer e^{-iu} par e^{iu} .

Note : comme α est réel on a $|P_\alpha(e^{iu})| = |\overline{P_\alpha(e^{iu})}| = |P_\alpha(e^{-iu})| = |P_\alpha(e^{-iu})|$.

Note : P_α est un polynôme ssi le support $\{k \in \mathbb{Z} : \alpha_k \neq 0\}$ est une partie finie de \mathbb{N} .

Démonstration. Par le théorème 2.2, $\gamma_{F_\alpha X}(h) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h + k - j)$, d'où

$$\begin{aligned} \gamma_{F_\alpha X}(h) &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \int_{[-\pi, \pi]} e^{iu(h+k-j)} \nu_X(du) \\ &= \int_{[-\pi, \pi]} \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \alpha_j \alpha_k e^{iu(h+k-j)} \nu_X(du) \\ &= \int_{[-\pi, \pi]} e^{iuh} \underbrace{\left| \sum_{j \in \mathbb{Z}} \alpha_j e^{-iuj} \right|^2}_{\substack{P_\alpha(e^{-iu}) \\ \nu_{F_\alpha X}(du)}} \nu_X(du). \end{aligned}$$

Note : si on avait utilisé $h + j - k$ (et pas $h + k - j$) alors on aurait eu e^{iu} (et pas e^{-iu}).

Note : si $\gamma, \alpha \in \ell^1(\mathbb{Z})$ alors $h \mapsto \sum_{j, k \in \mathbb{Z}} \alpha_j \alpha_k \gamma(h + j - k) \in \ell^1(\mathbb{Z})$ car

$$\forall h \in \mathbb{Z}, \quad \sum_{j, k \in \mathbb{Z}} \alpha_j \alpha_k \gamma(h + j - k) = \sum_{j \in \mathbb{Z}} \alpha_j (\underbrace{\alpha * \gamma}_{\in \ell^1(\mathbb{Z})})_{h+j} = (\underbrace{\alpha * \gamma}_{\in \ell^1(\mathbb{Z})})(h).$$

□

Exemple 4.13 : Bruits blancs et processus linéaires

Si $Z = (Z_t)_{t \in \mathbb{Z}}$ est un BB(0, σ^2) alors $\gamma_Z(h) = \sigma^2 \mathbf{1}_{h=0}$ et donc Z admet pour densité

spectrale la fonction constante

$$u \in [-\pi, \pi] \mapsto f_Z(u) = \frac{\sigma^2}{2\pi}.$$

En voyant chaque fréquence entre $-\pi$ et π comme une couleur, on a un poids uniforme sur toutes les couleurs, ce qui explique le terme **bruit blanc**. Si $\alpha = (\alpha_k)_{k \in \mathbb{Z}} \in \ell^1(\mathbb{Z})$ alors d'après le théorème 4.12 le processus linéaire $F_\alpha Z$ a pour densité spectrale

$$u \in [-\pi, \pi] \mapsto \frac{\sigma^2}{2\pi} |P_\alpha(e^{-iu})|^2.$$

Par conséquent, si la mesure spectrale n'admet pas de densité (présence d'une partie étrangère à la mesure de Lebesgue, comme des masses de Dirac par exemple), alors il ne peut pas s'agir d'un processus linéaire, obtenu en filtrant un bruit blanc. D'un autre côté, en combinant les théorèmes 1.18 et 4.6 on obtient que toute mesure positive finie est la mesure spectrale d'un processus gaussien fortement stationnaire.

Exemple 4.14 : Densité spectrale d'un ARMA

Si $X = (X_t)_{t \in \mathbb{Z}}$ est un processus ARMA(p, q) solution de l'équation

$$\Phi(B)X = \Theta(B)Z \quad \text{où} \quad Z = (Z_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$$

et si $\Phi(z) = 1 - (\varphi_1 z + \dots + \varphi_p z^p)$ et $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ n'ont pas de racine commune de module 1, alors $X = F_\alpha Z$ où $\alpha \in \ell^1(\mathbb{Z})$ et $\Theta(z)/\Phi(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k$ pour tout $z \in \mathbb{C}$ tel que $|z| = 1$, et donc X , en combinant l'exemple 4.13 avec le théorème 3.6, on obtient que X a pour densité spectrale

$$\forall u \in [-\pi, \pi], \quad f_X(u) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-iu})|^2}{|\Phi(e^{-iu})|^2}.$$

Cette quantité est plus facile à manier que l'autocovariance pour les ARMA. Autrement dit, la densité spectrale d'un ARMA est proportionnelle au carré du module de sa **fraction rationnelle** sur le cercle unité. En particulier, un MA(1) d'équation $X_t = Z_t + \theta Z_{t-1}$ a pour densité spectrale

$$u \in [-\pi, \pi] \mapsto \frac{\sigma^2}{2\pi} |1 + \theta e^{-iu}|^2 = \frac{\sigma^2}{2\pi} (1 + 2\theta \cos(u) + \theta^2),$$

tandis qu'un AR(1) d'équation $X_t - \varphi X_{t-1} = Z_t$ a pour densité spectrale

$$u \in [-\pi, \pi] \mapsto \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-iu}|^2} = \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\varphi \cos(u) + \varphi^2}.$$

Pour un processus stationnaire ARMA(p, q) donné, peut-on construire un processus stationnaire ARMA(p, q), causal et inversible, de même fonction d'autocovariance? Le théorème suivant affirme que oui, quitte à changer la variance du bruit blanc. D'après le théorème 4.6 de Herglotz, deux processus stationnaires centrés ont la même fonction d'autocovariance ssi ils ont la même mesure spectrale. **La fonction d'autocovariance d'un processus stationnaire ARMA ne dépend que du module de sa fraction**

rationnelle sur le cercle unité. L'idée est donc de construire une fraction rationnelle dont le module coïncide avec celle du processus d'origine sur le cercle unité, et dont les racines du numérateur et du dénominateur (pôles) sont de module > 1 .

Théorème 4.15 : ARMA : représentation isospectrale causale inversible

Soit X le processus linéaire solution de $\text{ARMA}(p, q)$ $\Phi(B)X = \Theta(B)Z$ où Z est un $\text{BB}(0, \sigma^2)$, et où Φ et Θ n'ont pas de racines de module 1. Soient a_1, \dots, a_p et b_1, \dots, b_q les racines de Φ et Θ , avec

$$|a_1| \leq \dots \leq |a_r| < 1 < |a_{r+1}| \leq \dots \leq |a_p|$$

et

$$|b_1| \leq \dots \leq |b_s| < 1 < |b_{s+1}| \leq \dots \leq |b_q|$$

où $0 \leq r \leq p$ et $0 \leq s \leq q$, de sorte que pour tout $z \in \mathbb{C}$,

$$\Phi(z) = \prod_{j=1}^p (1 - a_j^{-1}z) \quad \text{et} \quad \Theta(z) = \prod_{j=1}^q (1 - b_j^{-1}z).$$

Soient Φ_* et Θ_* les polynômes définis par

$$\Phi_*(z) = \prod_{j=1}^r (1 - \overline{a_j}z) \prod_{j=r+1}^p (1 - a_j^{-1}z) \quad \text{et} \quad \Theta_*(z) = \prod_{j=1}^s (1 - \overline{b_j}z) \prod_{j=s+1}^q (1 - b_j^{-1}z).$$

Soit à présent Z_* un $\text{BB}(0, \sigma_*^2)$ où (avec la convention $\prod_{\emptyset} = 1$)

$$\sigma_*^2 = \sigma^2 \frac{\prod_{j=1}^r |a_j|^2}{\prod_{j=1}^s |b_j|^2}.$$

Alors Θ_* et Φ_* n'ont pas de racines de module ≤ 1 et le processus linéaire X_* solution de $\text{ARMA}(p, q)$ $\Phi_*(B)X_* = \Theta_*(B)Z_*$ est causal et inversible, et possède la même densité spectrale (donc la même autocovariance) que le processus d'origine X .

Note : on a $\Phi_*(0) = \Theta_*(0) = 1$ grâce à la manière de prendre en compte les racines.

Démonstration (facultative). Les polynômes Φ_* et Θ_* ont pour racines respectives

$$\frac{1}{\overline{a_1}}, \dots, \frac{1}{\overline{a_r}}, a_{r+1}, \dots, a_p \quad \text{et} \quad \frac{1}{\overline{b_1}}, \dots, \frac{1}{\overline{b_s}}, b_{s+1}, \dots, b_q,$$

qui sont toutes de module > 1 car l'application $z \mapsto 1/\overline{z}$ conserve l'argument et inverse le module. Par définition, on a

$$\Phi(z) = 1 - \sum_{j=1}^p \varphi_j z^j = -\varphi_p \prod_{j=1}^p (z - a_j) \quad \text{et} \quad \Theta(z) = 1 + \sum_{j=1}^q \theta_j z^j = \theta_q \prod_{j=1}^q (z - b_j).$$

Comme $\Phi(0) = \Theta(0) = 1$, on a $|a_1 \cdots a_p \varphi_p| = |b_1 \cdots b_q \theta_q| = 1$, et

$$|\Phi(z)| = \prod_{j=1}^p \left| \frac{z - a_j}{a_j} \right| = \prod_{j=1}^p |1 - a_j^{-1}z| \quad \text{et} \quad |\Theta(z)| = \prod_{j=1}^q \left| \frac{z - b_j}{b_j} \right| = \prod_{j=1}^q |1 - b_j^{-1}z|.$$

D'après l'exemple 4.14, on a, pour tout $u \in [-\pi, \pi]$,

$$f_X(u) = \frac{\sigma^2 |\Theta(e^{-iu})|^2}{2\pi |\Phi(e^{-iu})|^2} = \frac{\sigma^2 \prod_{j=1}^q |1 - b_j^{-1} e^{-iu}|^2}{2\pi \prod_{j=1}^p |1 - a_j^{-1} e^{-iu}|^2} = \frac{\sigma_*^2 |\Theta_*(e^{-iu})|^2}{2\pi |\Phi_*(e^{-iu})|^2} = f_{X_*}(u),$$

où on a utilisé, pour tout $z \in \mathbb{C}$ tel que $|z| = 1$ (notons que $z^{-1} = \bar{z}$ dans ce cas!),

$$|1 - c^{-1}z| = |c|^{-1}|c - z| = |c|^{-1}|\bar{z}c - 1| = |c|^{-1}|1 - \bar{c}z|$$

pour déplacer les racines de l'intérieur vers l'extérieur du disque unité, sans perturber la valeur du module! À présent, comme les racines de Φ_* et Θ_* sont de module > 1 , le processus ARMA(p, q) X_* est causal et inversible (théorème 3.9). \square

Exemple 4.16 : Représentation causale inversible d'un processus ARMA

Considérons l'équation ARMA(1, 1) suivante où Z est un BB(0, σ^2) :

$$X_t - 2X_{t-1} = Z_t + 3Z_{t-1}.$$

On a $\Phi(z) = 1 - 2z$ dont l'unique racine est $a_1 = 1/2$, et $\Theta(z) = 1 + 3z$ dont l'unique racine est $b_1 = -1/3$. L'équation ARMA(1, 1) ci-dessus a une solution X dans l'ensemble des processus linéaires d'après le théorème 3.6, mais elle n'est ni causale ni inversible d'après le théorème 3.9. D'après le théorème 4.15, $\Phi_*(z) = 1 - z/2$ et $\Theta_*(z) = 1 + z/3$, et $\sigma_*^2 = \frac{9}{4}\sigma^2$, et le processus X_* solution de l'équation ARMA(1, 1)

$$X_{*t} - \frac{1}{2}X_{*(t-1)} = Z_t + \frac{1}{3}Z_{t-1}$$

est causal et inversible et de même densité spectrale que X (donc même autocov.).

8→9

4.4 Compléments

Il existe un analogue complexe du théorème de Herglotz, dans lequel les suites γ hermitiennes de type positif sont les transformées de Fourier de mesures positives finies (plus forcément paires).

Remarque 4.17 : Symétrie

Si une mesure signée finie ν sur $[-\pi, \pi]$ est paire^a, alors $\hat{\nu}$ est paire et réelle : $\forall h \in \mathbb{Z}$, $\hat{\nu}(h) = \hat{\nu}(-h) \in \mathbb{R}$. De même, si $f \in L^1([-\pi, \pi], du)$ est paire^b alors \hat{f} est réelle et paire : $\forall h \in \mathbb{Z}$, $\hat{f}(h) = \hat{f}(-h) \in \mathbb{R}$.

a. C.-à-d. que $\int_{[-\pi, \pi]} g(u) \nu(du) = \int_{[-\pi, \pi]} g(-u) \nu(du)$, $\forall g : [-\pi, \pi] \rightarrow \mathbb{R}$ mesurable bornée.

b. C.-à-d. que $f(u) = f(-u)$ pour presque tout $u \in [-\pi, \pi]$.

Remarque 4.18 : Aspects hilbertiens

Soit $f \in L^2([-\pi, \pi], du) \subset L^1([-\pi, \pi], du)$. Alors on a $\widehat{f}(h) = \langle \varphi_{-h}, f \rangle$. Pour tous $h, h' \in \mathbb{Z}$, $\varphi_{-h} = \varphi_h$ et

$$\langle \varphi_h, \varphi_{h'} \rangle = \int_{-\pi}^{\pi} \varphi_h(u) \varphi_{h'}(u) du = \int_{-\pi}^{\pi} e^{it(h-h')} du = 2\pi \mathbf{1}_{h=h'}.$$

En posant $\psi_h = \frac{1}{\sqrt{2\pi}} \varphi_h$, ceci montre que la famille $(\psi_h)_{h \in \mathbb{Z}}$ est orthonormée dans $L^2_{\mathbb{C}}([-\pi, \pi], du)$. De plus, cette famille est dense dans $L^2_{\mathbb{C}}([-\pi, \pi], du)$ (ceci découle par exemple du théorème de Stone-Weierstrass, cf. preuve du théorème 4.3 plus loin). Cette famille constitue donc une base hilbertienne, et l'identité de Parseval donne

$$f = \sum_{h \in \mathbb{Z}} \langle \psi_h, f \rangle \psi_h = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \varphi_{-h} \widehat{f}(h).$$

La série converge dans $L^2_{\mathbb{C}}([-\pi, \pi], du)$, et ponctuellement absolument si $\widehat{f} \in \ell^1_{\mathbb{C}}(\mathbb{Z})$.

Remarque 4.19 : Coefficients de Fourier et transformée de Fourier

Pour une mesure finie ν sur \mathbb{R}^d , on considère en général la **transformée de Fourier** à paramètre $h \in \mathbb{R}^d$, définie par $h \in \mathbb{R}^d \mapsto \widehat{\nu}(h) = \int \varphi_h d\nu$. Le théorème de Herglotz possède un analogue pour les fonctions $\mathbb{R} \rightarrow \mathbb{C}$ dû à Bochner. Lorsque ν est définie sur l'ensemble compact $[-\pi, \pi]^d$, on peut se restreindre à un paramètre $h \in \mathbb{Z}^d$, ce qui donne la notion de **coefficient de Fourier**, et de **série de Fourier**. Dans ce cas, le théorème 4.3 d'injectivité peut être obtenu en régularisant par convolution avec une mesure gaussienne, point fixe de la transformée de Fourier, ce qui produit de la densité, puis en inversant la transformée de Fourier.

Remarque 4.20 : Séries de Fourier

Le théorème 4.4 fournit une condition suffisante pour la reconstruction de fonction intégrable à partir de ses coefficients de Fourier : si $f \in L^1([-\pi, \pi], du)$ vérifie $\widehat{f} \in \ell^1_{\mathbb{C}}(\mathbb{Z})$, alors pour presque tout $u \in [-\pi, \pi]$,

$$f(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-iuh} \widehat{f}(h).$$

La série du membre de droite (on parle de **série de Fourier**) converge absolument. Cette formule exprime f comme une superposition (un mélange) de fonctions trigonométriques élémentaires $u \mapsto e^{ihu} = \cos(hu) + i \sin(hu)$, où h joue le rôle d'une fréquence (couleur), d'où les termes **synthèse spectrale** et **analyse harmonique**. Cette représentation permet d'effectuer un filtrage, consistant à éliminer certaines fréquences ou plages de fréquences (filtre passe-haut, passe-bas, passe-bande, etc).

Théorème 4.21 : Lemme de Riemann-Lebesgue

Si $f \in L^1([-\pi, \pi], du)$ alors

$$\lim_{|h| \rightarrow \infty} \widehat{f}(h) = 0.$$

Note : $f \in L^1([-\pi, \pi], du) \mapsto \widehat{f} \in \{\alpha \in \mathbb{C}^{\mathbb{Z}} : \lim_{|h| \rightarrow \infty} \alpha_h = 0\}$ n'est pas surjective, cf. théorème 5.15 p. 99 du livre d'analyse réelle et complexe de Rudin.

Démonstration. D'après la preuve du théorème 4.3, l'ensemble \mathcal{T} des polynômes trigonométriques est dense, pour $\|\cdot\|_{\infty}$, dans l'ensemble $\mathcal{C}([-\pi, \pi])$ des fonctions continues sur $[-\pi, \pi]$. Par ailleurs, $\mathcal{C}([-\pi, \pi])$ est dense, pour $\|\cdot\|_1$, dans $L^1([-\pi, \pi], du)$. Or $\|\cdot\|_1 \leq \|\cdot\|_{\infty}$, et donc \mathcal{T} est dense, pour $\|\cdot\|_1$, dans $L^1([-\pi, \pi], du)$. Ainsi, pour tout $f \in L^1([-\pi, \pi], du)$ et tout $\varepsilon > 0$, il existe $P = P_{f, \varepsilon} \in \mathcal{T}$ tel que $\|f - P\|_1 \leq \varepsilon$. Or pour tout $h \in \mathbb{Z}$ avec $|h| > \deg(P)$, on a $\langle e^{iuh}, P \rangle = 0$ et donc

$$|\widehat{f}(h)| = \left| \int_{-\pi}^{\pi} e^{itu} (f(u) - P(u)) du \right| \leq \|f - P\|_1 \leq \varepsilon.$$

□

Remarque 4.22 : Fourier et convolution

Le passage en Fourier transforme la convolution en produit. Il est possible de le voir sur les mesures ou sur les fonctions. Si par exemple f et g sont deux fonctions intégrables sur $[-\pi, \pi]$, prolongées par périodicité à tout \mathbb{R} , alors leur convolution $f_1 * f_2$, définie pour tout $u \in \mathbb{R}$ par

$$(f_1 * f_2)(u) = \int_{-\pi}^{\pi} f_1(v) f_2(u - v) dv,$$

est une fonction intégrable sur $[-\pi, \pi]$, et $\widehat{f_1 * f_2} = \widehat{f_1} \widehat{f_2}$. En effet, grâce au théorème de Fubini-Tonelli (\diamond) et par périodicité ($\diamond\diamond$), pour tout $h \in \mathbb{Z}$,

$$\begin{aligned} \widehat{f_1 * f_2}(h) &= \int_{-\pi}^{\pi} e^{iuh} \int_{-\pi}^{\pi} f_1(v) f_2(u - v) dv du \\ &\stackrel{\diamond}{=} \int_{-\pi}^{\pi} e^{ivh} f_1(v) \int_{-\pi}^{\pi} e^{i(u-v)h} f_2(u - v) du dv \\ &\stackrel{\diamond\diamond}{=} \int_{-\pi}^{\pi} e^{ivh} f_1(v) \int_{-\pi}^{\pi} e^{iwh} f_2(w) dw dv \\ &= \widehat{f_1}(h) \widehat{f_2}(h). \end{aligned}$$

Le résultat suivant a été obtenu à l'origine par Carathéodory et Toeplitz. Il apparaît ici comme un corollaire du théorème 4.6 de Herglotz. Il existe également une preuve algorithmique, tout à fait différente, obtenue par Schur, liée à l'algorithme de Bareiss de calcul de la **décomposition de Cholesky** des matrices de Toeplitz. Ce bouillonnement a eu lieu au début du vingtième siècle, cf. [Am].

Théorème 4.23 : Carathéodory-Toeplitz

Si $\gamma \in \ell_{\mathbb{C}}^1(\mathbb{Z})$ est hermitienne, c'est-à-dire que

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty \quad \text{et} \quad \forall h \in \mathbb{Z}, \quad \gamma(-h) = \overline{\gamma(h)}.$$

alors les propriétés suivantes sont équivalentes :

1. γ est de type positif : pour tout $n \geq 1$ et tout $v \in \mathbb{C}^n$, $\sum_{j,k=1}^n v_j \overline{v_k} \gamma(j-k) \geq 0$;
2. L'image du disque unité ouvert $\{z \in \mathbb{C} : |z| < 1\}$ par $g(z) = \sum_{n=0}^{\infty} \gamma(n) z^n$ est incluse dans le demi-plan droit fermé $\{z \in \mathbb{C} : \Re(z) \geq 0\}$.

Démonstration. Comme γ est hermitienne, on a $\gamma(0) \in \mathbb{R}$. D'autre part g est bien définie sur le disque unité fermé car $\gamma \in \ell_{\mathbb{C}}^1(\mathbb{Z})$. On pose, pour tout $u \in [-\pi, \pi]$,

$$2\pi f(u) = \sum_{h \in \mathbb{Z}} e^{-iuh} \gamma(h) = \gamma(0) + \sum_{h=1}^{\infty} \left(e^{iuh} \overline{\gamma(h)} + e^{-iuh} \gamma(h) \right) = 2\Re(g(e^{-iu})) - \gamma(0).$$

Supposons que γ est de type positif. Pour $n = 1$ on obtient $\gamma(0) \geq 0$. D'autre part, d'après le théorème 4.6, γ est la transformée de Fourier d'une mesure positive finie sur $[-\pi, \pi]$, qui admet f pour densité car $\gamma \in \ell_{\mathbb{C}}^1(\mathbb{Z})$. Donc $f \geq 0$ p.p. et donc partout par continuité. Il en découle que $\Re g(z) \geq 0$ pour tout $z \in \mathbb{Z}$ tel que $|z| = 1$. Comme g est holomorphe, sa partie réelle est harmonique, et comme $g(0) = \gamma(0) \geq 0$, il en découle que $\Re g(z) \geq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| \leq 1$.

Réciproquement, supposons que $\Re g(z) \geq 0$ pour tout $z \in \mathbb{Z}$ tel que $|z| < 1$. En prenant $z = 0$ on obtient $\Re(\gamma(0)) \geq 0$. Comme γ est hermitienne, $\gamma(0)$ est réel, et donc $\gamma(0) \geq 0$. Par ailleurs, par continuité, $\Re g(z) \geq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| = 1$. Il en découle que $f \geq 0$. En passant en transformée de Fourier inverse (possible car $\gamma \in \ell_{\mathbb{C}}^1(\mathbb{Z})$), on obtient que γ est la transformée de Fourier d'une mesure positive finie, et donc γ est de type positif grâce au théorème 4.6. \square

Prédiction linéaire (x3)

Mots-clés : prédiction ; équations de Yule-Walker ; matrices de Toeplitz ; décomposition de Cholesky ; algorithme de Levinson-Durbin ; algorithme de Gram-Schmidt (ou des innovations).

Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire, de moyenne nulle, et d'autocovariance γ vérifiant $\gamma(0) > 0$ (le processus n'est pas identiquement nul). Dans ce chapitre, on s'intéresse à la meilleure approximation, au sens des moindres carrés, de X_t avec une combinaison linéaire de X_{t-1}, \dots, X_{t-p} . Cette approximation est une projection orthogonale, dont les coefficients ne dépendent que de l'autocovariance γ . La stationnarité du processus permet d'estimer γ dans une plage temporelle donnée, et d'utiliser son invariance par translation pour faire de la prédiction. Ce mécanisme d'estimation-prédiction rend naturelle l'hypothèse de stationnarité.

Dans ce chapitre, on utilise à la fois la structure hilbertienne de L^2 et de \mathbb{R}^p , et on note $\langle \cdot, \cdot \rangle$ et $\|\cdot\|_2$ le produit scalaire et la norme, par abus de notation.

5.1 Prédicteur linéaire optimal

Pour tous entiers $p \geq 0$ et $t \in \mathbb{Z}$, on note

$$H_{t-1,p} = \text{vect}\{X_{t-1}, \dots, X_{t-p}\}$$

le sous-espace vectoriel de L^2 engendré par X_{t-1}, \dots, X_{t-p} , avec pour convention $H_{t-1,0} = \{0\}$ si $p = 0$. Ce sous-espace, de dimension au plus p , représente le passé de profondeur p de X_t . Le carré de la distance dans L^2 de X_t à $H_{t-1,p}$ est donné par

$$\sigma_p^2 = \text{dist}_{\|\cdot\|_2}(X_t, H_{t-1,p})^2 = \inf_{Y \in H_{t-1,p}} \|X_t - Y\|_2^2.$$

L'infimum est atteint en un unique $Y \in H_{t-1,p}$, noté $\text{proj}(X_t, H_{t-1,p})$, projection orthogonale dans L^2 de X_t sur $H_{t-1,p}$. On dit qu'il s'agit du **prédicteur progressif** ou **direct** de X_t par une combinaison linéaire de X_{t-1}, \dots, X_{t-p} . Il est optimal au sens des **moindres carrés**. Comme $\text{proj}(X_t, H_{t-1,p}) \in H_{t-1,p}$, il existe $\varphi_p = (\varphi_{p,1}, \dots, \varphi_{p,p})^\top \in \mathbb{R}^p$ tel que

$$\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-k}.$$

Le vecteur des coefficients φ_p est unique ssi les vecteurs X_{t-1}, \dots, X_{t-p} sont linéairement indépendants dans L^2 , c'est-à-dire que $\dim(H_{t-1,p}) = p$. C'est le cas ssi la matrice de

covariance $\Gamma_p = \text{Cov}(X_{t-1}, \dots, X_{t-p})$ est inversible (matrice de Gram). La projection $\text{proj}(X_t, H_{t-1,p})$ et l'**erreur de prédiction progressive** (ou directe)

$$E_{t,p}^+ = X_t - \text{proj}(X_t, H_{t-1,p})$$

sont centrées, et la variance σ_p^2 de l'erreur de prédiction $E_{t,p}^+$ vérifie

$$\sigma_p^2 = \|E_{t,p}^+\|_2^2 = \|X_t - \text{proj}(X_t, H_{t-1,p})\|_2^2 = \mathbb{E}((X_t - \text{proj}(X_t, H_{t-1,p}))^2).$$

La figure 5.1 résume ces formules en une.

Note : φ_p et σ_p^2 ne dépendent pas de t car le processus est stationnaire : en effet

$$\begin{aligned} \text{proj}(X_t, H_{t-1,p}) &= \arg \inf_{Y \in H_{t-1,p}} \|X_t - Y\|_2^2 \\ &= \arg \inf_{\varphi \in \mathbb{R}^p} \left\| X_t - \sum_{k=1}^p \varphi_k X_{t-k} \right\|_2^2 \\ &= \arg \inf_{\varphi \in \mathbb{R}^p} \sum_{j,k=0}^p \varphi_j \varphi_k \gamma_X(j-k) \end{aligned}$$

(avec la convention $\varphi_0 := -1$), formule qui ne dépend plus du tout de t , et de même

$$\sigma_p^2 = \inf_{Y \in H_{t-1,p}} \|X_t - Y\|_2^2 = \inf_{\varphi \in \mathbb{R}^p} \sum_{j,k=0}^p \varphi_j \varphi_k \gamma_X(j-k).$$

Note : si $p = 0$ alors $\text{proj}(X_t, H_{t-1,0}) = \text{proj}(X_t, \{0\}) = 0$, $E_{t,p}^+ = X_t$, et $\sigma_p^2 = \gamma(0)$.

Note : si X est gaussien alors, en notant $\mathcal{F}_{t-1,p}$ la tribu engendrée par X_{t-1}, \dots, X_{t-p} ,

$$\text{proj}(X_t, H_{t-1,p}) = \text{proj}(X_t, L^2(\mathcal{F}_{t-1,p})) = \mathbb{E}(X_t | \mathcal{F}_{t-1,p})$$

ce qui est tout à faire remarquable car $H_{t-1,p} \subsetneq L^2(\mathcal{F}_{t-1,p})$.

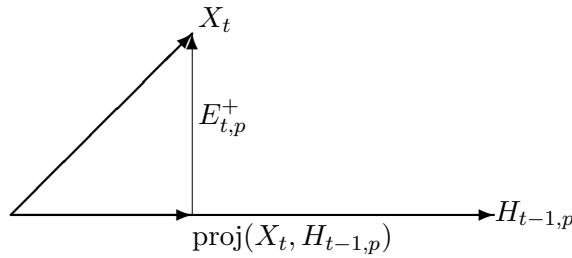


FIGURE 5.1 – Prédiction progressive et erreur progressive.

Exemple 5.1 : Processus AR(m) causal

Soit Z un $\text{BB}(0, \sigma^2)$ et soit X un processus $\text{AR}(m)$ linéaire causal solution de

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_m X_{t-m} + Z_t,$$

où $1 - (\varphi_1 z + \dots + \varphi_m z^m) \neq 0$ pour tout $z \in \mathbb{C}$ avec $|z| \leq 1$. Alors, comme X est causal, on a $\mathbb{E}(Z_t X_{t-h}) = 0$ pour tout $h \geq 1$, et l'équation $\text{AR}(m)$ donne l'identité $\mathbb{E}((X_t - \sum_{k=1}^m \varphi_k X_{t-k}) X_{t-h}) = 0$ pour tout $h \geq 1$. Ceci implique que $X_t - \sum_{k=1}^m \varphi_k X_{t-k} \perp H_{t-1,p}$ pour tout $p \geq 1$. De plus $\sum_{k=1}^m \varphi_k X_{t-k} \in H_{t-1,p}$ pour

tout $m \leq p$. Donc ^a pour tout $p \geq m$, on a $\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_k X_{t-k}$ avec

$$\varphi_{p,k} = \begin{cases} \varphi_k & \text{si } 1 \leq k \leq m \\ 0 & \text{si } m < k \leq p. \end{cases}$$

Ainsi, pour notre processus $\text{AR}(m)$ causal, les coefficients de prédiction linéaires de profondeur p coïncident avec les coefficients de l'équation d'autorégression, dès que $p \geq m$. Ceci explique la notation φ utilisée pour les coefficients de prédiction. Nous allons voir plus loin le cas des processus $\text{MA}(q)$ et un algorithme lié à la notation θ .

^a. Caractérisation du projeté orthogonal : $v = \text{proj}(u, H)$ ssi $v \in H$ et $u - v \perp H$.

5.2 Équations de Yule-Walker

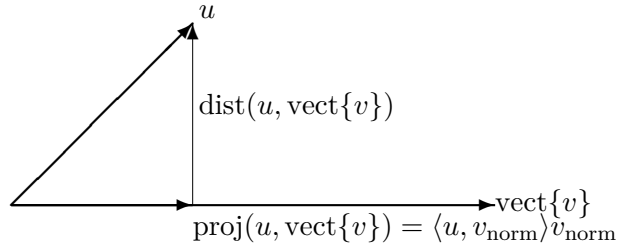


FIGURE 5.2 – Intuition géométrique : projection du vecteur u sur le sous-espace $\text{vect}\{v\} = \text{vect}\{v_{\text{norm}}\}$ de dimension 1. Ici $v_{\text{norm}} = \frac{v}{\|v\|}$ de sorte que $v = \|v\|v_{\text{norm}}$ et $\|v_{\text{norm}}\| = 1$. Le théorème de Pythagore s'écrit ici : $\|u\|^2 = \|\text{dist}(u, \text{vect}\{v\})\|^2 + \|\text{proj}(u, \text{vect}\{v\})\|^2$.

Pour $p = 1$, les choses sont simples et directes (figure 5.2) :

$$\text{proj}(X_t, H_{t-1,1}) = \text{proj}(X_t, \text{vect}\{X_{t-1}\}) = \left\langle X_t, \frac{X_{t-1}}{\|X_{t-1}\|_2} \right\rangle \frac{X_{t-1}}{\|X_{t-1}\|_2}$$

ce qui donne

$$\varphi_{1,1} = \left\langle X_t, \frac{X_{t-1}}{\|X_{t-1}\|_2} \right\rangle \frac{1}{\|X_{t-1}\|_2} = \langle X_t, X_{t-1} \rangle \frac{1}{\|X_{t-1}\|_2^2} = \frac{\gamma(1)}{\gamma(0)}.$$

Plus généralement, le vecteur $\varphi_p = (\varphi_{p,k})_{1 \leq k \leq p}^\top$ vérifie les équations de Yule¹-Walker² :

Théorème 5.2 : Équations Yule-Walker

Si $\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-k}$ alors

$$\Gamma_p \varphi_p = \gamma_p$$

1. George Udny Yule (1871–1951).

2. Gilbert Thomas Walker (1868–1958).

où $\Gamma_p = (\gamma(k-j))_{1 \leq k, j \leq p}$ et où $\gamma_p = (\gamma(1), \dots, \gamma(p))^\top$. De plus

$$\sigma_p^2 = \gamma(0) - \varphi_p^\top \gamma_p = \gamma(0) - \varphi_p^\top \Gamma_p \varphi_p.$$

Note : $\Gamma_p = (\gamma(k-j))_{1 \leq k, j \leq p}$ est la matrice de covariance de $(X_{t-1}, \dots, X_{t-p})$. Il s'agit d'une magnifique matrice de Toeplitz symétrique à valeurs propres ≥ 0 :

$$\Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \ddots & & \vdots \\ \gamma(2) & \gamma(1) & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \gamma(1) & \gamma(2) \\ \vdots & & \ddots & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(p-1) & \dots & \dots & \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix}.$$

Note : si Γ_p est inversible alors $\Gamma_p \varphi_p = \gamma_p$ a une unique solution $\varphi_p = \Gamma_p^{-1} \gamma_p$. Ceci permet de calculer les coefficients de prédiction φ_p à partir de l'autocovariance γ puisque Γ_p et γ_p sont des fonctions de γ . D'autre part, on a aussi

$$\sigma_p^2 = \gamma(0) - \gamma_p^\top \Gamma_p^{-1} \gamma_p.$$

Il s'agit du complément de Schur du bloc Γ_p dans la décomposition par bloc

$$\Gamma_{p+1} = \begin{pmatrix} \Gamma_p & w_p^\top \\ w_p & \gamma(0) \end{pmatrix} \quad \text{où} \quad w_p := (\gamma(p), \dots, \gamma(1))^\top.$$

On peut aussi écrire $\gamma_p^\top \Gamma_p^{-1} \gamma_p = \gamma(0) - \sigma_p^2 = \sigma_0^2 - \sigma_p^2$.

Démonstration. L'orthogonalité $X_t - \text{proj}(X_t, H_{t-1,p}) \perp H_{t-1,p}$ (dessin) donne

$$\forall 1 \leq j \leq p, \quad \left\langle X_t - \sum_{k=1}^p \varphi_{p,k} X_{t-k}, X_{t-j} \right\rangle = 0.$$

Cela s'écrit également au moyen de la fonction d'autocovariance γ ,

$$\forall 1 \leq j \leq p, \quad \sum_{k=1}^p \varphi_{p,k} \gamma(k-j) = \gamma(j).$$

d'où $\Gamma_p \varphi_p = \gamma_p$. D'autre part, $\text{proj}(X_t, H_{t-1,p}) \perp X_t - \text{proj}(X_t, H_{t-1,p})$, donc

$$\sigma_p^2 = \langle X_t - \text{proj}(X_t, H_{t-1,p}), X_t \rangle = \gamma(0) - \sum_{k=1}^p \varphi_{p,k} \gamma(k) = \gamma(0) - \varphi_p^\top \gamma_p.$$

□

Notons que grâce au théorème de Pythagore on obtient que

$$\|\text{proj}(X_t, H_{t-1,p})\|_2^2 = \|X_t\|_2^2 - \sigma_p^2 = \gamma(0) - \sigma_p^2 = \varphi_p^\top \gamma_p$$

tandis que par ailleurs

$$\|\text{proj}(X_t, H_{t-1,p})\|_2^2 = \sum_{j,k=1}^p \varphi_{p,j} \varphi_{p,k} \gamma(j-k) = \varphi_p^\top \Gamma_p \varphi_p.$$

Définition 5.3 : Décroissance et convergence de l'erreur de prédiction

On a $H_{t-1,p} \subset H_{t-1,p+1}$ et donc, comme $E \mapsto \inf E$ est décroissante pour l'inclusion,

$$\sigma_{p+1}^2 = \inf_{Y \in H_{t-1,p+1}} \|X_t - Y\|_2^2 \leq \inf_{Y \in H_{t-1,p}} \|X_t - Y\|_2^2 = \sigma_p^2.$$

La suite $(\sigma_p^2)_{p \geq 1}$ décroît et est minorée par 0, donc elle converge vers une limite ≥ 0 :

$$\sigma_\infty^2 := \lim_{p \rightarrow \infty} \sigma_p^2 = \inf_{p \geq 1} \sigma_p^2 \geq 0.$$

Définition 5.4 : Processus réguliers et déterministes

Un processus est **régulier** si $\sigma_\infty^2 > 0$ et **déterministe** si $\sigma_\infty^2 = 0$ (dans ce cas, X_t est une combinaison linéaire infinie à coefficients déterministes de X_{t-1}, X_{t-2}, \dots).

Note : la quantité σ_∞^2 est égale à la quantité σ^2 du lemme 1.21.

Théorème 5.5 : Condition suffisante d'inversibilité

Si $\gamma(h) \rightarrow 0$ quand $h \rightarrow \infty$ alors Γ_p est inversible pour tout $p \geq 1$.

Note : on rappelle que dans tout ce chapitre on suppose que $\gamma(0) > 0$.

Note : si $\gamma \in \ell^1(\mathbb{Z})$, i.e. $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$, alors $\gamma(h) \rightarrow 0$.

Note : si X a une densité spectrale alors $\gamma(h) \rightarrow 0$ (lemme de Riemann-Lebesgue 4.21).

Note : si $\gamma \in \ell^1(\mathbb{Z})$ alors X a une densité spectrale (théorème de Herglotz 4.6).

Démonstration. La démonstration fait appel au lemme suivant.

Lemme : Inversibilité des matrices symétriques de type positif

Soit A une matrice $n \times n$ symétrique réelle et de type positif. Alors A est inversible si et seulement si $u^\top A u \neq 0$ (c'est-à-dire > 0) pour tout $u \in \mathbb{R}^n$ tel que $u \neq 0$.

Notons que pour toute matrice A carrée réelle, les propriétés suivantes sont équivalentes :

- A est symétrique réelle de type positif
- A est diagonalisable en base orthonormée et ses valeurs propres sont ≥ 0
- A est la matrice de covariance d'un vecteur aléatoire.

Preuve du lemme. Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres de A et v_1, \dots, v_n une base orthonormée de vecteurs propres. On a $\lambda_k \geq 0$, $Av_k = \lambda_k v_k$, et $\langle v_j, v_k \rangle = \mathbf{1}_{j=k}$ pour tous $1 \leq j, k \leq n$. Pour tout $u \in \mathbb{R}^n$, $u = \sum_{j=1}^n \langle u, v_j \rangle v_j$, $Au = \sum_{k=1}^n \lambda_k \langle u, v_k \rangle v_k$, et

$$u^\top A u = \langle u, Au \rangle = \sum_{j,k=1}^n \lambda_k \langle u, v_j \rangle \langle u, v_k \rangle \langle v_j, v_k \rangle = \sum_{k=1}^n \lambda_k \langle u, v_k \rangle^2 \geq 0.$$

Maintenant, si A est inversible alors $\lambda_k > 0$ pour tout $1 \leq k \leq n$ et donc pour tout $u \in \mathbb{R}^n$ avec $u \neq 0$, en choisissant $1 \leq k \leq n$ tel que $\langle u, v_k \rangle \neq 0$, il vient $u^\top A u \geq \lambda_k \langle u, v_k \rangle^2 > 0$.

Réciproquement, si $u^\top A u > 0$ pour tout $u \in \mathbb{R}^n$ avec $u \neq 0$ alors pour tout $1 \leq k \leq n$, en prenant $u = v_k$ il vient $\lambda_k = \lambda_k \langle v_k, v_k \rangle = v_k^\top A v_k > 0$ et donc A est inversible. \square

Démontrons à présent le théorème. Supposons que $\gamma(p) \rightarrow 0$ quand $p \rightarrow \infty$. Pour $p = 1$, $\Gamma_1 = \gamma(0) > 0$ est inversible. À présent, supposons que $\Gamma_1, \dots, \Gamma_p$ sont inversibles pour $p \geq 1$ et montrons que Γ_{p+1} est alors inversible.

D'après le lemme, Γ_{p+1} est inversible ssi $u^\top \Gamma_{p+1} u \neq 0$ (> 0 en fait) pour tout $u \in \mathbb{R}^{p+1}$ avec $u \neq 0$. Raisonnons par l'absurde. Supposons que $u^\top \Gamma_{p+1} u = 0$ pour un $u \in \mathbb{R}^{p+1}$ avec $u \neq 0$. Si $u_{p+1} = 0$ alors $(u_1, \dots, u_p) \neq 0$, et la décomposition par blocs

$$\Gamma_{p+1} = \begin{pmatrix} \Gamma_p & w_p \\ w_p^\top & \gamma(0) \end{pmatrix} \quad \text{où} \quad w_p = (\gamma(p), \dots, \gamma(1))^\top$$

donne $(u_1, \dots, u_p)^\top \Gamma_p (u_1, \dots, u_p) = 0$ ce qui est impossible car Γ_p est inversible. Supposons à présent que $u_{p+1} \neq 0$. En notant $\mathbf{X}_{p+1} = (X_1, \dots, X_{p+1})^\top$, il vient

$$0 = u^\top \Gamma_{p+1} u = u^\top \mathbb{E}(\mathbf{X}_{p+1}^\top \mathbf{X}_{p+1}) u = \mathbb{E}(u^\top \mathbf{X}_{p+1}^\top \mathbf{X}_{p+1} u) = \mathbb{E}(\langle u, \mathbf{X}_{p+1} \rangle^2).$$

Donc $\langle u, \mathbf{X}_{p+1} \rangle = 0$ p.s. d'où, p.s. (la division par u_{p+1} est licite car $u_{p+1} \neq 0$)

$$X_{p+1} = - \sum_{k=1}^p \frac{u_k}{u_{p+1}} X_k.$$

Par stationnarité, le même procédé avec $(X_{h+1}, \dots, X_{h+p+1})$ donne, p.s. pour tout $h \in \mathbb{Z}$,

$$X_{h+p+1} = - \sum_{k=1}^p \frac{u_k}{u_{p+1}} X_{h+k}.$$

Par récurrence, pour tout $h \geq p+1$, il existe un vecteur déterministe de coefficients $v_h = (v_{h,1}, \dots, v_{h,p})^\top$ tel que p.s. pour tout $h \geq p+1$,

$$X_h = \sum_{k=1}^p v_{h,k} X_k = \langle v_h, \mathbf{X}_p \rangle.$$

À présent, en notant $\lambda_{\min}(\Gamma_p)$ la plus petite valeur propre de Γ_p , on a d'une part

$$\gamma(0) = \mathbb{E}(X_h^2) = \mathbb{E}(\langle v_h, \mathbf{X}_p \rangle^2) = v_h^\top \Gamma_p v_h \geq \lambda_{\min}(\Gamma_p) \|v_h\|_2^2.$$

Comme Γ_p est inversible on a $\lambda_{\min}(\Gamma_p) > 0$ et donc $\sup_{h,k} |v_{h,k}| < \infty$. D'autre part

$$\gamma(0) = \text{Cov}\left(X_h, \sum_{k=1}^p v_{h,k} X_k\right) = \sum_{k=1}^p v_{h,k} \gamma(k-h)$$

et donc

$$\gamma(0) \leq \sum_{k=1}^p |v_{h,k}| |\gamma(k-h)| \leq \left(\sup_{h,k} |v_{h,k}| \right) \sum_{k=1}^p |\gamma(k-h)|.$$

Comme $\lim_{h \rightarrow \infty} \gamma(h) = 0$, on obtient $\gamma(0) \leq 0$, ce qui est impossible car $\gamma(0) > 0$. \square

5.3 Algorithme de Cholesky

Lorsque $A = \Gamma_p$ est inversible, le coefficient φ_p de prédiction est unique et peut être calculé en résolvant l'équation de Yule-Walker $\Gamma_p \varphi_p = \gamma_p$. Cela revient à déterminer A^{-1} . Comme les formules de Cramer sont trop coûteuses en temps de calcul, la résolution pratique du système d'équations linéaires $Ax = b$ est plutôt faisable en utilisant un algorithme comme le pivot de Gauss (ou élimination de Gauss), ou encore une décomposition LU (pour Lower-triangular \times Upper-triangular). La plupart des algorithmes réduisent le problème aux matrices triangulaires. Le coût du calcul peut être amélioré en utilisant un algorithme qui tire partie de la structure de A . L'**algorithme de Cholesky** utilise le fait que A est symétrique positive. La **décomposition de Cholesky**³ de A , introduite par Cholesky vers 1924, affirme l'existence d'une matrice triangulaire inférieure L à termes diagonaux ≥ 0 , facile à calculer, telle que $A = LL^\top$. Si A est inversible, alors L est inversible, unique, et ses termes diagonaux sont > 0 . Une fois L calculée, on résout l'équation $Ax = b$ en x en résolvant l'équation triangulaire $Ly = b$ en y , puis l'équation triangulaire $L^\top z = y$ en z .

Sans perte de généralité, on peut utiliser une décomposition de la forme,

$$A = LDL^\top$$

où L est triangulaire inférieure à termes diagonaux = 1, et où D est diagonale à termes diagonaux ≥ 0 . Cette décomposition modifiée évite le calcul de racines carrées. Le lien entre les deux décompositions est

$$LDL^\top = LD^{1/2}(LD^{1/2})^\top,$$

où $D^{1/2} = \text{diag}(D_{1,1}^{1/2}, \dots, D_{p,p}^{1/2})$. Les colonnes de $LD^{1/2}$ s'obtiennent en multipliant celles de L par la racine des coefficients diagonaux de D .

Théorème 5.6 : Décomposition et algorithme de Cholesky

Si $A \in \mathcal{M}_{p,p}(\mathbb{R})$ est symétrique à valeurs propres > 0 alors

$$A = LDL^\top$$

où L est une matrice triangulaire inférieure à diagonale = 1 et D diagonale à diagonale > 0 , calculables par récurrence avec l'algorithme de Cholesky :

$$D_{k,k} = A_{k,k} - \sum_{r=1}^{k-1} L_{k,r}^2 D_{r,r} \quad \text{pour } 1 \leq k \leq p$$

$$L_{j,k} = \frac{1}{D_{k,k}} \left(A_{j,k} - \sum_{r=1}^{k-1} L_{j,r} L_{k,r} D_{r,r} \right) \quad \text{pour } 1 \leq k < j \leq p.$$

La décomposition s'étend aux matrices A à valeurs propres ≥ 0 en posant $L_{j,k} = 0$ si $D_{k,k} = 0$ (se produit si A est singulière), et dans ce cas L a une diagonale dans $\{0, 1\}$.

On prendra garde à ne pas confondre la décomposition de Cholesky $A = LDL^\top$ avec la diagonalisation de A en base orthonormée. Dans la décomposition de Cholesky, la matrice L n'est pas orthogonale en général, et les termes diagonaux de la matrice D ne sont pas les valeurs propres de A en général. La décomposition de Cholesky est moins coûteuse

3. On dit parfois **factorisation de Cholesky** («*factorization*» en anglais).

que la diagonalisation en base orthonormée. Elle permet de résoudre plus efficacement des systèmes d'équations linéaires mais aussi de simuler efficacement des vecteurs gaussiens.

Démonstration. Supposons qu'il existe L et D comme indiqué telles que $A = LDL^\top$. La définition du produit matriciel et la structure de L et D donnent, pour tout $1 \leq k \leq p$,

$$A_{k,k} = \sum_{r=1}^p L_{k,r} D_{r,r} L_{r,k}^\top = \sum_{r=1}^{k-1} L_{k,r}^2 D_{r,r} + D_{k,k},$$

tandis que pour tout $1 \leq k < j \leq p$,

$$A_{j,k} = \sum_{r=1}^p L_{j,r} D_{r,r} L_{r,k}^\top = \sum_{r=1}^{k-1} L_{j,r} L_{k,r} D_{r,r} + L_{j,k} D_{k,k}.$$

On retrouve bien les formules de Cholesky. Réciproquement, remarquablement, ces formules permettent, à partir de la matrice A , le calcul des matrices D et L en partant de $L_{1,1} = 1$ et $D_{1,1} = A_{1,1}$, et en procédant ligne par ligne ou colonne par colonne dans L .

La matrice D est à diagonale > 0 car semblable au sens des formes quadratiques à la matrice symétrique A à valeurs propres > 0 : pour tout vecteur colonne $u \in \mathbb{R}^p$,

$$\langle u, Au \rangle = u^\top Au = \langle L^\top u, DL^\top u \rangle.$$

La matrice L est inversible car triangulaire inférieure à diagonale > 0 (en fait $= 1$). \square

La complexité de l'algorithme de Cholesky est $\mathcal{O}(p^3)$. Bareiss⁴ a montré vers 1969 qu'il est possible de faire chuter la complexité à $\mathcal{O}(p^2)$ en tenant compte de la structure Toeplitz de $A = \Gamma_{p+1}$ (provient de la stationnarité).

Il se trouve que la décomposition de Cholesky $L_{p+1} D_{p+1} L_{p+1}^\top$ de Γ_{p+1} possède une expression remarquable en fonction des variances de prédiction et des coefficients de prédiction, comme le montre le théorème suivant. Cette expression remarquable provient de la stationnarité, et reste valable pour toute matrice de covariance Toeplitz.

Théorème 5.7 : Décomposition de Cholesky de Γ_{p+1}

Pour tout $p \geq 0$ on a

$$\Gamma_{p+1} = L_{p+1} D_{p+1} L_{p+1}^\top,$$

où

$$D_{p+1} = \text{diag}(\sigma_0^2, \dots, \sigma_p^2) \quad \text{et} \quad L_{p+1} = \Phi_{p+1}^{-1}$$

et où Φ_{p+1} est la matrice triangulaire inférieure suivante :

$$\Phi_{p+1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\varphi_{1,1} & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ -\varphi_{p,p} & -\varphi_{p,p-1} & \cdots & -\varphi_{p,1} & 1 \end{pmatrix}.$$

De plus Γ_{p+1} est inversible si et seulement si D_{p+1} est inversible, ce qui se produit si

4. Cet algorithme est aussi une version d'un algorithme de Schur, datant de 1917, qui constitue à son tour une preuve algorithmique des théorèmes 4.23 de Carathéodory-Toeplitz et 4.6 de Herglotz, cf. [Am].

et seulement si $\sigma_p > 0$, et dans ce cas

$$\Gamma_{p+1}^{-1} = \Phi_{p+1}^\top D_{p+1}^{-1} \Phi_{p+1} \quad \text{où} \quad D_{p+1}^{-1} = \text{diag}(\sigma_0^{-2}, \dots, \sigma_p^{-2}).$$

Note : à partir de Γ_{p+1} , on peut calculer L_{p+1} en utilisant l'algorithme de Cholesky, puis on peut inverser L_{p+1} , ce qui fournit $(\varphi_{r,k})_{1 \leq k \leq r \leq p}$, sans équations de Yule-Walker !

Démonstration. La matrice Φ_{p+1} est inversible car triangulaire à diagonale $\neq 0$ (car $= 1$). La matrice L_{p+1} hérite de Φ_{p+1} le caractère triangulaire inférieur à diagonale $= 1$. Par convention $\sigma_0^2 = \gamma(0)$, et on a $\Gamma_1 = \gamma(0) = \sigma_0^2 = D_1$. La matrice Γ_{p+1} est inversible ssi D_{p+1} est inversible, c'est-à-dire ssi $\min\{(D_{p+1})_{k,k} : 1 \leq k \leq p+1\} = \min\{\sigma_0^2, \dots, \sigma_p^2\} = \sigma_p^2 > 0$.

Établissons à présent que $\Phi_{p+1} \Gamma_{p+1} \Phi_{p+1}^\top = D_{p+1}$. Comme Φ_{p+1} est déterministe, on a, par définition de Γ_{p+1} , en posant $\mathbf{X}_{p+1} = (X_1, \dots, X_{p+1})^\top$,

$$\Phi_{p+1} \Gamma_{p+1} \Phi_{p+1}^\top = \Phi_{p+1} \mathbb{E}(\mathbf{X}_{p+1} \mathbf{X}_{p+1}^\top) \Phi_{p+1}^\top = \mathbb{E}(\Phi_{p+1} \mathbf{X}_{p+1} (\Phi_{p+1} \mathbf{X}_{p+1})^\top).$$

Calculons à présent $\Phi_{p+1} \mathbf{X}_{p+1}$. On a

$$\begin{aligned} \Phi_{p+1} \mathbf{X}_{p+1} &= \Phi_{p+1} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{p+1} \end{pmatrix} \\ &= \begin{pmatrix} X_1 \\ X_2 - \varphi_{1,1} X_1 \\ \vdots \\ X_{p+1} - \sum_{k=1}^p \varphi_{p,k} X_{p+1-k} \end{pmatrix} \\ &= \begin{pmatrix} X_1 - \text{proj}(X_1, H_{0,0}) \\ X_2 - \text{proj}(X_2, H_{1,1}) \\ \vdots \\ X_{p+1} - \text{proj}(X_{p+1}, H_{p,p}) \end{pmatrix} \\ &=: \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_{p+1} \end{pmatrix} \end{aligned}$$

car par définition des coefficients de prédiction $\varphi_1, \dots, \varphi_p$,

$$X_{k+1} - \text{proj}(X_{k+1}, \underbrace{H_{k,k}}_{H_{k+1-1,k}}) = X_{k+1} - \sum_{j=1}^k \varphi_{k,j} X_{k+1-j} = X_{k+1} - \sum_{j=1}^k \varphi_{k,k+1-j} X_j$$

(par convention $H_{0,0} = \{0\}$ ce qui donne $\text{proj}(X_1, H_{0,0}) = 0$). À présent, si $j < k$ alors

$$V_j = X_j - \text{proj}(X_j, H_{j-1,j-1}) \in H_{k-1,k-1} \perp X_k - \text{proj}(X_k, H_{k-1,k-1}) = V_k,$$

et donc V_1, \dots, V_{p+1} sont orthogonaux dans L^2 . De plus, comme on a (prendre $t = k$)

$$\sigma_{k-1}^2 = \|X_t - \text{proj}(X_t, H_{t-1,k-1})\|_2^2 = \|X_k - \text{proj}(X_k, H_{k-1,k-1})\|_2^2 = \|V_k\|_2^2,$$

il vient

$$\mathbb{E}(\Phi_{p+1} \mathbf{X}_{p+1} (\Phi_{p+1} \mathbf{X}_{p+1})^\top) = \mathbb{E}((\langle V_j, V_k \rangle)_{1 \leq j, k \leq p+1}) = \text{diag}(\sigma_0^2, \dots, \sigma_p^2) = D_{p+1}.$$

□

Exemple 5.8 : Processus AR(1) causal

Lorsque X est un processus AR(1) causal solution de $X_t = \varphi X_{t-1} + Z_t$ avec $|\varphi| < 1$ et $Z \sim \text{BB}(0, \sigma^2)$, alors d'après la remarque 5.1, $\varphi_p = (\varphi, 0, \dots, 0)^\top$ pour tout $p \geq 1$, et comme $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \varphi^j \varphi^{j+h} = \sigma^2 \varphi^h / (1 - \varphi^2)$ pour tout $h \geq 0$, on obtient $\sigma_0^2 = \sigma^2 / (1 - \varphi^2)$ et $\sigma_p^2 = \gamma(0) - \varphi_p^\top \gamma_p = \gamma(0) - \varphi \gamma(1) = \sigma^2$ pour tout $p \geq 1$, d'où

$$D_p = \text{diag}\left(\frac{\sigma^2}{1 - \varphi^2}, \sigma^2, \dots, \sigma^2\right) \quad \text{et} \quad \Phi_p = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\varphi & 1 & \ddots & & \vdots \\ 0 & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & -\varphi & 1 \end{pmatrix}.$$

pour tout $p \geq 1$. De ces formules on tire la formule $\Phi_p \gamma_p = (\gamma(1), 0, \dots, 0)^\top$ puis $D_p^{-1} \Phi_p \gamma_p = (\varphi, 0, \dots, 0)^\top$ et on retrouve alors bien la formule de Yule-Walker :

$$\varphi_p = \Gamma_p^{-1} \gamma_p = \Phi_p^\top D_p^{-1} \Phi_p \gamma_p = (\varphi, 0, \dots, 0)^\top.$$

10→11

5.4 Algorithme de Levinson-Durbin

L'**algorithme de Levinson-Durbin** a le mérite d'être récursif, ce qui permet de calculer des prédicteurs successifs sans avoir besoin de tout recalculer. Proposé par Levinson vers 1947 et amélioré par Durbin vers 1960, il s'agit d'un algorithme d'algèbre linéaire permettant de résoudre des équations liées à une matrice de Toeplitz. Il est appliqué ici au cadre spécial des processus stochastiques stationnaires, pour lesquels la matrice de covariance est une matrice de Toeplitz.

Rappelons qu'on souhaite calculer $(\varphi_p)_{p \geq 1}$ et $(\sigma_p^2)_{p \geq 1}$ telles que

$$\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-k} \quad \text{et} \quad \sigma_p^2 = \|E_{t,p}^+\|_2^2 = \text{Var}(E_{t,p}^+),$$

où

$$E_{t,p}^+ = X_t - \text{proj}(X_t, H_{t-1,p})$$

est l'erreur de prédiction progressive (ou directe). L'algorithme de Levinson-Durbin exploite une symétrie par retournement du temps. Pour tous $t \in \mathbb{Z}$ et $p \geq 0$,

$$E_{t,p}^- = X_t - \text{proj}(X_t, H_{t+p,p})$$

est appelée **erreur de prédiction rétrograde** car X_t est prédit par sa projection sur le futur $H_{t+p,p} = \text{vect}\{X_{t+1}, \dots, X_{t+p}\}$ et non pas sur le passé $H_{t-1,p}$ comme dans l'erreur de prédiction progressive $E_{t,p}^+ = X_t - \text{proj}(X_t, H_{t-1,p})$. Si $p = 0$ alors $\text{proj}(X_t, H_{t,0}) = \text{proj}(X_t, \{0\}) = 0$ et $E_{t,p}^- = X_t$.

Théorème 5.9 : Algorithme de Levinson-Durbin

Si Γ_p est inversible pour tout $p \geq 1$, alors les suites $(\varphi_p)_{p \geq 1}$ et $(\sigma_p^2)_{p \geq 1}$ sont calculables

par récurrence en utilisant les formules suivantes :

$$\begin{aligned} \varphi_{p+1,p+1} &= \kappa_{p+1} \quad \text{et} \quad \varphi_{p+1,k} = \varphi_{p,k} - \kappa_{p+1}\varphi_{p,p-(k-1)} \quad \text{pour } 1 \leq k \leq p \\ \sigma_{p+1}^2 &= \sigma_p^2(1 - \kappa_{p+1}^2) \end{aligned}$$

où κ_{p+1} est une quantité auxiliaire calculée par récurrence par

$$\kappa_{p+1} = \frac{\gamma(p+1) - \sum_{k=1}^p \varphi_{p,k} \gamma(p-(k-1))}{\sigma_p^2}$$

avec les valeurs initiales $\kappa_1 = \gamma(1)/\gamma(0)$, $\varphi_{1,1} = \kappa_1$, $\sigma_1^2 = \sigma_0^2(1 - \kappa_1^2)$ (et $\sigma_0^2 = \gamma(0)$).

Note : ne surtout pas apprendre ces formules par cœur !

Note : pour $p = 0$ la meilleure prédiction de X_t est $0 = \mathbb{E}(X_t)$ tandis que la variance de l'erreur de prédiction est $\sigma_0^2 = \mathbb{E}((X_t - 0)^2) = \gamma(0)$.

Démonstration. Formules pour φ . Considérons la décomposition orthogonale « rétrograde » suivante (penser à la suite $X_t, X_{t-1}, \dots, X_{t-p}, X_{t-(p+1)}$)

$$H_{t-1,p+1} = H_{t-1,p} \oplus \text{vect} \left\{ \underbrace{X_{t-(p+1)} - \text{proj}(X_{t-(p+1)}, H_{t-1,p})}_{E_{t-(p+1),p}^-} \right\}.$$

Cette décomposition orthogonale rétrograde permet d'écrire⁵

$$\text{proj}(X_t, H_{t-1,p+1}) = \text{proj}(X_t, H_{t-1,p}) + \underbrace{\text{proj}(X_t, \text{vect}\{E_{t-(p+1),p}^-\})}_{=\kappa_{p+1} E_{t-(p+1),p}^-}$$

où⁶ $\kappa_{p+1} := \langle X_t, E_{t-(p+1),p}^- \rangle / \|E_{t-(p+1),p}^-\|_2^2$, ou encore, grâce à la définition de $E_{t-(p+1),p}^-$,

$$\text{proj}(X_t, H_{t-1,p+1}) = \text{proj}(X_t, H_{t-1,p}) + \kappa_{p+1} (X_{t-(p+1)} - \text{proj}(X_{t-(p+1)}, H_{t-1,p})).$$

D'autre part, par définition de φ_p on a

$$\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-k}.$$

Par stationnarité, les équations de Yule-Walker (théorème 5.2) sont réversibles, et donc les coefficients de prédiction progressifs et rétrogrades coïncident (propriété \star) :

$$\begin{aligned} \text{proj}(X_{t-(p+1)}, H_{t-1,p}) &\stackrel{\star}{=} \sum_{k=1}^p \varphi_{p,k} X_{t-(p+1)+k} = \sum_{k=1}^p \varphi_{p,p-(k-1)} X_{t-k} \\ &\quad \dots, X_{t-(p+1)}, \underbrace{X_{t-p}, \dots, X_{t-1}}_{H_{t-1,p}}, \underbrace{X_t, X_{t+1}, \dots, X_{t+p}}_{H_{t+p,p}}, \dots \end{aligned}$$

On a donc, en utilisant les trois dernières équations ci-dessus,

$$\text{proj}(X_t, H_{t-1,p+1}) = \sum_{k=1}^p (\varphi_{p,k} - \kappa_{p+1} \varphi_{p-(k-1),p}) X_{t-k} + \kappa_{p+1} X_{t-(p+1)}.$$

5. On utilise le fait que la projection orthogonale sur la somme de sous-espaces orthogonaux est égale à la somme des projections orthogonales sur chacun des sous-espaces (ici le dernier est de dimension 1).

6. La formule pour la projection sur un sous-espace de dimension 1 est expliquée dans la figure 5.2.

Comme Γ_p est inversible, les coefficients sont uniques. En identifiant les coefficients ci-dessus avec $\text{proj}(X_t, H_{t-1,p+1}) = \sum_{k=1}^{p+1} \varphi_{p+1,k} X_{t-k}$ il vient $\varphi_{p+1,p+1} = \kappa_{p+1}$ et

$$\varphi_{p+1,k} = \varphi_{p,k} - \kappa_{p+1} \varphi_{p-(k-1),p} \quad (1 \leq k \leq p).$$

Formule pour κ . Rappelons que $\kappa_{p+1} = \langle X_t, E_{t-(p+1),p}^- \rangle / \|E_{t-(p+1),p}^-\|_2^2$. D'une part,

$$\begin{aligned} \langle X_t, E_{t-(p+1),p}^- \rangle &= \langle X_t, X_{t-(p+1)} - \text{proj}(X_{t-(p+1)}, H_{t-1,p}) \rangle \\ &= \gamma(p+1) - \langle X_t, \sum_{k=1}^p \varphi_{p,k} X_{t-(p+1)+k} \rangle \\ &= \gamma(p+1) - \sum_{k=1}^p \varphi_{p,k} \gamma(p-(k-1)), \end{aligned}$$

et d'autre part, comme $\text{proj}(X_{t-(p+1)}, H_{t-1,p}) \perp E_{t-(p+1),p}^-$,

$$\begin{aligned} \|E_{t-(p+1),p}^-\|_2^2 &= \langle X_{t-(p+1)}, X_{t-(p+1)} - \sum_{k=1}^p \varphi_{p,k} X_{t-(p+1)+k} \rangle \\ &= \gamma(0) - \sum_{k=1}^p \varphi_{p,k} \gamma(k) \\ &\stackrel{**}{=} \sigma_p^2 = \|E_{t,p}^+\|_2^2, \end{aligned}$$

où on a utilisé en $(**)$ les équations de Yule-Walker du théorème 5.2, ce qui donne

$$\kappa_{p+1} = \frac{\gamma(p+1) - \sum_{k=1}^p \varphi_{p,k} \gamma(p-(k-1))}{\sigma_p^2}.$$

Formule pour σ (facultatif). Rappelons que $\sigma_p^2 = \|E_{t,p}^+\|_2^2$. On écrit

$$\begin{aligned} E_{t,p+1}^+ &= X_t - \text{proj}(X_t, H_{t-1,p+1}) \\ &= X_t - \text{proj}(X_t, H_{t-1,p}) - \kappa_{p+1} E_{t-(p+1),p}^- \\ &= E_{t,p}^+ - \kappa_{p+1} E_{t-(p+1),p}^- \end{aligned}$$

ce qui donne, par le théorème de Pythagore,

$$\sigma_{p+1}^2 = \|E_{t,p+1}^+\|_2^2 = \|E_{t,p}^+\|_2^2 + \kappa_{p+1}^2 \|E_{t-(p+1),p}^-\|_2^2 - 2\kappa_{p+1} \langle E_{t,p}^+, E_{t-(p+1),p}^- \rangle.$$

Comme $\text{proj}(X_t, H_{t-1,p}) \perp E_{t-(p+1),p}^-$ on a, grâce aux formules de κ_{p+1} et $\|E_{t-(p+1),p}^-\|_2^2$,

$$\begin{aligned} \langle E_{t,p}^+, E_{t-(p+1),p}^- \rangle &= \langle X_t - \text{proj}(X_t, H_{t-1,p}), E_{t-(p+1),p}^- \rangle \\ &= \langle X_t, E_{t-(p+1),p}^- \rangle \\ &= \kappa_{p+1} \|E_{t-(p+1),p}^-\|_2^2 \\ &= \kappa_{p+1} \sigma_p^2, \end{aligned}$$

où la dernière égalité provient d'un calcul déjà effectué plus haut, d'où enfin

$$\sigma_{p+1}^2 = \sigma_p^2 (1 - \kappa_{p+1}^2).$$

□

5.5 Algorithme de Gram-Schmidt (ou des innovations)

La prédiction linéaire optimale ne doit pas être confondue avec la méthode d'orthonormalisation de Gram-Schmidt, que nous étudions à présent. Dans cette section, $(X_n)_{n \geq 1}$ est un processus du second ordre de moyenne nulle, pas forcément stationnaire. Pour simplifier, on note, dans toute cette section, pour tout $n \geq 0$,

$$H_n := \text{vect}\{X_1, \dots, X_n\} \quad \text{et} \quad \sigma_n^2 := \|X_{n+1} - \text{proj}(X_{n+1}, H_n)\|_2^2,$$

avec la convention naturelle $H_0 := \{0\}$. Les vecteurs $(X_k - \text{proj}(X_k, H_{k-1}))_{k \geq 1}$ sont deux à deux orthogonaux dans L^2 car pour tous $1 \leq j < k \leq n$ on a

$$X_j - \text{proj}(X_j, H_{j-1}) \in H_j \subset H_{k-1} \perp X_k - \text{proj}(X_k, H_{k-1}).$$

Considérons la représentation suivante en termes orthogonaux :

$$\begin{aligned} H_n &= \text{vect}\{X_1, X_2 - \text{proj}(X_2, H_1), \dots, X_n - \text{proj}(X_n, H_{n-1})\} \\ &= \bigoplus_{1 \leq k \leq n}^\perp \text{vect}\{X_k - \text{proj}(X_k, H_{k-1})\} \\ &= \bigoplus_{1 \leq k \leq n}^\perp \text{vect}\{X_{n-k+1} - \text{proj}(X_{n-k+1}, H_{n-k})\}. \end{aligned}$$

Or la projection orthogonale sur la somme de sous-espaces deux à deux orthogonaux est égale à la somme des projections orthogonales sur chacun des sous-espaces. Ici les sous-espaces sont tous de dimension 1. Il existe donc un vecteur de coefficients déterministe $\theta_n = (\theta_{n,1}, \dots, \theta_{n,n}) \in \mathbb{R}^n$ tel que

$$\text{proj}(X_{n+1}, H_n) = \sum_{k=1}^n \theta_{n,k} (X_{n-k+1} - \text{proj}(X_{n-k+1}, H_{n-k})).$$

La numérotation des coordonnées de θ_n est similaire à celle des φ_p : on part de X_n , proche de X_{n+1} , pour arriver enfin à X_1 . Le calcul des projetés orthogonaux successifs peut être mené en utilisant l'**algorithme de Gram-Schmidt**. Cet algorithme, qui date du dix-neuvième siècle, est purement hilbertien. Nous l'appliquons ici à l'exemple particulier des processus stochastiques du second ordre (l'espace de Hilbert est L^2), où il est connu également sous le nom d'**algorithme des innovations**. Il est disponible pour tout processus du second ordre, stationnaire ou pas, et ne repose donc pas sur la nature Toeplitz de la matrice de covariance des processus stationnaires, contrairement à l'algorithme de Levinson-Durbin.

Théorème 5.10 : Algorithme d'orthonormalisation de Gram-Schmidt

Soit $(X_n)_{n \geq 1}$ un processus du second ordre de moyenne nulle, **pas forcément stationnaire**. Soit $\gamma(j, k) = \text{Cov}(X_j, X_k) = \mathbb{E}(X_j X_k) = \langle X_j, X_k \rangle$. Si la matrice

$$(\gamma(j, k))_{1 \leq j, k \leq n}$$

est inversible pour tout $n \geq 1$, alors les suites $(\sigma_n^2)_{n \geq 0}$ et $(\theta_n)_{n \geq 1}$ sont calculables **par**

récurrence à partir de $n = 1$ et $\sigma_0^2 = \gamma(1, 1)$ avec les formules :

$$\theta_{n,n-k} = \frac{\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2}{\sigma_k^2}$$

$$\sigma_n^2 = \gamma(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \sigma_j^2,$$

où pour tout $n \geq 1$, le calcul se fait pour $k = 0$, puis $k = 1, \dots$, puis enfin $k = n - 1$.

Note : ne surtout pas apprendre ces formules par cœur !

Note : ici on utilise la convention $\sum_{\emptyset} = 0$.

Note : l'algorithme de Gram-Schmidt consiste à calculer successivement les coefficients θ_n de $\text{proj}(X_{n+1}, H_n)$ sur $X_1, X_2 - \text{proj}(X_2, H_1), \dots, X_n - \text{proj}(X_n, H_{n-1})$. Il est **récuratif**. Il est cependant différent de l'algorithme de Levinson-Durbin, récuratif aussi, qui consiste à calculer les coefficients φ_n de $\text{proj}(X_{n+1}, H_n)$ sur X_1, \dots, X_n .

Démonstration. Le théorème de Pythagore donne

$$\begin{aligned} \sigma_n^2 &= \|X_{n+1} - \text{proj}(X_{n+1}, H_n)\|_2^2 \\ &= \|X_{n+1}\|_2^2 - \|\text{proj}(X_{n+1}, H_n)\|_2^2 \\ &= \gamma(n+1, n+1) - \sum_{k=1}^n \theta_{n,k}^2 \sigma_{n-k}^2 \\ &= \gamma(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 \sigma_k^2 \end{aligned}$$

qui est bien la formule annoncée pour σ_n^2 . La matrice covariance $(\gamma(j, k))_{1 \leq j, k \leq n+1}$ du vecteur aléatoire (X_1, \dots, X_{n+1}) est la matrice de Gram des vecteurs X_1, \dots, X_{n+1} dans L^2 . Elle est inversible ssi ces vecteurs sont linéairement indépendants, ce qui signifie que le parallélépipède engendré par X_1, \dots, X_{n+1} est de dimension pleine $n+1$, ce qui signifie que ses hauteurs $\sigma_0, \sigma_1, \dots, \sigma_n$ sont > 0 . On peut donc diviser par σ_k^2 dans les formules sans souci. Établissons à présent la formule pour $\theta_{n,n-k}$. À nouveau grâce à la même orthogonalité, on obtient, pour tout $0 \leq k < n$,

$$\begin{aligned} &\langle \text{proj}(X_{n+1}, H_n), X_{k+1} - \text{proj}(X_{k+1}, H_k) \rangle \\ &= \sum_{k'=1}^n \theta_{n,k'} \langle X_{n-k'+1} - \text{proj}(X_{n-k'+1}, H_{n-k'}), X_{k+1} - \text{proj}(X_{k+1}, H_k) \rangle \\ &= \theta_{n,n-k} \sigma_k^2. \end{aligned}$$

Or $\langle X_{n+1} - \text{proj}(X_{n+1}, H_n), X_{k+1} - \text{proj}(X_{k+1}, H_k) \rangle = 0$, et donc

$$\begin{aligned} \theta_{n,n-k} &= \frac{\langle X_{n+1}, X_{k+1} - \text{proj}(X_{k+1}, H_k) \rangle}{\sigma_k^2} \\ &= \frac{\gamma(n+1, k+1) - \langle X_{n+1}, \text{proj}(X_{k+1}, H_k) \rangle}{\sigma_k^2}. \end{aligned}$$

Dans cette formule, on peut remplacer les projections par l'expression suivante

$$\begin{aligned}\text{proj}(X_{k+1}, H_k) &= \sum_{j=1}^k \theta_{k,j} (X_{k+1-j} - \text{proj}(X_{k+1-j}, H_{k-j})) \\ &= \sum_{j=0}^{k-1} \theta_{k,k-j} (X_{j+1} - \text{proj}(X_{j+1}, H_j))\end{aligned}$$

ce qui donne

$$\begin{aligned}\theta_{n,n-k} &= \frac{\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \text{proj}(X_{j+1}, H_j) \rangle}{\sigma_k^2} \\ &= \frac{\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2}{\sigma_k^2}.\end{aligned}$$

Ceci est bien la formule annoncée pour $\theta_{n,n-k}$ (par convention, $\sum_{\emptyset} = 0$). □

11→12

Exemple 5.11 : Processus MA(1)

Si X est un processus MA(1) solution de $X_t = Z_t + \theta Z_{t-1}$ où $Z \sim \text{BB}(0, \sigma^2)$, alors

$$\gamma(j, k) = \sigma^2(1 + \theta^2) \mathbf{1}_{j=k} + \theta \sigma^2 \mathbf{1}_{|j-k|=1}.$$

Le théorème 5.10 donne alors après calcul

$$\theta_{n,k} = \frac{\theta \sigma^2}{\sigma_{n-1}^2} \mathbf{1}_{k=1} \quad \text{et} \quad \sigma_n^2 = \sigma^2(1 + \theta^2) - \frac{\theta^2 \sigma^4}{\sigma_{n-1}^2} \mathbf{1}_{n>0}.$$

Si $v_n := \sigma_n^2 / \sigma^2$ alors $v_0 = 1 + \theta^2$, $v_{n+1} = 1 + \theta^2 - \theta^2 / v_n$, et $\theta_{n,1} = \theta / v_{n-1}$, et le prédicteur linéaire optimal $\text{proj}(X_n, H_{n-1})$ vérifie donc la récurrence affine suivante :

$$\begin{aligned}\text{proj}(X_{n+1}, H_n) &= \sum_{k=1}^n \theta_{n,k} (X_{n+1-k} - \text{proj}(X_{n+1-k}, H_{n-k})) \\ &= \frac{\theta(X_n - \text{proj}(X_n, H_{n-1}))}{v_{n-1}}.\end{aligned}$$

5.6 Compléments

Définition 5.12 : Coefficient d'autocorrélation partielle

Le coefficient $\kappa_p (= \varphi_{p,p})$ de l'algorithme de Levinson-Durbin est appelé **coefficient d'autocorrélation partielle**.

Théorème 5.13 : Coefficient d'autocorrélation partielle

Pour tous $p \geq 0$ et $t \in \mathbb{Z}$,

$$\kappa_{p+1}(= \varphi_{p+1,p+1}) = \frac{\langle E_{t,p}^+, E_{t-(p+1),p}^- \rangle}{\|E_{t,p}^+\|_2 \|E_{t-(p+1),p}^-\|_2} = \frac{\text{Cov}(E_{t,p}^+, E_{t-(p+1),p}^-)}{\sqrt{\text{Var}(E_{t,p}^+)} \sqrt{\text{Var}(E_{t-(p+1),p}^-)}}.$$

En particulier, on a $|\kappa_p| \leq 1$ (inégalité de Cauchy-Schwarz).

Note : $E_{t,p}^+$ et $E_{t-(p+1),p}^-$ sont relatives à des projection sur le même espace $H_{t-1,p}$.

Note : $\kappa_1 = \frac{\text{Cov}(X_t, X_{t-1})}{\sqrt{\text{Var}(X_t)} \sqrt{\text{Var}(X_{t-1})}} = \rho(1)$ (fonction d'autocorrélation du processus).

Note : pour un AR(m) causal on trouve $\kappa_p = \varphi_{p,p} \mathbf{1}_{1 \leq p < m} + \varphi_m \mathbf{1}_{p=m}$.

Démonstration. La définition de κ_{p+1} donnée dans la preuve du théorème 5.9, le fait que $X_t = E_{t,p}^+ + \text{proj}(X_t, H_{t-1,p})$, et $\text{proj}(X_t, H_{t-1,p}) \in H_{t-1,p} \perp E_{t-(p+1),p}^-$, donnent

$$\kappa_{p+1} = \frac{\langle X_t, E_{t-(p+1),p}^- \rangle}{\|E_{t-(p+1),p}^-\|_2^2} = \frac{\langle E_{t,p}^+, E_{t-(p+1),p}^- \rangle}{\|E_{t-(p+1),p}^-\|_2^2}.$$

Or la preuve du théorème 5.9 donne $\|E_{t,p}^+\|_2^2 = \|E_{t-(p+1),p}^-\|_2^2$ ($= \sigma_p^2$). □

Estimation (x1)

Mots-clés : Estimation ; moyenne empirique ; autocovariance empirique.

Dans ce chapitre, $X = (X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire réel de moyenne μ et d'autocovariance γ . On cherche à estimer μ et γ à partir de l'observation de X_1, \dots, X_n . Pour cela, on construit des estimateurs, c'est-à-dire des fonctions mesurables de X_1, \dots, X_n , qui approchent les quantités d'intérêt, et dont on étudie les propriétés quand $n \gg 1$. Soulignons que les variables aléatoires X_1, \dots, X_n sont dépendantes. Cependant, elles ont même moyenne, et la matrice de covariance du vecteur aléatoire (X_1, \dots, X_n) est de type Toeplitz : $\Gamma_{j,k} = \text{Cov}(X_j, X_k) = \gamma(j - k)$. Tout se passe comme si nous devions estimer le vecteur moyenne et la matrice de covariance d'un vecteur aléatoire (X_1, \dots, X_n) à partir d'un échantillon de taille 1. On tire partie de la stationnarité, qui fait chuter le nombre de paramètres : la moyenne (μ, \dots, μ) ne dépend que d'un seul paramètre μ au lieu de n paramètres en général, tandis que la matrice de covariance $\Gamma_n = (\gamma(j - k))_{1 \leq j, k \leq n}$ ne dépend que de n paramètres $\gamma(0), \dots, \gamma(n - 1)$ au lieu des $n(n - 1)/2$ paramètres en général. Principe général : on peut estimer un signal de petite dimension plongé dans un espace de grande dimension à condition d'avoir une information structurelle (stationnarité ici).

6.1 Estimation de la moyenne

L'estimateur naturel de la moyenne μ est la **moyenne empirique** définie par

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Si $\hat{\theta}_n$ est un estimateur de θ alors on dispose de la **décomposition biais-variance** :

$$\mathbb{E}((\hat{\theta}_n - \theta)^2) = \text{Var}(\hat{\theta}_n) + (\mathbb{E}(\hat{\theta}_n) - \theta)^2.$$

Autrement dit l'**écart quadratique moyen** est la somme de la variance (erreur statistique) et du carré du biais (erreur systématique). La moyenne empirique \bar{X}_n est un estimateur sans biais de $\theta = \mu$ car grâce à la propriété de stationnarité on a $\mathbb{E}(X_1) = \dots = \mathbb{E}(X_n) = \mu$, ce qui donne, grâce à la linéarité de l'espérance,

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = \mu.$$

Ainsi, la variance et l'écart quadratique moyen coïncident : $\text{Var}(\bar{X}_n) = \mathbb{E}((\bar{X}_n - \mu)^2)$. Le biais de \bar{X}_n ne fait intervenir que les lois marginales de dimension 1 de X et n'est donc pas sensible à l'autocovariance.

On rappelle qu'un processus stationnaire $X = (X_t)_{t \in \mathbb{Z}}$ est un **bruit blanc très fort** $\text{BB}(\mu, \sigma^2)$ lorsque les variables $(X_t)_{t \in \mathbb{Z}}$ sont i.i.d. de moyenne μ et de variance σ^2 .

Théorème 6.1 : Estimation de la moyenne pour les BB très forts

Si $(X_t)_{t \in \mathbb{Z}}$ est un $\text{BB}(\mu, \sigma^2)$ très fort alors la moyenne empirique \bar{X}_n vérifie :

1. Consistance forte : $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mu$;
2. Convergence dans L^2 à vitesse $1/\sqrt{n}$:

$$\|\bar{X}_n - \mu\|_2 = \frac{\sigma}{\sqrt{n}} \quad \text{c'est-à-dire} \quad \mathbb{E}((\bar{X}_n - \mu)^2) = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} ;$$

3. Fluctuations asymptotiques gaussiennes de la version centrée et réduite :

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1)$$

(les deux membres, droite et gauche, sont de moyenne 0 et de variance 1).

Démonstration.

1. Il s'agit de la loi forte des grands nombres. Elle reste vraie quand X est un bruit blanc fort, mais cela n'est pas évident ;
2. Il s'agit d'une conséquence de la stationnarité et de l'absence de corrélation :

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{j,k=1}^n \text{Cov}(X_j, X_k) = \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{\sigma^2}{n}.$$

La propriété reste vraie de manière évidente si $X \sim \text{BB}(0, \sigma^2)$.

3. Il s'agit du théorème limite central.

□

La fluctuation gaussienne permet de construire des intervalles de confiance asymptotiques pour μ . La proximité non asymptotique de \bar{X}_n est quantifiable à l'aide d'inégalités de concentration de la mesure, comme celle de Bienaymé-Tchebychev :

$$\forall r > 0, \quad \mathbb{P}(|\bar{X}_n - \mu| > r) \leq \frac{\sigma^2}{nr^2}.$$

La majoration en $1/r^2$ peut être remplacée par une majoration exponentielle lorsque la loi commune des X_t possède des moments exponentiels finis. Cette inégalité permet de construire des intervalles de confiance non-asymptotiques.

D'autre part, il est possible de quantifier la vitesse de convergence des fluctuations vers la loi gaussienne. Si par exemple $\tau := \mathbb{E}(|X_1|^3) < \infty$, alors le théorème de Berry-Esseen affirme qu'il existe une constante $c > 0$ telle que pour tout $n \geq 1$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq t\right) - \mathbb{P}(Z \leq t) \right| \leq \frac{c\tau}{\sigma^3 \sqrt{n}}, \quad \text{où } Z \sim \mathcal{N}(0, 1).$$

Lorsque σ n'est pas connu, les intervalles de confiance pour μ construits avec l'inégalité de Tchebychev ou avec la fluctuation asymptotique gaussienne ne sont pas utilisables car ils font intervenir σ . Si $\hat{\sigma}_n$ est un estimateur de σ tel que $\hat{\sigma}_n \rightarrow \sigma$ en probabilité quand $n \rightarrow \infty$, alors le lemme de Slutsky¹ permet de remplacer σ par $\hat{\sigma}_n$ dans le résultat de fluctuation asymptotique :

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

Théorème 6.2 : Moyenne empirique et écart quadratique moyen

Soit X un processus stationnaire de moyenne μ et d'autocovariance γ .

1. Si $\lim_{h \rightarrow \infty} \gamma(h) = 0$ alors

$$\text{Var}(\bar{X}_n) = \mathbb{E}((\bar{X}_n - \mu)^2) \xrightarrow[n \rightarrow \infty]{} 0$$

c'est-à-dire

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{L^2} \mu;$$

2. Si $\gamma \in \ell^1(\mathbb{Z})$ c'est-à-dire si $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$, alors

$$\text{Var}(\bar{X}_n) = \mathbb{E}((\bar{X}_n - \mu)^2) \underset{n \rightarrow \infty}{\sim} \frac{\sum_{h \in \mathbb{Z}} \gamma(h)}{n} = \frac{2\pi f(0)}{n}$$

c'est-à-dire

$$\|\bar{X}_n - \mu\|_2 \underset{n \rightarrow \infty}{\sim} \sqrt{\frac{2\pi f(0)}{n}}$$

où f est la densité spectrale du processus X . En particulier, si $X = \mu + F_\alpha Z$ avec $Z = (Z_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$ et $\alpha \in \ell^1(\mathbb{Z})$ alors

$$\text{Var}(\bar{X}_n) \sim \frac{\sigma^2}{n} \left(\sum_{h \in \mathbb{Z}} \alpha_h \right)^2.$$

Note : ici les observations X_1, \dots, X_n ne sont **pas indépendantes** en général.

Démonstration. Grâce à la stationnarité, on a

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \frac{1}{n^2} \sum_{j,k=1}^n \mathbb{E}((X_j - \mu)(X_k - \mu)) \\ &= \frac{1}{n^2} \sum_{j,k=1}^n \gamma(j-k) \\ &= \frac{1}{n^2} \sum_{h=-(n-1)}^{n-1} (n-|h|)\gamma(h) \\ &= \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h). \end{aligned}$$

1. Lemme de Slutsky : si $U_n \rightarrow c$ en probabilité avec c constante et si $V_n \rightarrow V$ en loi alors $f(U_n, V_n) \rightarrow f(c, V)$ en loi, pour toute fonction f continue. Exemple : $f(u, v) = uv$.

Si $\lim_{h \rightarrow \infty} \gamma(h) = 0$ alors $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=-(n-1)}^{n-1} |\gamma(h)| = 0$ (critère de Cesàro), d'où

$$\text{Var}(\bar{X}_n) \leq \frac{1}{n} \sum_{h=-(n-1)}^{n-1} |\gamma(h)| \xrightarrow{n \rightarrow \infty} 0.$$

Si $\gamma \in \ell^1(\mathbb{Z})$, c'est-à-dire que $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$, alors, par convergence dominée,

$$n \text{Var}(\bar{X}_n) = \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sum_{h \in \mathbb{Z}} g_n(h) \xrightarrow{n \rightarrow \infty} \sum_{h \in \mathbb{Z}} \gamma(h)$$

où $g_n(h) := \left(1 - \frac{|h|}{n}\right) \gamma(h) \mathbf{1}_{-n < |h| < n}$ car pour tout $h \in \mathbb{Z}$ on a $\lim_{n \rightarrow \infty} g_n(h) = \gamma(h)$ avec $|g_n(h)| \leq |\gamma(h)|$ et $\gamma \in \ell^1(\mathbb{Z})$. De plus le théorème 4.6 de Herglotz donne

$$\sum_{h \in \mathbb{Z}} \gamma(h) = 2\pi \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-ih0} \gamma(h) = 2\pi f(0).$$

Enfin, dans le cas d'un processus linéaire $X = \mu + F_\alpha Z$ avec $Z \sim \text{BB}(0, \sigma^2)$ et $\alpha \in \ell^1(\mathbb{Z})$ alors $\gamma_X = \gamma_{X-\mu} \in \ell^1(\mathbb{Z})$ et par le théorème 4.12 :

$$f(0) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_h = \frac{\sigma^2}{2\pi} \sum_{h \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_k \alpha_{k+h} = \frac{\sigma^2}{2\pi} \sum_{k, k' \in \mathbb{Z}} \alpha_k \alpha_{k'} = \frac{\sigma^2}{2\pi} \left(\sum_{h \in \mathbb{Z}} \alpha_h \right)^2$$

□

Théorème 6.3 : Moyenne empirique et normalité asymptotique

Soit $X = \mu + F_\alpha Z$ un processus linéaire obtenu à partir d'un bruit blanc **très fort** $Z = (Z_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, \sigma^2)$, et de coefficients $\alpha \in \ell^1(\mathbb{Z})$ tels que $\sum_{h \in \mathbb{Z}} \alpha_h \neq 0$, alors

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \tau^2) \quad \text{où} \quad \tau^2 = \sum_{h \in \mathbb{Z}} \gamma(h) = \sigma^2 \left(\sum_{h \in \mathbb{Z}} \alpha_h \right)^2.$$

Cela permet de construire des intervalles de confiance (asymptotiques) pour μ , à condition toutefois de connaître τ , ou un estimateur de τ (lemme de Slutsky).

Démonstration. On se contente ici du cas où Z est gaussien. Dans ce cas,

$$X_t = \mu + \sum_{h \in \mathbb{Z}} \alpha_h Z_{t-h} \sim \mathcal{N}\left(\mu, \sigma^2 \sum_{h \in \mathbb{Z}} \alpha_h^2\right)$$

et surtout

$$\sqrt{n}(\bar{X}_n - \mu) \sim \mathcal{N}\left(0, \sum_{|h| < n} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right).$$

Or pour les variables aléatoires gaussiennes centrées, la convergence en loi est équivalente à la convergence de la variance, ce qui est immédiat en utilisant les fonctions caractéristiques. Au delà du cas gaussien, on procède également en utilisant les fonctions caractéristiques, mais il faut contrôler des termes de reste. □

6.2 Estimation de l'autocovariance

Dans le cas où les variables aléatoires $(X_t)_{t \in \mathbb{Z}}$ sont i.i.d. de moyenne μ et de variance σ^2 , si μ est connue alors l'estimateur naturel de la variance σ^2 est donné par $\frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2$, qui est convergent d'après la loi des grands nombres, et sans biais, tandis que si μ n'est pas connue alors $\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ est un estimateur sans biais.

Dans le cas d'un processus stationnaire $X = (X_t)_{t \in \mathbb{Z}}$ général, on estime l'autocovariance γ avec l'**autocovariance empirique** définie par

$$\hat{\gamma}_n(h) = \begin{cases} \frac{1}{n} \sum_{k=1}^{n-h} (X_{k+h} - \bar{X}_n)(X_k - \bar{X}_n) & \text{si } 0 \leq h < n; \\ \hat{\gamma}_n(-h) & \text{si } -n < h \leq 0, \end{cases}$$

Ceci conduit à l'**autocorrélation empirique**, définie pour tout h tel que $|h| < n$ par

$$\hat{\rho}(h) = \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}.$$

La **matrice de covariance empirique** $\hat{\Gamma}_n = (\hat{\gamma}(j-k))_{1 \leq j, k \leq n}$ est symétrique et de type positif. En général, l'autocovariance empirique est un estimateur biaisé. Lorsque X est un bruit blanc fort alors il est bien connu qu'on peut rendre $\hat{\gamma}$ sans biais en remplaçant le coefficient de normalisation $1/n$ par $1/(n-h-1)$. L'autocovariance empirique est asymptotiquement sans biais pour une large classe de processus linéaires.

Théorème 6.4 : Autocorrélation empirique et normalité asymptotique

Si X est de la forme $X = \mu + F_\alpha Z$ où $Z = (Z_t)_{t \in \mathbb{Z}}$ est un BB(0, σ^2) **très fort** et $\alpha \in \ell^1(\mathbb{Z})$, et si $\mathbb{E}(Z_t^4) < \infty$ pour tout $t \in \mathbb{Z}$ ou si $\sum_{k \in \mathbb{Z}} |k| \alpha_k^2 < \infty$ alors pour tout $h > 0$ fixé,

$$\sqrt{n} \left(\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} - \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix} \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, T)$$

où la matrice de covariance asymptotique T est donnée par la **formule de Bartlett**

$$T_{j,k} = \sum_{r=1}^{\infty} [\rho(r+j) + \rho(r-j) - 2\rho(j)\rho(r)][\rho(r+k) + \rho(r-k) - 2\rho(k)\rho(r)].$$

En particulier $T_{k,k} = \sum_{r=1}^{\infty} [\rho(r+k) + \rho(r-k) - 2\rho(k)\rho(r)]^2$, et

$$\sqrt{n}(\hat{\rho}(k) - \rho(k)) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, T_{k,k}).$$

6.3 Application aux tests

Faisons l'hypothèse H suivante : les X_t sont i.i.d. d'espérance 0 et de variance σ^2 . Sous l'hypothèse H , on a $\rho(h) = \mathbf{1}_{h=0}$, ce qui donne $T_{j,k} = \mathbf{1}_{j=k}$, et par conséquent, d'après le théorème 6.4, $\sqrt{n}\hat{\rho}(1), \dots, \sqrt{n}\hat{\rho}(h)$ sont approximativement i.i.d. de loi gaussienne $\mathcal{N}(0, 1)$ quand n est grand. Il n'y a plus de σ car il s'agit ici de l'autocorrélation. Il en découle que sous l'hypothèse H , pour tout $0 < \alpha < 1$, quand $n \gg 1$, la probabilité que

$$\sqrt{n}\hat{\rho}(1), \dots, \sqrt{n}\hat{\rho}(h)$$

appartiennent tous à l'intervalle $I_{n,\alpha} = [-q_{1-\alpha/2}, q_{1-\alpha/2}]$ est proche de $(1-\alpha)^h$, où $q_{1-\alpha/2}$ est le quantile $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. Pour $\alpha = 0.05$ on a $q_{1-0.05/2} = q_{0.975} = 1.96$.

Le test du Portmanteau consiste à considérer une statistique moyennisée

$$S_n = n \sum_{k=1}^h \hat{\rho}_k^2.$$

Sous l'hypothèse H , la variable S_n converge en loi quand $n \rightarrow \infty$ vers une loi $\chi^2(h)$. On peut donc choisir $0 < \alpha < 1$ et tester si S_n appartient à l'intervalle $] -\infty, q_{1-\alpha}]$ où $q_{1-\alpha}$ est le quantile $1-\alpha$ de la loi $\chi^2(h)$.

Rappels d'analyse et probabilités

A.1 Espaces de suites

On dit que $(x_k)_{k \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ est sommable lorsque $\lim_{n \rightarrow \infty} \sum_{-n \leq k \leq n} |x_k| < \infty$. Dans ce cas, $\sum_{k \in \mathbb{Z}} x_k = \lim_{m, n \rightarrow \infty} \sum_{-m \leq k \leq n} x_k$ existe. Pour tout $p \in [1, \infty]$ on définit

$$\ell^p(\mathbb{Z}) = \begin{cases} \{x \in \mathbb{R}^{\mathbb{Z}} : \sum_{k \in \mathbb{Z}} |x_k|^p < \infty\} & \text{si } p < \infty \\ \{x \in \mathbb{R}^{\mathbb{Z}} : \sup_{k \in \mathbb{Z}} |x_k| < \infty\} & \text{si } p = \infty. \end{cases}$$

C'est un espace vectoriel normé complet (espace de Banach) pour la norme

$$\|x\|_p = \begin{cases} \left(\sum_{k \in \mathbb{Z}} |x_k|^p \right)^{\frac{1}{p}} & \text{si } p < \infty \\ \sup_{k \in \mathbb{Z}} |x_k| & \text{si } p = \infty. \end{cases}$$

Ces espaces sont emboîtés de manière croissante : pour tous $p, q \in [1, \infty]$,

$$p \leq q \Rightarrow \ell^p(\mathbb{Z}) \subset \ell^q(\mathbb{Z}).$$

Une astuce mnémotechnique pour le retrouver rapidement consiste à penser que le terme général d'une série convergente est borné, de module < 1 pour $|k| \gg 1$.

L'inégalité de Hölder affirme que pour tous $p, q \in [1, \infty]$ avec $1/p + 1/q = 1$,

$$\forall x \in \ell^p(\mathbb{Z}), \forall y \in \ell^q(\mathbb{Z}), \quad \|xy\|_1 \leq \|x\|_p \|y\|_q,$$

où $(xy)_k = x_k y_k$ pour tout $k \in \mathbb{Z}$. Il est parfois commode de définir les espaces $\ell^p(\mathbb{Z})$ pour les suites à valeurs dans \mathbb{C} plutôt que dans \mathbb{R} . L'espace $\ell^2(\mathbb{Z})$ possède une structure d'espace de Hilbert donnée par le produit scalaire $\langle x, y \rangle = \sum_{k \in \mathbb{Z}} x_k \overline{y_k}$. On peut noter si nécessaire $\ell_{\mathbb{C}}^p(\mathbb{Z})$ pour indiquer qu'il s'agit de la version complexe.

A.2 Espaces de fonctions

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et L^0 l'ensemble des variables aléatoires $(\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ (quotient par l'égalité presque sûre). Pour tout $p \in [1, \infty]$ on définit

$$L^p = \begin{cases} \{X \in L^0 : \mathbb{E}(|X|^p) < \infty\} & \text{si } p < \infty \\ \{X \in L^0 : \text{ess sup } |X| < \infty\} & \text{si } p = \infty \end{cases}$$

où $\text{esssup}(X) = \inf\{c \in \mathbb{R} \cup \{+\infty\} : X \leq c \text{ p.s.}\}$. L'ensemble L^p est un espace vectoriel normé complet (on dit espace de Banach) pour la norme

$$\|X\|_p := \begin{cases} \mathbb{E}(|X|^p)^{1/p} & \text{si } p < \infty \\ \text{ess sup}(|X|) & \text{si } p = \infty. \end{cases}$$

L'inégalité de Hölder affirme que si $X \in L^p$ et $Y \in L^q$ avec $p, q \in [1, \infty]$ vérifiant $r = 1/(1/p + 1/q) \geq 1$ avec la convention $1/\infty = 0$ et $1/0 = \infty$, alors $XY \in L^r$ et

$$\|XY\|_r \leq \|X\|_p \|Y\|_q.$$

En posant $Y = 1 \in L^q$ pour tout $q \in [1, \infty]$ on obtient que les espaces sont emboîtés de manière décroissante : pour tous $p, q \in [1, \infty]$, si $p \leq q$ alors $L^q \subset L^p$ (pour le retrouver rapidement, penser au fait que toute v.a. bornée est intégrable). Pour $p = q = 1/2$ on obtient $r = 1/(1/p + 1/q) = 1$ et l'inégalité de Hölder devient :

$$\forall X, Y \in L^2, \quad XY \in L^2 \quad \text{et} \quad \mathbb{E}(|XY|) = \|XY\|_1 \leq \|X\|_2 \|Y\|_2.$$

Notons que $\ell^p(\mathbb{Z}) = L^p(\mathbb{Z})$ où \mathbb{Z} est muni de la tribu de ses parties et de la mesure de comptage. D'autre part, tout comme pour les espace ℓ^p , il est parfois commode de définir les espaces L^p pour les fonctions à valeurs dans \mathbb{C} plutôt que dans \mathbb{R} , et on peut noter si nécessaire $L_{\mathbb{C}}^p$ pour indiquer qu'il s'agit de la version complexe.

L'inégalité précédente permet de définir l'application

$$(X, Y) \in L^2 \times L^2 \mapsto \langle X, Y \rangle = \mathbb{E}(XY) \in \mathbb{R}.$$

Elle constitue un produit scalaire qui fait de L^2 un espace de Hilbert. L'inégalité de Cauchy-Schwarz affirme que pour tous $X, Y \in L^2$,

$$|\mathbb{E}(XY)| = |\langle X, Y \rangle| \leq \|X\|_2 \|Y\|_2 = \sqrt{\mathbb{E}(|X|^2)} \sqrt{\mathbb{E}(|Y|^2)}.$$

Le théorème de Pythagore s'écrit

$$\|X + Y\|_2^2 = \|X\|_2^2 + \|Y\|_2^2 + 2\langle X, Y \rangle.$$

ce qui donne l'identité du parallélogramme :

$$\|X - Y\|_2^2 + \|X + Y\|_2^2 = 2\|X\|_2^2 + 2\|Y\|_2^2.$$

L'inégalité de Jensen affirme que si φ est convexe alors pour tout $X \in L^1$, on a

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

Exemples courants : $|\mathbb{E}(X)|^p \leq \mathbb{E}(|X|^p)$, pour tout $p \geq 1$, $e^{\mathbb{E}(X)} \leq \mathbb{E}(e^X)$ si X réelle, $-\log(\mathbb{E}(X)) \leq -\mathbb{E}(\log(X))$ si $X > 0$.

Deux variables aléatoires X et Y à valeurs dans K sont indépendantes ssi

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$$

pour tous $f, g \in \mathcal{F}$, où \mathcal{F} est une classe de fonctions assez grande et pour lesquelles toutes les espérances dans cette formule ont un sens. Par exemple, on peut prendre pour \mathcal{F} l'ensemble des fonctions mesurables bornées, ou l'ensemble des fonctions mesurables positives. On peut aussi prendre les fonctions trigonométriques (fonctions caractéristiques!).

Si $X, Y \in L^2$ alors $XY \in L^1$, grâce à l'inégalité de Cauchy-Schwarz. D'autre part, si $X, Y \in L^1$ sont indépendantes, alors $XY \in L^1$ car $\mathbb{E}|XY| = \mathbb{E}|X|\mathbb{E}|Y| < \infty$.

A.3 Covariance

La variance de $X \in L^2$ est définie par

$$\text{Var}(X) = \|X - \mathbb{E}(X)\|_2^2 = \mathbb{E}(|X - \mathbb{E}(X)|^2) = \mathbb{E}(|X|^2) - |\mathbb{E}(X)|^2.$$

On a aussi la formule variationnelle

$$\text{Var}(X) = \inf_{c \in \mathbb{R}} \|X - c\|_2^2 \quad \text{atteint pour } c = \mathbb{E}(X).$$

La covariance de $X, Y \in L^2$ est définie par

$$\begin{aligned} \text{Cov}(X, Y) &= \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle \\ &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

On a

$$\text{Cov}(X, X) = \text{Var}(X) \quad \text{et} \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

On dit que X et Y ne sont pas corrélées lorsque $\text{Cov}(X, Y) = 0$. Si X et Y sont indépendantes alors X et Y ne sont pas corrélées, mais la réciproque est fausse en général, sauf si le couple (X, Y) est un vecteur gaussien.

Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire de \mathbb{C}^n alors

$$\text{Var}(X_1 + \dots + X_n) = \sum_{1 \leq j, k \leq n} \text{Cov}(X_j, X_k) = \sum_{j=1}^n \text{Var}(X_j) + \sum_{1 \leq j \neq k \leq n} \text{Cov}(X_j, X_k).$$

Géométriquement, pour les vecteurs aléatoires centrés de carré (du module) intégrable, covariance nulle signifie orthogonalité. Si X_1, \dots, X_n ne sont centrés, de variance unité, et non corrélées, alors $\text{Var}(X_1 + \dots + X_n) = \|X_1 + \dots + X_n\|_2^2 = n$. Penser au fait que la diagonale $[(0, \dots, 0), (1, \dots, 1)]$ du cube $[0, 1]^n$ de \mathbb{R}^n a pour longueur \sqrt{n} .

A.4 Vecteur moyenne et matrice de covariance

Le **vecteur moyenne** μ_X et la matrice de covariance Γ_X d'un vecteur aléatoire (colonne) de \mathbb{R}^n sont définis par

$$\mu_X = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))^T \quad \text{et} \quad \Gamma_X = (\text{Cov}(X_j, X_k))_{1 \leq j, k \leq n}.$$

La matrice de covariance Γ_X est réelle et symétrique.

Pour toute matrice carrée $A \in \mathcal{M}_n(\mathbb{R})$, les propriétés suivantes sont équivalentes :

1. A est symétrique de type positif sur \mathbb{R} : pour tous indices j, k , tout vecteur $v \in \mathbb{R}^n$,

$$A(j, k) = A(k, j) \quad \text{et} \quad \langle Av, v \rangle = \sum_{1 \leq j, k \leq n} v_j v_k A(j, k) \geq 0;$$

2. A est symétrique de type positif sur \mathbb{C} : pour tous indices j, k , tout vecteur $v \in \mathbb{C}^n$,

$$A(j, k) = A(k, j) \quad \text{et} \quad \langle Av, v \rangle = \sum_{1 \leq j, k \leq n} v_j \overline{v_k} A(j, k) \geq 0;$$

3. A est symétrique et toutes ses valeurs propres sont ≥ 0 ;
4. A est la matrice de covariance d'un vecteur aléatoire de \mathbb{R}^n ;
5. $A = LL^\top$ où $L \in \mathcal{M}_n(\mathbb{R})$ est triangulaire inférieure (décomposition Cholesky).

La décomposition de Cholesky est un cas particulier de la décomposition LU.

Le **spectre** $\text{spec}(A) \subset \mathbb{C}$ de $A \in \mathcal{M}_n(\mathbb{C})$ est l'ensemble des valeurs propres de A . Les **Valeurs propres** de A sont les racines dans \mathbb{C} de son **polynôme caractéristique**

$$P(z) = \det(A - zI).$$

On a $\lambda \in \text{spec}(A)$ ssi

$$\ker(A - \lambda I) \neq \{0\},$$

c'est-à-dire ssi il existe un vecteur $v \in \mathbb{C}^n$, appelé **vecteur propre** associé à λ , tel que $v \neq 0$ et $Av = \lambda v$. Le **théorème spectral** de l'algèbre linéaire assure que toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ symétrique admet une base orthonormale de vecteurs propres réels : il existe une matrice orthogonale $O \in \mathcal{M}_n(\mathbb{R})$, c'est-à-dire telle que $O^{-1} = O^\top$, telle que

$$D = O^\top A O = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathcal{M}_n(\mathbb{R}), \quad \text{c'est-à-dire} \quad A = O D O^\top.$$

Ici les réels $\lambda_1, \dots, \lambda_n$ sont les valeurs propres de A , tandis que les colonnes v_1, \dots, v_n de O sont les vecteurs propres orthonormés associés, et pour tout vecteur $x \in \mathbb{R}^n$,

$$\langle Ax, x \rangle = \langle D O^\top x, O^\top x \rangle = \sum_{k=1}^n \lambda_k \langle x, v_k \rangle v_k.$$

Les **formules variationnelles de Rayleigh-Courant-Fischer** assurent que pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ symétrique,

$$\min\{\lambda \text{ valeur propre de } A\} = \min_{x: \|x\|_2=1} \langle Ax, x \rangle$$

et

$$\max\{\lambda \text{ valeur propre de } A\} = \max_{x: \|x\|_2=1} \langle Ax, x \rangle.$$

Bibliographie

- [Ar] Y. Aragon. *Séries temporelles avec R* (bibliothèque d'étude 519.5 ARA) (disponible sur Internet)
- [Am] G. Ammar. *Classical foundations of algorithms for solving positive definite Toeplitz equations* CALCOLO (1996), volume 33, issue 1-2, pp 99–113 (disponible sur Internet)
- [C] A. Charpentier. *Cours de séries temporelles – Théorie et application* (disponible sur Internet)
- [LL] C. Lévy-Leduc. *Introduction à l'étude des séries temporelles* (disponible sur Internet)
- [OPV] G. Oppenheim, A. Philippe, M.-C. Viano. *Cours de séries temporelles* (dispo. sur Internet)
- [LMR] C. Lévy-Leduc, É. Moulines, F. Roueff. *Séries temporelles : théorie et méthodes* (dispo. sur Internet)
- [BD1] P. Brockwell, R. Davis. *Time Series : Theory and Methods* (bib. rech. 519.5 BRO)
- [BD2] P. Brockwell, R. Davis. *Intro. to time series and forecasting* (bib. rech. 519.5 BRO)
- [CC] J. Cryer, K.-S. Chan. *Time Series Analysis – With Applications in R* (bib. étu. 519.5 CRY)
- [BJ] G. Box, G. Jenkins. *Time Series Analysis – Forecasting and Control* (bib. rec. 519.5 BRO)