

Université Paris-Sud
Master Ingénierie Mathématiques, deuxième année
Année scolaire 2010–2011

Séries chronologiques

Volume 1: cours et exercices

Responsable des cours:

Michel Prenat

Responsables des travaux dirigés:

Christine Keribin

Raphaël Rossignol

Ce volume est complété par un autre volume qui contient les énoncés des TD sur machine.

Date de mise à jour: 27 septembre 2010

Table des matières

Table des matières	iii
Avant-propos	vii
1 Introduction	1
1.1 Les pré-requis	1
1.2 Définition d'une série chronologique univariée	2
1.3 Classification des séries chronologiques	2
1.3.1 Domaines d'application	2
1.3.2 Séries réelles / complexes	3
1.3.3 Stationnarité, tendances et saisonnalité	3
1.3.4 Représentation temporelle / représentation spectrale	7
1.4 Pourquoi étudier une série chronologique ?	10
1.4.1 Quelques opérations de base	11
1.4.2 Quelques exemples d'utilisation de modèles de séries chronologiques	12
1.5 Méthodologie générale d'analyse	13
1.6 Exercice complémentaire	14
2 Représentation spectrale	15
2.1 Densité et distribution spectrales	15
2.1.1 Définition et propriétés de la fonction de covariance	15
2.1.2 Densité spectrale d'un processus aléatoire stationnaire	16
2.1.3 Distribution spectrale d'un processus aléatoire stationnaire	17
2.1.4 Filtre linéaire appliqué à un processus aléatoire : transformation de la densité spectrale	17
2.2 Densité spectrale d'un processus ARMA	17
2.3 Décomposition de Wold	17
2.3.1 Distribution spectrale	17
2.3.2 Densité spectrale	18
2.3.3 Décomposition de Wold	18
2.4 Estimation spectrale non paramétrique	19
2.4.1 Le périodogramme : propriétés, limitations	19
2.4.2 Estimateurs dérivés du périodogramme	22
2.5 Estimation spectrale paramétrique	22
2.5.1 Identification à un processus ARMA	23
2.5.2 Multiple Signal Classification Method (MUSIC)	23
2.6 Analyse des séries chronologiques	23

3 Tendances et facteurs saisonniers	25
3.1 Définitions et traitements	25
3.1.1 Les modèles	25
3.1.2 Les phases de traitement	26
3.2 Analyse	26
3.2.1 Analyse visuelle	26
3.2.2 Analyse par fonction de covariance échantillonnée	26
3.2.3 Analyse spectrale	28
3.3 Traitement des tendances	29
3.3.1 Estimation de tendance par la méthode des moindres carrés	29
3.3.2 Lissage par la méthode de la Moyenne Mobile	31
3.3.3 Elimination par différentiation	32
3.3.4 Tendances de formes spécifiques	33
3.3.5 Traitement des tendances et estimation de fonctions	35
3.4 Traitement global	36
3.4.1 Méthode des "tendances faibles"	36
3.4.2 Estimation par la méthode de la moyenne mobile	36
3.4.3 Elimination par différentiation	37
3.5 Elimination de composantes spectrales	38
3.6 Exercices complémentaires	40
4 Prédiction linéaire - Modèles d'état - Filtrage de Kalman	41
4.1 Prédiction linéaire	41
4.1.1 Prédiction linéaire d'une variable aléatoire réelle	41
4.1.2 Prédiction linéaire d'un vecteur aléatoire	42
4.1.3 Minimiser les erreurs de prédiction, maximum de vraisemblance et cadre gaussien	42
4.2 Les modèles d'état	43
4.3 Stationnarité	44
4.4 Le filtre de Kalman	45
4.4.1 Equations du filtre	46
4.4.2 Le filtre d'information	47
4.5 Observations manquantes	48
4.6 Observations irrégulières	49
4.7 Filtre de Kalman étendu	49
5 Les processus ARMA	51
5.1 Définitions et importance pratique	51
5.2 Covariance, corrélation, corrélation partielle	53
5.2.1 Formulation générale de la fonction de covariance d'un processus <i>ARMA</i> causal	54
5.2.2 Fonction de corrélation partielle	54
5.3 Modèle d'état pour un processus ARMA	56
5.3.1 Modèle d'état pour un AR(p)	56
5.3.2 Modèle d'état pour un ARMA(p,q)	57
5.4 Densité spectrale d'un processus ARMA	58
5.5 Calcul de la covariance d'un processus ARMA	58
5.5.1 Méthode directe	59
5.5.2 Equation aux différences	59
5.6 Identification des processus ARMA	60
5.6.1 Définition	60

5.6.2	Critères de présomption de processus ARMA	60
5.6.3	Principes d'identification	60
5.6.4	Détermination de l'ordre d'un MA et d'un AR	62
5.6.5	Identification "robuste" d'un processus ARMA	65
5.6.6	Identification paramétrique d'un processus ARMA	71
5.6.7	Validation d'un modèle : critères de contrôle sur le résidu	72
5.6.8	Méthodes de sélection	74
5.7	Estimation spectrale des processus ARMA	75
5.8	Prédiction linéaire à h pas	76
5.8.1	Filtrage de Kalman	76
5.8.2	Combinaison linéaire des X_i	76
5.8.3	La méthode des innovations	76
5.9	Exercices complémentaires	77
6	Processus ARIMA et SARIMA - Processus à corrélation périodique	79
6.1	Les processus ARIMA	79
6.1.1	Définitions, exemples, propriétés	79
6.1.2	Prédiction des processus ARIMA	80
6.1.3	Principes d'identification	81
6.1.4	Représentation par modèle d'état	82
6.2	Les processus SARIMA	82
6.2.1	Définitions, exemples, propriétés	82
6.2.2	Principes d'identification	83
6.3	Les processus à corrélation périodique	84
7	Processus à mémoire longue	87
7.1	Quelques rappels	87
7.1.1	Décroissance de la fonction de corrélation des processus ARMA	87
7.1.2	Densité spectrale des processus ARMA	89
7.1.3	Les processus ARIMA	89
7.2	Les processus à mémoire longue	90
7.2.1	Définition des processus à mémoire longue	90
7.2.2	Bruit blanc intégré avec un ordre fractionnaire : $ARIMA(0, d, 0)$	90
7.2.3	Les processus ARFIMA (p,d,q)	93
7.2.4	Quelques principes d'identification	94
8	Processus GARCH	101
8.1	Rappel sur les séries chronologiques	101
8.2	Processus GARCH - Processus ARMA/GARCH	101
8.3	Propriétés des processus ARMA/GARCH	104
8.4	Principes d'identification	105
8.4.1	Tester la présence d'un effet "GARCH" sur le résidu	107
8.4.2	Calcul des paramètres d'un processus "GARCH"	108
8.4.3	Tester la validité du modèle (Model Checking)	109
8.4.4	Estimation globale d'un processus ARMA/GARCH	109
8.5	Les processus M-GARCH	110
	Conclusion : prédiction et synthèse	113
	Correction des exercices	115
	Bibliographie	143

Avant-propos

Ce cours est une introduction à l'étude des séries chronologiques, appelées encore séries temporelles ("time series" en anglais). Les séries chronologiques, c'est-à-dire les collections de mesures ordonnées dans le temps, sont présentes un peu partout. En conséquence, elles intéressent beaucoup de gens différents : différents par la nature des phénomènes qu'ils étudient et par les buts qu'ils se fixent dans leur étude. Ces différences ne sont pas anodines, et nécessitent des théories différentes : les séries temporelles univariées ou multivariées, linéaires ou non linéaires, etc. Cet avant-propos est avant tout un avertissement au lecteur : si de nombreux modèles et types de traitements existent, on n'explorera dans ce cours que les modèles les plus simples, et on se restreindra essentiellement au cadre *linéaire, stationnaire (ou stationnaire après dérivation ou extraction d'une partie déterministe), ergodique et univarié*, ce qui fera des processus *ARMA* (et de leurs variantes) l'objet d'étude principal. Un processus (stationnaire) $(Y_t)_{t \in \mathbb{Z}}$ est dit *ARMA*(p, q) s'il existe une série $(Z_t)_{t \in \mathbb{Z}}$ décorrélée, de moyenne nulle et de variance finie constante, telle que :

$$\forall t \in \mathbb{Z}, Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

où $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ sont des coefficients réels. Le bruit Z est appelé bruit blanc faible, et on dit qu'il pilote l'*ARMA* Y . Comme toujours en statistique, nous voulons insister sur la nécessité de distinguer deux situations d'utilisation des *modèles* de ce cours :

- la première situation est celle où il y a de très bonnes raisons de se limiter à ce type de modèles. Par exemple, le modèle a une justification physique, ce qui peut arriver avec les modèles *ARMA* en traitement du signal. On peut alors avoir confiance dans notre modèle, *il a en fait déjà été validé*. Remarquons que dans ce cadre, en général, les paramètres ont un sens très précis,
- la seconde situation est celle où il n'y a pas de raison de se limiter à ce type de modèles. Dans ce cas, même si divers tests basés sur l'étude des résidus ne rejettent pas le modèle, il ne faudrait surtout pas s'en contenter : ces tests sont peu puissants contre une alternative générale. Dans ce genre de situation, les paramètres du modèle, et donc le modèle lui-même, n'ont pas, en général, de sens solide. Le but de l'étude n'est alors pas l'estimation des paramètres, mais plus souvent un problème de prédiction ou de filtrage. Bref, il sera indispensable de ne pas oublier qu'il existe d'autres modèles, non linéaires en particulier, et de confronter divers candidats par des méthodes de validation (cf. [14], chapitre 7).

Bien sûr, la première situation est assez rare, et il sera donc toujours sage de valider un modèle, même lorsqu'on a relativement confiance en lui (cf. section 5.6.7).

Venons-en à une description plus détaillée du contenu. Ce cours comprend huit chapitres. On trouve dans chaque chapitre des énoncés d'exercices de cours, et à la fin de certains des exercices complémentaires. Les corrigés de la plupart des exercices sont reportés à la fin du cours.

Après une introduction détaillée dans le premier chapitre, le second est consacré à l'analyse spectrale des processus et de leurs réalisations. Cette analyse est utile à la fois pour détecter d'éventuels phénomènes périodiques et pour reconnaître un phénomène *ARMA*, puisque les modèles *ARMA* sont caractérisés par leur spectre rationnel.

Le troisième chapitre complète l'analyse préliminaire d'une série univariée en présentant les tendances et facteurs saisonniers : il vise à apprendre à décrire et éliminer les composantes déterministes d'une série temporelle, ce qui est souvent nécessaire pour, plus tard, envisager de modéliser sa partie aléatoire. Notons déjà que les modèles *SARIMA*, présentés au chapitre 6, pourront inclure une partie de ces composantes déterministes, court-circuitant en quelque sorte cette étape.

Le chapitre 4 est consacré aux modèles d'états et au filtrage de Kalman. Les modèles d'états forment une classe particulièrement intéressante et générale de modèles de séries temporelles, contenant notamment les processus *ARMA* et *SARIMA*. Ce chapitre nous sera d'ailleurs utile pour l'identification des processus *ARMA*.

Les chapitres 5 et 6 sont consacrés respectivement aux processus *ARMA* et *SARIMA*, dont les *ARMA* forment un sous-groupe. On présente les méthodes classiques d'estimation et de prédiction. Les *SARIMA* sont des *ARMA* qui peuvent être intégrés et saisonniers et donc ne sont plus stationnaires. La plupart des résultats requièrent que le bruit blanc pilotant le *SARIMA* soit *i.i.d.*

Dans les deux derniers chapitres, on sort du cadre classique des *SARIMA* avec bruit *i.i.d.* en considérant deux directions possibles.

Dans le chapitre 7, c'est essentiellement la relation d'autorégression que l'on modifie en nous intéressant aux processus *ARFIMA*, qui sont des processus à mémoire longue, et n'ont pas de représentation en modèle d'état de dimension finie.

Dans le chapitre 8, ce sont les propriétés du bruit blanc faible pilotant l'*ARMA* qui sont modifiées : on ne suppose plus qu'il est *i.i.d.*, mais qu'il est conditionnellement hétéroscédastique. On obtient les processus *ARMA/GARCH*.

Pour ce qui est de la bibliographie, ce cours est basé essentiellement sur le livre [5], sauf pour le chapitre consacré aux processus (G)ARCH ([17]). D'autres sources sont utilisées, en particulier pour le chapitre d'introduction (exemples de séries chronologiques, paragraphe 1.4 "Pourquoi étudier une série chronologique?") et également dans la logique générale du cours, qui fait une place plus importante que [5] aux modèles d'état et au filtrage de Kalman. Ce choix a été fait car le cours est orienté vers la pratique plus que vers la théorie : le filtrage de Kalman permet de mettre en oeuvre de façon aisée l'estimation des modèles de nombreuses séries chronologiques.

Pour la même raison, le cours lui-même comprend un nombre limité de démonstrations, et les exercices permettent d'approfondir certains aspects théoriques.

Les techniques d'extrêmes appliquées aux séries chronologiques sont importantes, par exemple pour déterminer la probabilité de survenue d'événements rares ou leurs durées (dépassement de températures élevées). Ces techniques ne sont pas développées dans ce cours, mais dans un autre module de la formation.

De la même façon, les méthodes de recherche de variables explicatives ne sont pas étudiées ici, ce qui est une des raisons pour lesquelles l'accent est mis sur les séries univariées.

Ce cours correspond à un volume de 30 heures. Il est accompagné d'une série de Travaux Dirigés (30 heures également) pour lesquels le logiciel R sera utilisé. Les énoncés des Travaux Dirigés sont présentés dans un autre volume.

Chapitre 1

Introduction

Ce chapitre a pour objectif de commencer à familiariser le lecteur avec les principaux aspects des séries chronologiques. Il commence par une liste des pré-requis supposés assimilés par le lecteur. On y donne ensuite la définition des séries chronologiques (multivariées en général), des éléments de classification des séries chronologiques selon plusieurs critères, des exemples pratiques, quelques notions utiles pour la suite (processus *ARMA*, fonctions de covariance et de corrélation, représentation spectrale), ainsi que les raisons qui motivent l'étude des séries chronologiques. Ces notions sont reprises en détail dans les chapitres suivants. Le cours est focalisé sur les séries univariées ; les séries multivariées ne sont évoquées que pour les processus *PARMA* et dans le chapitre consacré aux processus *GARCH*.

1.1 Les pré-requis

D'un point de vue *théorique*, ce cours de séries chronologiques peut être considéré comme un aperçu de la statistique des processus à temps discret. Les pré-requis essentiels sont donc naturellement les bases des probabilités et de la statistique. Plus précisément, en probabilités :

- variable aléatoire, moyenne, variance, indépendance, loi d'un vecteur aléatoire, matrice de covariance, vecteur gaussien,
- loi conditionnelle,
- convergence en proba, presque sûre, L^2 et en loi (notions peu utilisées, surtout dans les exercices complémentaires),
- ergodicité,
- loi des grands nombres et théorème central limite.

et en statistique :

- estimateur, biais, et intervalle de confiance,
- test d'hypothèse,
- modèle linéaire gaussien.

Toutes ces notions (et bien plus !) peuvent être trouvées dans le livre [6]. Quelques connaissances de bases d'algèbre et d'analyse sont essentielles. Plus précisément, on peut citer, pour l'algèbre :

- pour les équations d'état (chapitre 4) : valeurs propres, vecteurs propres, déterminant,
- polynômes.

Et pour l'analyse :

- séries entières et fonctions génératrices,
- transformée de Fourier sur $l^1(\mathbb{Z})$.

Pour ces notions d'algèbre et d'analyse, n'importe quel bon cours de première et deuxième années de Licence de mathématiques fera l'affaire, par exemple [12] et [13].

D'un point de vue *pratique*, on utilisera les logiciels *Matlab* et *R*, et il est vivement conseillé d'être familier avec les bases de la programmation sur l'un de ces logiciels au moins. Pour *Matlab*, et pour ne citer qu'une référence, on pourra consulter :

<http://www.ann.jussieu.fr/~postel/matlab/>

Pour R, l'ouvrage d'Emmanuel Paradis fait figure d'excellente introduction :

http://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf

1.2 Définition d'une série chronologique univariée

Un *processus aléatoire* X indexé par un ensemble T (a priori quelconque) est une famille X_t , où $t \in T$, de vecteurs aléatoires à valeurs dans l'espace d'états $E = \mathbb{R}^k$ ou $E = \mathbb{C}^k$. Si $E = \mathbb{R}$ ou $E = \mathbb{C}$ (i.e $k = 1$), le processus est dit *unidimensionnel* (ou *univarié*).

Un processus aléatoire univarié X_t indexé par T est *du second ordre* si $X \in L^2(\Omega, P)$ pour tout t dans T (les moyennes $\mathbb{E}(X_t)$ et covariances $\mathbb{E}(X_t X_s) - \mathbb{E}(X_t)\mathbb{E}(X_s)$ existent et sont finies pour tout couple (t, s) de valeurs de T).

Si de plus $T = \mathbb{Z}$ ou $T = \mathbb{N}$, X est une *série chronologique* (ou *temporelle*) *univariée*, car les éléments de T correspondent implicitement à des instants. Cette définition permet de couvrir le cas des séries non régulièrement espacées dans le temps.

La définition des séries chronologiques multidimensionnelles ou multivariées se déduit facilement ; ces séries ne sont considérées dans la suite que de façon exceptionnelle (processus GARCH multivariés).

Dans la pratique, une *réalisation* sur une *durée finie* d'un tel processus est également appelée *série chronologique* par abus de langage ; cette réalisation correspond à l'observation à des instants discrets d'une quantité variable, réelle ou complexe. Si le processus sous-jacent est ergodique (voir en particulier les processus ARMA), on peut déduire ses caractéristiques de l'observation d'une réalisation.

1.3 Classification des séries chronologiques

On peut classer les séries chronologiques selon des critères variés : domaines d'application, séries réelles / complexes, séries stationnaires ou non stationnaires (avec tendances, avec facteurs saisonniers, processus intégrés), séries représentées de façon temporelle ou spectrale.

1.3.1 Domaines d'application

On trouve des exemples de séries chronologiques univariées dans de très nombreux domaines. La liste suivante n'est qu'un échantillon :

- finance et économétrie : évolution des indices boursiers, des prix, des données économiques des entreprises, des ventes et achats de biens, des productions agricoles ou industrielles,
- assurance : analyse des sinistres,
- médecine / biologie : suivi des évolutions des pathologies, analyse d'électro-encéphalogrammes et d'électrocardiogrammes,
- sciences de la Terre et de l'Espace : indices de marées, variations des phénomènes physiques (météorologie), évolution des taches solaires, phénomènes d'avalanches,
- traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole,
- traitement des données : mesures successives de position ou de direction d'un objet mobile (trajectographie),
- métrologie : variation de phase ou de fréquence des oscillateurs (voir par exemple [16] où l'on voit que dans un laser, un bruit stationnaire de position sur un miroir conduit à un bruit de phase stationnaire sur l'onde produite, alors qu'un bruit stationnaire sur la longueur de la cavité laser se traduit par un bruit stationnaire sur la fréquence de l'onde, c'est-à-dire une marche au hasard sur sa phase), dérive et bruit des capteurs inertiels.

On voit aux Figures 1.1 à 1.5 quelques exemples de séries chronologiques.

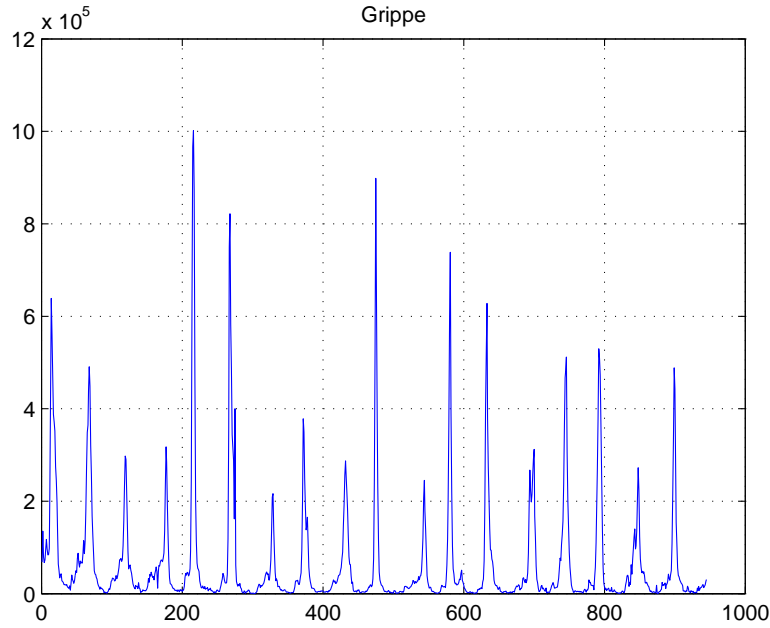


FIG. 1.1 – Evolution des cas de grippe en France, par semaine, de 1984 (semaine 44) à 2002 (semaine 50), soit 945 valeurs.

1.3.2 Séries réelles / complexes

La plupart des applications énumérées ci-dessus conduisent à des séries chronologiques réelles. Les séries complexes se rencontrent essentiellement en traitement du signal.

1.3.3 Séries stationnaires, séries non stationnaires, séries avec tendances et facteurs saisonniers

Ces définitions sont données dans un premier temps pour des processus réels (à valeurs dans \mathbb{R}).

Fonction de covariance d'un processus aléatoire :

$$\text{Cov}(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}([X_t - \mathbb{E}(X_t)][X_s - \mathbb{E}(X_s)])$$

Rappels sur la stationnarité d'un processus aléatoire :

- stationnarité au sens large (“weak stationarity”) :

$$\begin{cases} \forall t \in \mathbb{Z}, \mathbb{E}(|X_t|^2) < \infty, \\ t \mapsto \mathbb{E}(X_t) \text{ est constante}, \\ \forall r, s, t \in \mathbb{Z}, \text{Cov}(X_r, X_s) = \text{Cov}(X_{r+t}, X_{s+t}). \end{cases}$$

- stationnarité au sens strict : les distributions de probabilité conjointes de X_{t_1}, \dots, X_{t_k} et de $X_{t_1+h}, \dots, X_{t_k+h}$ sont les mêmes pour tout k de \mathbb{N}^* et pour tous t_1, \dots, t_k de T^k et h de \mathbb{Z} .
- la stationnarité au sens strict implique la stationnarité au sens large si le processus est du second ordre (cas des séries chronologiques).



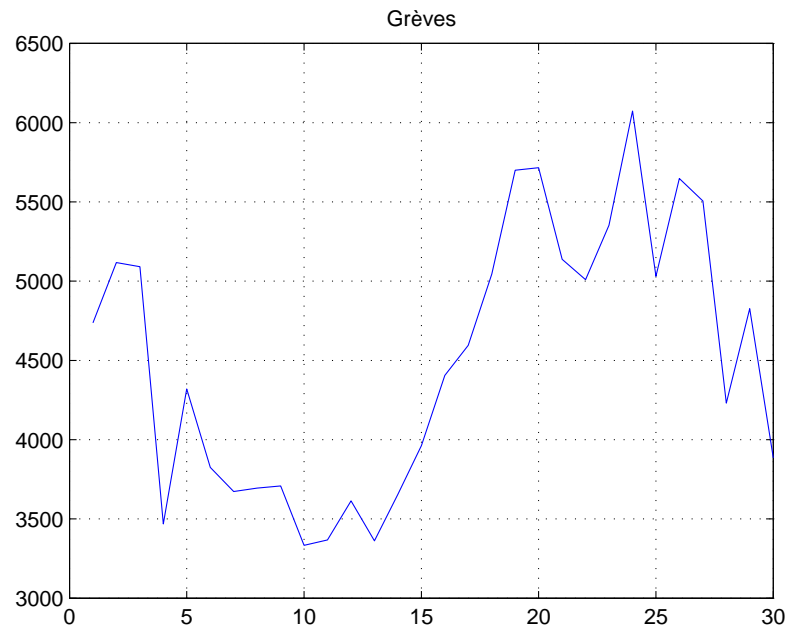


FIG. 1.4 – Nombre annuel de grèves aux Etats-Unis de 1951 à 1980, soit 30 valeurs.

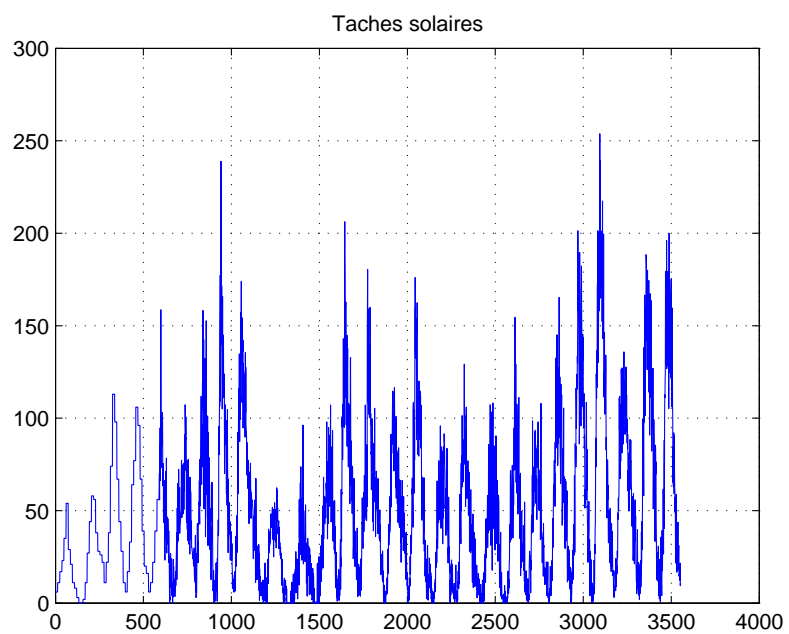


FIG. 1.5 – Nombre mensuel de taches solaires de 1700 à 1995, soit 3552 valeurs.

Pour un processus stationnaire, on peut réécrire la fonction de covariance comme une fonction d'une seule variable :

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

On définit la fonction de corrélation ρ par :

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)},$$

avec

$$\gamma(0) = V_X,$$

où V_X est la variance du processus. Par définition, on a donc $\rho(0) = 1$.

On a par ailleurs, pour les processus à valeurs dans \mathbb{R} :

$$\gamma(-\tau) = \gamma(\tau) \quad \text{et} \quad \rho(-\tau) = \rho(\tau).$$

Pour les processus à valeurs dans \mathbb{C} , on a par définition :

$$\gamma(\tau) = \text{Cov}(X_t, X_{t+\tau}^*),$$

soit :

$$\gamma(-\tau) = \gamma(\tau)^* \quad \text{et} \quad \rho(-\tau) = \rho(\tau)^*.$$

Pour les processus *univariés*, on parlera indifféremment de covariance ou d'autocovariance, et de corrélation ou d'autocorrélation. Les termes *intercovariance* et *intercorrélation* sont réservés aux séries multivariées.

Une série en général peut être décomposée en la somme d'une partie *déterministe* (une fonction suffisamment "régulière" pour être identifiée et extraite) et d'une partie *aléatoire* à moyenne nulle, stationnaire ou non.

La partie *déterministe* est soit une tendance (évolution à long terme), soit une fluctuation saisonnière, soit une composition des deux.

Les notions de *processus réguliers* et de *processus singuliers* (ou *prédictibles*), et le théorème de décomposition de Wold, qui seront abordés dans la section 2.3, permettent d'affiner ces définitions.

La partie *aléatoire* d'une série chronologique peut être le reflet :

- de variations aléatoires du phénomène observé, autour de sa tendance et de ses évolutions saisonnières,
- de variations de l'erreur de mesure si le phénomène est observé par un appareil qui ajoute du bruit à la mesure (cas des récepteurs de radars ou de communications par exemple),
- ou des deux dans le cas le plus général, par exemple les échos de sol observés par un radar ou une vibration aléatoire observée à l'aide d'un laser cohérent ; il est alors important de savoir séparer les deux phénomènes.

Notion de stationnarité pour une observation. Dans la pratique, on observe une « trajectoire » d'une série chronologique, et on aimerait pouvoir la qualifier de stationnaire s'il est vraisemblable qu'elle soit la réalisation d'un processus stationnaire. Ceci suppose que l'on puisse dire quelque chose sur la loi du processus à partir d'une seule réalisation, et fait donc l'hypothèse d'une certaine *ergodicité* du processus, et même d'une *mémoire courte par rapport à la longueur de la série* puisque la réalisation est de longueur finie. On dira donc dans ce cours qu'une *série chronologique semble stationnaire lorsqu'il est vraisemblable qu'elle soit la réalisation d'un processus stationnaire dont la mémoire est courte devant la longueur de la série* (cette définition

sera à nuancer dans les chapitres 7 et 8). Il existe des tests statistiques de stationnarité ou de non-stationnarité (Dickey-Fuller ou KPSS notamment), mais ils sont orientés contre des alternatives trop peu générales pour être intéressants à ce stade. Pour répondre à la question de la stationnarité d'une observation, on se laissera guider par l'intuition et l'idée suivante : si le processus sous-jacent est à mémoire courte devant la longueur de la série, il doit exister trois échelles de temps dans l'intervalle de temps $[1, n]$ d'observation de la série.

- Le court terme qui correspond à la longueur de la mémoire du processus : deux valeurs dont les dates diffèrent d'un court terme sont corrélées,
- le moyen terme à l'échelle duquel la mémoire s'efface : si l'on calcule une moyenne empirique sur le moyen terme, cette moyenne est très proche de la valeur théorique, (environ un dixième de la longueur de la série en pratique)
- le long terme, qui est l'ordre de la longueur de la série, et qui doit être grand devant le moyen terme.

L'idée est alors la suivante : on déplace mentalement une fenêtre de largeur le moyen terme sur toute la longueur de la série. A chaque position de cette fenêtre, on observe l'échantillon qui se trouve à l'intérieur. Si tous ces échantillons semblent similaires au second ordre, au sens où leurs moyennes et covariances empiriques à court terme semblent à peu près les mêmes, alors on en conclura que la série semble stationnaire.

Exercice 1.1 Soit $(Z_t)_{t \in \mathbb{Z}}$ une suite de gaussiennes i.i.d de moyenne nulle et variance 1, et a , b et c des constantes réelles. Les processus suivants sont-ils stationnaires (au second ordre) ? Si oui, donner leur moyenne et leur fonction d'autocovariance.

- | | |
|--|---|
| <p>(a) $X_t = a + bZ_t + cZ_{t-1}$,</p> <p>(c) $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$</p> <p>(e) $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$</p> | <p>(b) $X_t = a + bZ_0$</p> <p>(d) $X_t = Z_0 \cos(ct)$</p> <p>(f) $X_t = Z_t Z_{t-1}$</p> |
|--|---|

Exercice 1.2 Considérer les séries suivantes :

- les premières correspondent aux cinq séries représentées à la section 1.3 ci-dessus ; que peut-on dire sur leur aspect stationnaire / tendance / facteurs saisonniers par une simple analyse visuelle ?
- la dernière correspond au processus de 'marche au hasard' (random walk). Donner une définition de ce processus (cas univarié) et un exemple de phénomène pratique correspondant à ce modèle. Caractériser ce processus d'un point de vue statistique (moyenne, covariance), démontrer que c'est bien une série chronologique et préciser si cette série est :
 - avec tendance
 - stationnaire ou non stationnaire

1.3.4 Représentation temporelle / représentation spectrale

1.3.4.1 Différents types de représentations temporelles

La représentation générale d'une série avec tendance et facteur saisonnier est, dans le cas du *modèle additif*, la suivante :

$$x_t = g_t + s_t + w_t$$

où x désigne la série, g la tendance, s le facteur saisonnier, et w la partie aléatoire.

Le détail des modèles et des types de tendances et de facteurs saisonniers sera vu au chapitre 3.

Les séries stationnaires peuvent avoir des représentations temporelles variées. Ce cours aborde principalement les séries stationnaires suivantes :

- les séries *ARMA* (Auto Regressive Moving Average) dont le modèle est linéaire et dont la fonction de corrélation décroît exponentiellement, cf. chapitre 5,
- les séries linéaires à mémoire longue, dont la fonction de corrélation décroît 'lentement', qui peuvent trouver un modèle *ARMA* d'ordre infini ou plus simplement un modèle *ARMA* intégré avec un ordre *fractionnaire*, cf. chapitre 7,
- les séries non linéaires de type *(G)ARCH* ((Generalized) Auto Regressive Conditionnally Heteroskedastic) largement rencontrées dans le domaine de la finance, cf. chapitre 8.

Les séries non stationnaires et sans composante déterministe peuvent souvent être représentées par des modèles *ARIMA* (Auto Regressive Integrated Moving Average), qui sont encore linéaires, cf. chapitre 6.

Un mode de représentation temporelle largement utilisé est le modèle d'état, support en particulier du filtrage de Kalman. Le modèle d'état permet de représenter et traiter de nombreuses séries chronologiques, cf. chapitre 4.

Les propriétés statistiques d'une série *stationnaire gaussienne* sont entièrement définies par sa moyenne m et sa fonction de covariance ou d'autocovariance, qui vaut dans le cas d'une série à valeurs dans \mathbb{R} :

$$\gamma(\tau) = \mathbb{E}((X_t - m)(X_{t+\tau} - m)) ,$$

et est indépendante de t .

1.3.4.2 Définition des modèles ARMA

Le processus (X_t) , où $t \in \mathbb{Z}$, est un *ARMA*(p, q) de moyenne nulle si :

$$\begin{cases} (X_t) \text{ est stationnaire de moyenne nulle,} \\ X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} , \end{cases}$$

où Z_t est un *bruit blanc faible*, c'est-à-dire ici un processus aléatoire stationnaire de moyenne nulle avec $\gamma_Z(0) = \sigma^2$ et $\gamma_Z(h) = 0$ pour tout $h \neq 0$.

Dans le cadre de ce cours, Z_t sera par défaut un *bruit blanc fort*, c'est-à-dire que la séquence Z_t est de plus *i.i.d.*. Dans le chapitre 8 consacré aux processus *ARMA* / *(G)ARCH*, les propriétés de Z (appelé ϵ) sont spécifiques.

Une représentation équivalente et plus concise est la suivante :

$$\Phi(B)X_t = \Theta(B)Z_t ,$$

où Φ et Θ sont des polynômes de degrés respectifs p et q :

$$\begin{aligned} \Phi(z) &= 1 - \phi_1 z - \dots - \phi_p z^p , \\ \Theta(z) &= 1 + \theta_1 z + \dots + \theta_q z^q , \end{aligned}$$

et B est l'opérateur "retard" défini par :

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

Le processus X_t est un *ARMA*(p, q) de moyenne m si le processus $X_t - m$ est un *ARMA*(p, q) de moyenne nulle.

Propriété importante. Pour tout processus aléatoire stationnaire dont la fonction de corrélation $\gamma(\cdot)$ tend vers 0 à l'infini, et pour tout entier $k > 0$, il est possible de trouver un processus *ARMA* dont la fonction de corrélation est égale à celle du processus aléatoire jusqu'à l'ordre k . Cette propriété est l'une des raisons de l'intérêt pratique des modèles *ARMA* ; une autre raison est leur facilité de synthèse.

Symétrie des processus ARMA. Lorsque l'on observe un processus ARMA en “retournant” le temps, on obtient un nouveau processus qui a les mêmes propriétés statistiques ; la représentation ARMA ne permet donc pas de modéliser un processus dont les décroissances sont en moyenne plus rapides que les croissances (cas des indices boursiers par exemple).

On dit que X_t est un processus MA d'ordre q si $\Phi(z) \equiv 1$, soit :

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1.1)$$

On dit que X_t est un processus AR d'ordre p si $\Theta(z) \equiv 1$, soit :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \quad (1.2)$$

Exercice 1.3 Démontrer qu'un processus défini par $X_t = \Theta(B)Z_t$ pour $t \in \mathbb{Z}$ et où Z est un bruit blanc, est toujours stationnaire (quelque soit Θ).
Quelle remarque peut-on faire si le processus est défini pour $t \in \mathbb{N}$?

Exercice 1.4 Considérer le processus défini par l'équation (1.2) pour $p = 1$:

$$X_t - \phi_1 X_{t-1} = Z_t, \quad \forall t \in \mathbb{Z}$$

Ce processus est-il stationnaire ? (calculer la moyenne et la variance des échantillons successifs)
A quel processus correspond le cas $\phi_1 = 1$?
Calculer la fonction de covariance du processus dans le cas stationnaire.

1.3.4.3 Définition des modèles d'état

Dans ce paragraphe, les lettres en gras représentent des vecteurs colonne.

Un modèle d'état *linéaire* est défini par :

- une équation d'évolution de l'état :

$$\mathbf{X}_{t+1} = F_t \mathbf{X}_t + \mathbf{V}_t \quad (1.3)$$

- une équation de mesure :

$$\mathbf{Y}_t = G_t \mathbf{X}_t + \mathbf{W}_t \quad (1.4)$$

Dans ces équations :

\mathbf{Y}_t est un vecteur de dimension w

\mathbf{X}_t est un vecteur de dimension v

G_t est une séquence temporelle de matrices (w, v)

F_t est une séquence temporelle de matrices (v, v)

\mathbf{W}_t est une séquence temporelle de vecteurs aléatoires indépendants de dimension w (bruit de mesure), de moyenne nulle et de matrice de covariance $R_t(w, w)$

\mathbf{V}_t est une séquence temporelle de vecteurs aléatoires indépendants de dimension v (bruit d'état), de moyenne nulle et de matrice de covariance $Q_t(v, v)$

Les séquences \mathbf{W}_t et \mathbf{V}_t sont indépendantes entre elles, et *décorrélées* avec X_1 .

L'équation (1.3) traduit l'évolution au cours du temps du vecteur \mathbf{X} , et l'équation (1.4) traduit la mesure (ou observation) de \mathbf{X} . Dans de nombreux cas pratiques, les matrices F_t , G_t , Q_t et R_t sont indépendantes du temps, et sont alors notées respectivement F , G , Q et R .

Tout processus ARMA causal peut être représenté par un modèle d'état (la définition de la causalité sera donnée dans un chapitre ultérieur) ; avec les notations précédentes (équations (1.3) et (1.4)), \mathbf{Y} est la série chronologique ($w = 1$ pour une série univariée). Cette représentation n'est pas unique en général. Les modèles d'état permettent de représenter des classes plus larges de processus, et en particulier de séries chronologiques.

Exercice 1.5 Donner une représentation par modèle d'état de quelques séries chronologiques simples :

1. une série chronologique égale à la somme d'une tendance linéaire et d'un bruit blanc (\mathbf{X} est de dimension 2, \mathbf{Y} est la série chronologique)
2. un processus $AR(1)$,
3. une série égale à la somme d'une tendance linéaire et d'un $AR(1)$.

On respectera la contrainte que les suites (\mathbf{V}_t) et (\mathbf{W}_t) doivent être de moyenne nulle, indépendantes entre elles et décorrélés de \mathbf{X}_1 .

Exercice 1.6 Soit Y un processus $ARMA(1, 1)$ défini par l'équation suivante :

$$Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1} .$$

Trouver un modèle d'état dont la série des observations est conforme à ce modèle.

1.3.4.4 Représentation spectrale d'un processus aléatoire stationnaire

La densité spectrale de puissance d'un processus aléatoire stationnaire discret est la transformée de Fourier (fonction continue) de sa fonction de covariance :

$$P_X(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) e^{-jk\omega} \quad \omega \in [-\pi, \pi] \quad (1.5)$$

On peut retrouver γ_X à partir de P_X par la relation :

$$\gamma_X(k) = \int_{-\pi}^{\pi} P_X(\omega) e^{jk\omega} d\omega \quad (1.6)$$

Exercice 1.7 Soit X un processus stationnaire. Soient γ_X sa fonction d'autocovariance et P_X sa densité spectrale de puissance. On suppose que γ_X est absolument sommable :

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

1. Montrer que P_X est à valeurs réelles. Quelle propriété supplémentaire présente P_X dans le cas d'un processus à valeurs dans \mathbb{R} ?
2. Prouver la relation (1.6). Quelle relation existe-t-il entre la variance d'un processus aléatoire stationnaire et sa densité spectrale ?
3. Calculer la densité spectrale d'un bruit blanc.

1.4 Pourquoi étudier une série chronologique ?

Ce paragraphe décrit :

- dans un premier temps, quelques opérations de base que l'on peut réaliser sur une série chronologique,

- dans un deuxième temps, des exemples d'utilisation des modèles de séries chronologiques.

1.4.1 Quelques opérations de base

1.4.1.1 Analyse / estimation

Cette opération consiste à *caractériser* une série chronologique :

- déterminer s'il existe des tendances et des facteurs saisonniers et si oui les caractériser de façon quantitative (tendance linéaire, polynomiale, période, etc),
- pour une série stationnaire, estimer sa fonction de covariance ou son spectre, l'identifier à un processus *ARMA*, etc,
- pour une série non stationnaire, chercher des modèles de type *ARIMA*, ou d'autres modèles représentatifs.

L'opération d'analyse / estimation est :

- soit une fin en soi, pour interpréter un phénomène,
- soit une étape nécessaire pour la mise en oeuvre d'autres opérations, dont certaines sont données comme exemples dans la suite de ce paragraphe.

La Figure 1.6 montre un exemple, qui sera détaillé par la suite, où la tendance est bien modélisée par une courbe du deuxième degré.

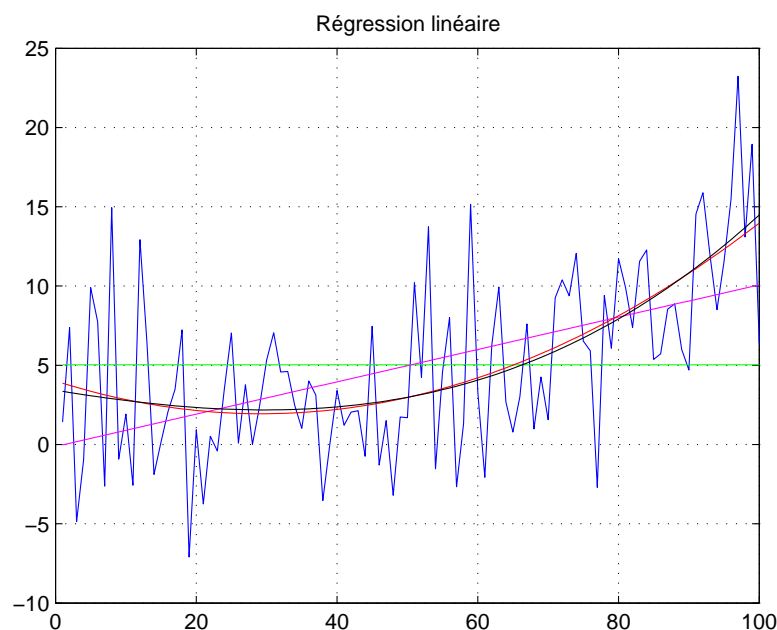


FIG. 1.6 – Régression linéaire sur une parabole

1.4.1.2 Filtrage / lissage / prédiction

C'est l'ensemble des opérations qui permettent de donner la meilleure valeur estimée d'un processus à un ou plusieurs instants différents ; généralement, les termes suivants sont utilisés :

- *filtrage* pour la valeur à l'instant courant compte tenu de l'ensemble des valeurs observées précédemment et à l'instant courant,

- *lissage* pour l'ensemble des valeurs comprises entre deux instants, compte tenu de l'ensemble des valeurs observées,
- *prédiction* pour une ou plusieurs valeurs futures, compte tenu des valeurs observées précédemment.

La Figure 1.6 suggère qu'il est possible de mettre en oeuvre une prédiction pour le processus considéré. En l'absence d'information particulière sur le bruit, cette prédiction consiste à prolonger la courbe du deuxième degré. Si des informations complémentaires sur le bruit étaient acquises (au moyen par exemple de modèles ARMA pour un bruit corrélé), la prédiction pourrait être plus précise, c'est-à-dire présenter une erreur minimale.

1.4.1.3 Synthèse / simulation

La synthèse ou simulation est une opération qui consiste à générer, de façon en général automatique, des séries chronologiques conformes à un modèle. La synthèse a généralement un but d'étude, en permettant de jouer à volonté sur divers types de modèles ainsi que sur leurs paramètres.

Les processus à tendances et facteurs saisonniers et / ou suivant un modèle *ARMA* ou *ARIMA* sont faciles à synthétiser, car ils sont décrits par des fonctions temporelles simples et / ou des relations de récurrence sur leurs valeurs successives. Les processus stationnaires définis par leur densité spectrale de puissance sont généralement plus difficiles à synthétiser.

1.4.2 Quelques exemples d'utilisation de modèles de séries chronologiques

1.4.2.1 Partitionnement des séries chronologiques

Soit par exemple un enregistrement d'électro-encéphalogramme (eeg). Selon l'état du patient (éveil, sommeil lent léger, sommeil lent profond, sommeil paradoxal, état pathologique), les caractéristiques de l'eeg varient.

La Figure 1.7 nous montre un extrait d'un enregistrement d'eeg, présentant une transition entre deux états au cours d'une crise d'épilepsie. Le "scoring", ou partitionnement, de l'eeg peut être fait manuellement par un opérateur spécialiste, mais cette opération est très fastidieuse pour des enregistrements longs (par exemple un partitionnement de huit heures d'eeg par tranches de cinq secondes) et peut conduire à des erreurs. On cherchera donc à mettre en place un scoring automatique basé sur le fait que le modèle de l'eeg (par exemple un processus autorégressif) change de paramètres lorsque le patient change d'état, voir par exemple [4]. La "bibliothèque" de jeux de paramètres à utiliser pour réaliser cette opération peut elle-même être issue d'une phase d'apprentissage statistique telle que décrite ci-dessous.

1.4.2.2 Apprentissage statistique

Dans l'exemple précédent, le scoring automatique peut être précédé d'une phase d'apprentissage. Il s'agira ici de prendre des enregistrements d'eeg pendant des phases où l'état du patient est connu, et d'élaborer un modèle du processus pour chacune de ces phases, par exemple un modèle moyen ou un modèle avec des intervalles possibles pour les paramètres, etc.

1.4.2.3 Détection d'un signal dans un bruit de densité spectrale inconnue

Prenons le cas d'un radar de type "Doppler" : le radar émet un signal dans les longueurs d'onde dites "radiofréquence", et le signal réfléchi sur tous les obstacles rencontrés revient vers le radar après avoir subi des transformations (dites "modulations") qui dépendent des mouvements de ces obstacles par rapport au radar (qui peut lui-même être fixe ou mobile).

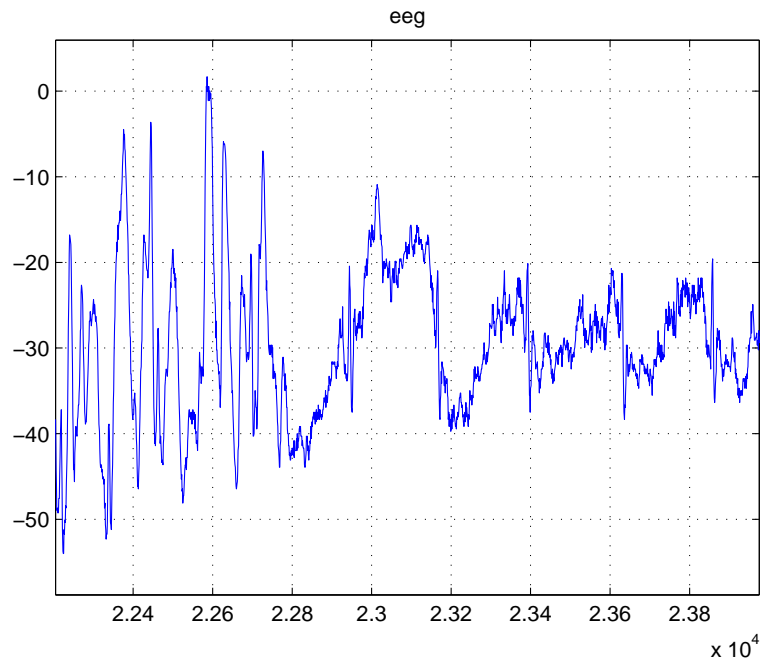


FIG. 1.7 – Transition dans un EEG

Dans le cas d'un radar aéroporté, les “échos de sol” peuvent dans certaines conditions se superposer au signal réfléchi par un objet que l'on souhaite détecter (appelé cible), ces échos de sol ayant des caractéristiques statistiques inconnues.

Dans ces conditions, élaborer un modèle statistique de ces échos de sol permet d'optimiser le processus de détection de la cible. Ce modèle sera obtenu en appliquant les techniques de modélisation des séries chronologiques.

1.4.2.4 Optimisation de fonctions complexes

Une centrale de navigation est un dispositif qui fournit la position et l'altitude, dans un repère inertiel, de l'objet sur lequel elle est montée (par exemple un avion ou un navire). Elle est composée très souvent de capteurs inertiels élémentaires : des gyromètres qui fournissent des dérivées angulaires, et des accéléromètres ; elle porte alors le nom de centrale inertielle. Elle peut également comporter un capteur de position appelé génériquement GNSS (Global Navigation Satellite System) dont l'exemple le plus connu est le GPS (Global Positioning System).

Pour réaliser la fonction de navigation, les informations élémentaires disponibles sont fusionnées au moyen d'un filtrage, le plus classiquement un filtrage de Kalman comportant plusieurs dizaines d'états. Il est pour cela nécessaire de disposer de modèles d'erreurs des capteurs élémentaires : bruit (blanc ou non), biais fixe ou dérive lente, marche au hasard. Ces modèles sont obtenus en appliquant les techniques de modélisation des séries chronologiques.

1.5 Méthodologie générale d'analyse

La méthode d'analyse d'une série chronologique se place dans le cadre général suivant : la série (généralement non stationnaire) est constituée de la somme d'une fonction 'moyenne' $m(t)$, encore appelée tendance, et d'un processus aléatoire à moyenne nulle (processus ARMA, processus ARMA intégré, ...), appelé résidu.

- la méthode consiste en général à déterminer successivement la fonction moyenne puis les caractéristiques du résidu,
- mais les deux opérations peuvent dans certains cas être imbriquées.

1.6 Exercice complémentaire

Exercice 1.8 Soit X_0 une variable aléatoire de carré intégrable. Soit (U_t) un bruit blanc fort d'espérance m et de variance σ^2 tel que U_t soit indépendant de X_0 pour tout $t \geq 1$. On définit le processus (X_t) par la relation suivante :

$$X_t = aX_{t-1} + U_t, \quad t \geq 1$$

avec $|a| < 1$.

1. Montrer qu'une condition nécessaire et suffisante pour que (X_t) soit stationnaire au sens faible est que :

$$\mathbb{E}(X_0) = \frac{m}{1-a},$$

et que :

$$\text{Var}(X_0) = \frac{\sigma^2}{1-a^2}.$$

2. Déterminer la fonction d'autocovariance $k \mapsto \gamma_X(k)$.

Chapitre 2

Représentation spectrale

L'analyse spectrale est une des composantes fondamentales de l'analyse des séries chronologiques. Ce chapitre décrit les notions de densité spectrale et de distribution spectrale des processus aléatoires *stationnaires*, ainsi que des estimateurs de la densité spectrale, en particulier le périodogramme et ses dérivés.

Bien que la densité spectrale soit définie pour des *processus aléatoires stationnaires*, ses estimateurs sont utilisables pour l'analyse des séries chronologiques en général.

2.1 Densité et distribution spectrales

2.1.1 Définition et propriétés de la fonction de covariance

Définition 2.1 *La fonction de covariance (ou d'autocovariance) d'un processus aléatoire complexe est donnée par :*

$$\text{Cov}(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}([X_t - \mathbb{E}(X_t)][X_s - \mathbb{E}(X_s)]^*) ,$$

où $'^*$ désigne le complexe conjugué. Pour un processus stationnaire, on peut réécrire la fonction de covariance comme une fonction d'une seule variable :

$$\gamma(\tau) = \text{Cov}(X_t, X_{t+\tau}) .$$

On définit alors la fonction de corrélation ρ par :

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} ,$$

où $\gamma(0) = V_X$ est la variance du processus.

Remarquons que par définition, on a $\rho(0) = 1$. On a par ailleurs, pour les processus à valeurs dans \mathbb{C} :

$$\gamma(-\tau) = \gamma(\tau)^* \quad \text{et} \quad \rho(-\tau) = \rho(\tau)^* .$$

Un processus aléatoire stationnaire dont la fonction de covariance est identiquement nulle sauf pour $\tau = 0$ est qualifié de “*bruit blanc faible*”.

On définit également la fonction de covariance mutuelle entre deux processus aléatoires stationnaires X et Y (pour deux processus réels) :

$$\gamma_{[X,Y]}(t_1, t_2) = \gamma_{[X,Y]}(t_2 - t_1) = \mathbb{E}([X_{t_1} - \mathbb{E}(X)][Y_{t_2} - \mathbb{E}(Y)])$$

Définition 2.2 Les deux processus aléatoires stationnaires X et Y sont décorrélés si leur fonction de covariance mutuelle est identiquement nulle :

$$\gamma_{[X,Y]}(t_1, t_2) = 0 \quad \text{pour tous } t_1 \text{ et } t_2$$

Proposition 2.3 La fonction de covariance d'un processus stationnaire égal à la somme de deux processus stationnaires décorrélés est égale à la somme des fonctions de covariance de chacun des deux processus :

$$X_t \text{ et } Y_t \text{ décorrélés} \Rightarrow \gamma_{X+Y}(\tau) = \gamma_X(\tau) + \gamma_Y(\tau).$$

Notons que la propriété précédente ne s'applique pas à la fonction de corrélation.

2.1.2 Densité spectrale d'un processus aléatoire stationnaire

Théorème 2.4 Soit une fonction K définie sur \mathbb{Z} à valeurs complexes, absolument intégrable :

$$\sum_{n=-\infty}^{\infty} |K(n)| < \infty.$$

Alors on peut écrire :

$$K(k) = \int_{-\pi}^{\pi} f(\omega) e^{jk\omega} d\omega \quad \text{avec } j^2 = -1 \quad (2.1)$$

avec :

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} K(k) e^{-jk\omega} \quad \omega \in [-\pi, \pi] \quad (2.2)$$

Exercice 2.1 Démontrer l'équivalence des relations 2.1 et 2.2 énoncées au Théorème 2.4.

La fonction f ainsi définie est la transformée de Fourier de K .

Nota : la fonction f est continue, la fonction K est discrète.

Définition 2.5 Soit X un processus aléatoire stationnaire discret tel que :

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

La densité spectrale de puissance P_X de X est la transformée de Fourier (fonction continue) de sa fonction de covariance γ_X :

$$P_X(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-jk\omega} \quad \omega \in [-\pi, \pi] \quad (2.3)$$

$$\gamma_X(k) = \int_{-\pi}^{\pi} P_X(\omega) e^{jk\omega} d\omega \quad (2.4)$$

Proposition 2.6 La densité spectrale de puissance d'un processus aléatoire stationnaire est positive ou nulle pour tout ω .

Proposition 2.7 (équivalente à la précédente) Une fonction $K(\cdot)$ définie sur \mathbb{Z} est la fonction de covariance d'une série temporelle (à valeurs dans \mathbb{C} en général) si et seulement si $K(\cdot)$ est hermitienne et définie non-négative, soit :

$$K(n) = K(-n)^*$$

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

pour tout n et tout vecteur $(a_1, \dots, a_n)'$ de \mathbb{C}^n .

Exercice 2.2 La fonction K suivante est-elle une fonction d'autocovariance ?

$$K(0) = 1, K(1) = K(-1) = r, K(h) = 0 \text{ sinon.}$$

2.1.3 Distribution spectrale d'un processus aléatoire stationnaire

La distribution spectrale d'un processus aléatoire stationnaire est la fonction $D(\omega)$ définie sur l'intervalle $[-\pi, \pi]$, telle que :

$$D(\omega) = \int_{-\pi}^{\omega} P(u) du$$

(la densité spectrale est la dérivée de la distribution spectrale).

2.1.4 Filtre linéaire appliqué à un processus aléatoire : transformation de la densité spectrale

Soit Y_t un processus aléatoire stationnaire à valeurs dans \mathbb{C} et de moyenne nulle avec la densité spectrale $P_Y(\omega)$.

Soit X_t le processus défini par :

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i Y_{t-i} \quad (2.5)$$

avec :

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty. \quad (2.6)$$

Alors X_t est stationnaire avec une densité spectrale donnée par :

$$P_X(\omega) = |\Psi(e^{-j\omega})|^2 P_Y(\omega) \quad (2.7)$$

avec :

$$\Psi(z) = \sum_{i=-\infty}^{\infty} \psi_i z^i \quad (2.8)$$

Exercice 2.3 Démontrer les relations exprimées dans les équations 2.7 et 2.8.

2.2 Densité spectrale d'un processus ARMA

Exercice 2.4 Calculer la densité spectrale et la distribution spectrale d'un processus $MA(1)$.

De façon plus générale, la densité spectrale d'un processus ARMA est égale au rapport de deux polynômes trigonométriques, on la qualifie de *densité spectrale rationnelle*. Ce point sera étudié dans la section 5.4.

2.3 Décomposition de Wold

2.3.1 Distribution spectrale

Soit X le processus défini par :

$$X_t = a_1 e^{j\omega_1 t} + \dots + a_p e^{j\omega_p t},$$

où les $a_i, i = 1, \dots, p$ sont des variables aléatoires complexes indépendantes centrées et gaussiennes telles que :

$$\mathbb{E}(a_i) = 0 \quad \text{et} \quad \mathbb{E}(a_i a_i^*) = \sigma_i^2,$$

et les ω_i sont des nombres réels donnés. Alors X_t est stationnaire avec :

$$\begin{aligned} \mathbb{E}(X_t) &= 0 \quad \text{pour tout } t, \\ \gamma_X(h) &= \mathbb{E}(X_{t+h} X_t^*) = \sum_{i=1}^p \sigma_i^2 e^{jh\omega_i} \quad (\text{indépendant de } t) \end{aligned}$$

La distribution spectrale de X est donnée par la formule suivante :

$$D_X(\omega) = \sum_{\{i\}/\omega_i \leq \omega} \sigma_i^2.$$

Cette distribution spectrale présente donc des discontinuités (positives) aux points d'abscisses ω_i .

2.3.2 Densité spectrale

La densité spectrale étant la dérivée de la distribution spectrale, elle se présente sous la forme de fonctions de Dirac d'énergie σ_i^2 aux points d'abscisses ω_i :

$$P_X(\omega) = \sum_i \sigma_i^2 \delta(\omega - \omega_i)$$

2.3.3 Décomposition de Wold

En termes de densité spectrale, la décomposition de Wold s'énonce de la façon suivante :

La densité spectrale de tout processus aléatoire stationnaire est composée de la somme d'une fonction continue positive et d'un nombre fini ou dénombrable d'impulsions de Dirac.

Introduisons les notions de processus singulier et réguliers. Un processus stationnaire X_t est *singulier* (ou *prédictible*) s'il existe un ensemble (fini ou dénombrable) de coefficients c_k tels que :

$$X_t = \sum_k c_k X_{t-k}.$$

Le processus décrit dans le paragraphe 2.3.1 :

$$X_t = a_1 e^{j\omega_1 t} + \dots + a_p e^{j\omega_p t}$$

est prédictible. Un processus qui n'est pas singulier est dit *régulier*. La décomposition de Wold s'énonce également de la façon équivalente suivante :

Proposition 2.8 (Décomposition de Wold) *Tout processus aléatoire stationnaire est la somme de deux processus :*

$$X_t = X_t^s + X_t^r,$$

où X_t^s est un processus singulier (ou prédictible) et X_t^r est un processus régulier (ou non prédictible), et ces deux processus sont orthogonaux :

$$\forall t_1, t_2 \in \mathbb{Z}, \quad \mathbb{E}(X_{t_1}^s X_{t_2}^r) = 0.$$

Dans cette décomposition, le processus régulier X_t^r correspond à la partie continue de la densité spectrale, et le processus singulier X_t^s correspond aux impulsions de Dirac de la densité spectrale.

2.4 Estimation spectrale non paramétrique

Les méthodes d'estimation non paramétrique ne font pas d'hypothèses a priori sur le signal à analyser, elles sont donc souvent utilisées pour une analyse préliminaire (se reporter au chapitre 'tendances' où l'on voit qu'un simple périodogramme peut mettre en évidence la présence de sinusoïdes cachées).

2.4.1 Le périodogramme : propriétés, limitations

2.4.1.1 Définition du périodogramme

Soit n un nombre entier et l'ensemble des vecteurs $\{\mathbf{e}_i :$

$$\mathbf{e}_i = n^{-\frac{1}{2}}(e^{j\omega_i}, e^{j2\omega_i}, \dots, e^{jn\omega_i}),$$

avec $\omega_i = \frac{2\pi i}{n}$, pour tout $i \in F_n = \{-E(\frac{n-1}{2}), \dots, E(\frac{n}{2})\}$.

Exercice 2.5 Montrer que la famille de vecteurs $\{\mathbf{e}_i, i \in F_n\}$ définie au paragraphe 2.4.1 constitue une base orthonormale de \mathbb{C}^n .

Il résulte de l'Exercice 2.5 que tout vecteur \mathbf{x} de \mathbb{C}^n peut s'écrire sous la forme :

$$\mathbf{x} = \sum_i a_i \mathbf{e}_i,$$

avec :

$$a_i = \langle \mathbf{x}, \mathbf{e}_i \rangle = n^{-\frac{1}{2}} \sum_{t=1}^n x_t e^{-j\omega_i t}.$$

La séquence des a_i , pour $i \in F_n$, est appelée Transformée de Fourier Discrète (TFD) du vecteur \mathbf{x} de \mathbb{C}^n .

Définition 2.9 (périodogramme) *Le périodogramme de la séquence \mathbf{x} de \mathbb{C}^n est la séquence des carrés des modules des a_i :*

$$I_n(\omega_i) = \|a_i\|^2 = |\langle \mathbf{x}, \mathbf{e}_i \rangle|^2 = n^{-1} \left| \sum_{t=1}^n x_t e^{-j\omega_i t} \right|^2$$

Notons tout de suite deux propriétés importantes du périodogramme :

Proposition 2.10 *Soit I_n le périodogramme de $\mathbf{x} \in \mathbb{C}^n$:*

$$\|\mathbf{x}\|^2 = \sum_{i \in F_n} I_n(\omega_i)$$

Proposition 2.11 *Soit une série temporelle à valeurs complexes X_t , $t = 1, \dots, n$ et I_n son périodogramme. Alors on a la relation suivante pour toute valeur de ω :*

$$I_n(\omega) = n^{-1} \left| \sum_{t=1}^n X_t e^{-j\omega t} \right|^2 = \sum_{|k| < n} \gamma_{nc}(k) e^{-jk\omega}, \quad (2.9)$$

où $\gamma_{nc}(\cdot)$ est la fonction de covariance empirique non centrée de la série temporelle, définie pour $0 \leq h < n$ par :

$$\gamma_{nc}(h) = n^{-1} \sum_{j=1}^{n-h} X_{j+h} X_j^* \quad (2.10)$$

Exercice 2.6 Démontrer la relation décrite par les équations (2.9) et (2.10).

Cette relation suggère de prendre, comme estimateur *non paramétrique* de la densité spectrale d'un processus aléatoire stationnaire, pour lequel on dispose d'une réalisation de taille n , le périodogramme divisé par 2π (voir la Définition 2.5), soit :

$$\hat{P}_X(\omega) = \frac{1}{2\pi} I_n(\omega)$$

D'après la proposition 2.11, à partir d'une série X_t , le périodogramme peut être calculé de deux façons équivalentes, selon le schéma de la Figure 2.1, où l'opération TF est donnée par la formule (2.9) calculée pour les valeurs ω_i de ω .

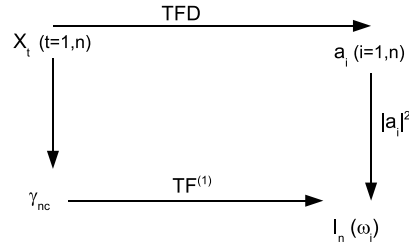


FIG. 2.1 – Illustration de la Proposition 2.11

2.4.1.2 Echelle des abscisses pour le périodogramme

Dans la formule qui donne le périodogramme :

$$I_n(\omega_i) = n^{-1} \left| \sum_{t=1}^n x_t e^{-j\omega_i t} \right|^2,$$

la période d'échantillonnage est égale à 1, et le pas en pulsation ω est égal à $\frac{2\pi}{n}$, soit un pas en fréquence de $\frac{1}{n}$. Dans les cas pratiques, la période d'échantillonnage t_e est exprimée en secondes, et $f_e = \frac{1}{t_e}$ en Hertz. On aura par exemple dans le cas d'un électro-encéphalogramme une fréquence d'échantillonnage égale à 100Hz , soit $t_e = 10^{-2}\text{s} = 10\text{ms}$. Dans ces conditions, le pas en fréquence du périodogramme vaut $\delta f = \frac{1}{nt_e} = \frac{1}{T}$ où $T = nt_e$ est la durée totale de la série, et le domaine de fréquence analysé est égal à $f_e = \frac{1}{t_e}$.

2.4.1.3 Propriétés asymptotiques du périodogramme

Théorème 2.12 Soit X un processus réel de la forme :

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}$$

où Z_t est un processus i.i.d. $(0, \sigma^2)$, avec :

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$$

Soit $I_n(\omega_i)$ le périodogramme de (X_1, \dots, X_n) et P la densité spectrale de X . On a les deux résultats suivants.

(i) Supposons que :

$$\forall \omega \in [-\pi, \pi], P(\omega) > 0.$$

Alors pour tous $0 < \omega_1 < \dots < \omega_m < \pi$, le vecteur aléatoire $(I_n(\omega_1), \dots, I_n(\omega_m))'$ converge en loi vers un vecteur de variables aléatoires indépendantes et distribuées exponentiellement, la moyenne (et donc aussi l'écart-type) de $I_n(\omega_i)$ étant égale à $2\pi P(\omega_i)$.

(ii) Si :

$$\sum_{i=-\infty}^{\infty} |\psi_i| |i|^{\frac{1}{2}} < \infty \text{ et } \mathbb{E}(Z^4) = \eta \sigma^4 < \infty,$$

pour $\omega_i = \frac{2i\pi}{n} \geq 0$ et $\omega_k = \frac{2k\pi}{n} \geq 0$, on a :

$$\text{Cov}(I_n(\omega_i), I_n(\omega_k)) = \begin{cases} 2(2\pi)^2 P^2(\omega_i) + O(n^{-1/2}) & \text{pour } \omega_i = \omega_k = 0 \text{ ou } \pi \\ (2\pi)^2 P^2(\omega_i) + O(n^{-1/2}) & \text{pour } 0 < \omega_i = \omega_k < \pi \\ O(n^{-1}) & \text{pour } \omega_i \neq \omega_k \end{cases} \quad (2.11)$$

Exercice 2.7 Interpréter les résultats donnés par l'équation (2.11).

Ces propriétés sont à reconsidérer dans le cas de processus complexes dont la partie réelle et la partie imaginaire ne sont pas indépendants.

Exercice 2.8 Réponse en fréquence du périodogramme

Calculer la réponse en fréquence du périodogramme, c'est-à-dire la valeur du périodogramme lorsque la série chronologique en entrée est une exponentielle complexe $X_t = e^{jt\omega}$ de pulsation ω .

Quelles conclusions peut-on tirer de ce résultat ?

L'exercice 2.9 a pour objectif de montrer quantitativement comment le périodogramme permet de déceler une sinusoïde de faible amplitude dans du bruit.

Exercice 2.9 Rapport signal sur bruit

Soit la série complexe $X_t = A e^{j\omega_0 t} + w_t$ (une exponentielle complexe plus un bruit blanc complexe dont les composantes sont indépendantes entre elles et ont chacune une variance σ_w^2). On définit le rapport signal sur bruit par $RSB = \frac{A^2}{2\sigma_w^2}$ (rapport entre l'énergie de l'exponentielle et celle du bruit).

Soit $I_n(\omega_i)$ le périodogramme de (X_1, \dots, X_n) et supposons que ω_0 soit égal à l'une des valeurs $\omega_i : \exists i / \omega_0 = \omega_i = \frac{2\pi i}{n}$.

Donner une définition du rapport signal sur bruit en sortie du périodogramme, soit RSB_s , et calculer le rapport $\frac{RSB_s}{RSB}$.

2.4.2 Estimateurs dérivés du périodogramme

Soit le périodogramme défini ci-dessus par la relation (2.9) :

$$I_n(\omega) = n^{-1} \left| \sum_{t=1}^n X_t e^{-j\omega t} \right|^2 = \sum_{|k| < n} \gamma_{nc}(k) e^{-jk\omega}.$$

Ce périodogramme, divisé par 2π , constitue un estimateur de la densité spectrale du processus X qui est sans biais, mais non consistant. On peut donc réaliser un “post-traitement” pour améliorer cet estimateur ; ce post-traitement s’effectue soit dans le domaine spectral, soit dans le domaine temporel.

2.4.2.1 Périodogramme lissé dans le domaine spectral

La propriété d’indépendance des erreurs d’estimation pour des valeurs voisines de ω suggère de lisser le périodogramme $I_n(\omega_i)$ obtenu pour des valeurs de ω qui sont des multiples de $\frac{2\pi}{n}$, par la formule :

$$\hat{P}(\omega_i) = \frac{1}{2\pi} \sum_{|k| \leq m(n)} W_n(k) I_n(\omega_{i+k})$$

Pour que cet estimateur soit consistant, il suffit que les conditions suivantes soient respectées :

$$\begin{cases} m \rightarrow \infty \text{ et } \frac{m}{n} \rightarrow 0 \text{ quand } n \rightarrow \infty \\ W_n(k) = W_n(-k) \text{ et } W_n(k) \geq 0 \text{ pour tout } k \\ \sum_{|k| \leq m(n)} W_n(k) = 1 \\ \sum_{|k| \leq m(n)} W_n^2(k) \rightarrow 0 \text{ quand } n \rightarrow \infty \end{cases}$$

Le choix de la fonction de pondération W résulte d’un compromis entre biais et variance de l’estimateur.

2.4.2.2 Pondération dans le domaine temporel

Une autre façon de lisser le périodogramme consiste à moyenner plusieurs périodogrammes construits à partir d’échantillons temporels successifs de la série à analyser. Lorsque l’on dispose d’un échantillon de taille donnée, ce qui est le cas dans la pratique, il faut donc diviser cet échantillon en plusieurs échantillons de tailles égales (de préférence), ce qui diminue la résolution fréquentielle.

La Figure 2.2 illustre le procédé, dans le cas d’un processus AR(1) simulé. La taille totale de l’échantillon est égale à 10000, et il a été décomposé en 100 échantillons de taille 100 pour le lissage du périodogramme. Il existe des équivalences entre les traitements dans le domaine spectral et dans le domaine temporel.

On peut également modifier la réponse fréquentielle du périodogramme, et en particulier minimiser les effets de “lobes secondaires”, en appliquant une pondération temporelle sur le signal.

2.5 Estimation spectrale paramétrique

Les méthodes paramétriques sont généralement plus efficaces lorsque l’on est capable de modéliser le processus grâce à un petit nombre de paramètres ; deux exemples sont donnés ci-dessous.

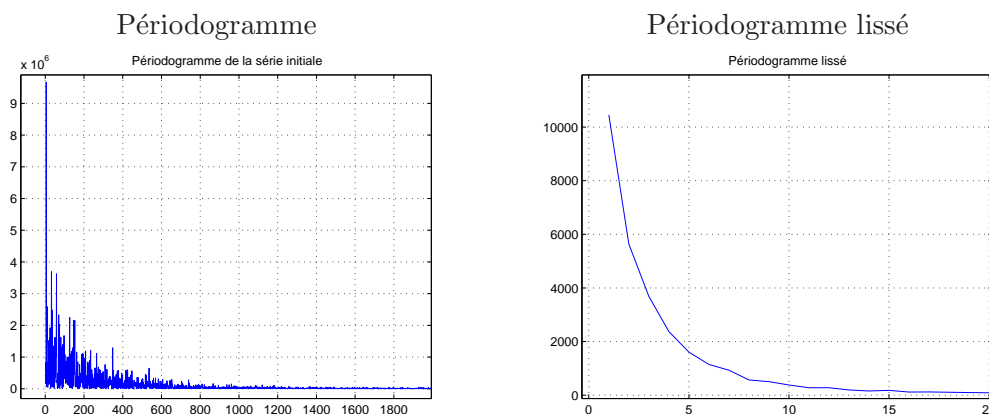


FIG. 2.2 – Lissage temporel du périodogramme

2.5.1 Identification à un processus ARMA

Les processus ARMA sont un exemple de processus qui se modélisent entièrement grâce à un petit nombre de paramètres. La méthode à utiliser est alors la suivante :

- estimation du processus par les méthodes temporelles étudiées plus loin dans le cours (détermination des ordres p et q , calcul des valeurs des ϕ et θ et de σ^2),
- calcul direct du spectre rationnel (cf. section 5.4).

Cette technique présente l'avantage de conduire à un spectre estimé qui est bien lissé, mais cependant affecté des erreurs de l'estimation temporelle.

2.5.2 Multiple Signal Classification Method (MUSIC)

Cette technique n'est pas décrite ici en détail. Elle modélise le signal comme étant la somme d'un nombre fini d'exponentielles complexes d'amplitudes aléatoires, et d'un bruit blanc. Elle opère à partir de la décomposition de l'espace d'observation en deux sous-espaces orthogonaux : l'espace signal et l'espace bruit.

Cette décomposition s'obtient à partir du classement par ordre décroissant des valeurs propres de la matrice de corrélation du signal ; si le nombre d'exponentielles complexes est égal à p , les p vecteurs propres correspondant aux p plus grandes valeurs propres engendrent l'espace signal.

Cette technique permet en particulier de détecter des sinusôides dont l'écart de fréquence est petit (c'est-à-dire non discernables par le périodogramme).

2.6 Analyse des séries chronologiques

Comme indiqué en introduction de ce chapitre, le périodogramme peut servir, non seulement à estimer la densité spectrale d'un *processus aléatoire stationnaire*, mais également pour mettre en évidence des composantes périodiques dans une série chronologique en général. La Figure 2.3 illustre cette possibilité, dans l'exemple des taches solaires (la figure de droite présente un zoom autour de la fréquence nulle du périodogramme en échelle logarithmique).

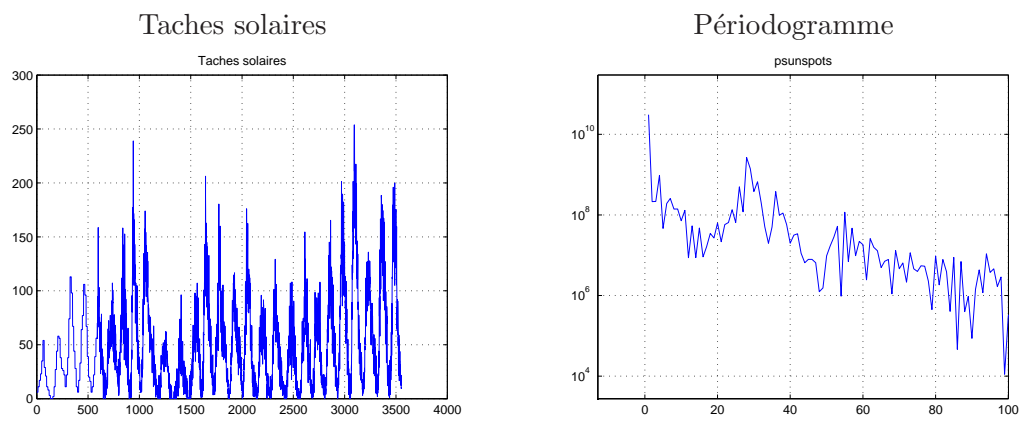


FIG. 2.3 – Le périodogramme pour analyser des composantes périodiques

Chapitre 3

Tendances et facteurs saisonniers

Ce chapitre décrit les modèles généraux “classiques” de séries chronologiques, composées de tendances “à long terme”, de composantes périodiques ou presque périodiques, et d’une partie aléatoire. Les techniques développées permettent d’identifier et éliminer les tendances et les composantes périodiques.

3.1 Définitions et traitements

3.1.1 Les modèles

Les tendances et facteurs saisonniers rencontrés le plus fréquemment peuvent être représentés par trois types de modèles :

3.1.1.1 Le modèle additif

$$x_i = g_i + s_i + w_i$$

Si le phénomène est observé à des instants “réguliers”, la valeur n° i est observée à l’instant it_e , où t_e est la période d’échantillonnage ou période d’observation.

x_i , $i = 1, \dots, N$, est la série observée, g_i la tendance, s_i le facteur saisonnier et w_i la composante aléatoire (ou irrégulière, ou fluctuation résiduelle). Par convention, on impose à s_i d’avoir une moyenne nulle, pour séparer de façon non ambiguë la tendance et le facteur saisonnier :

$$\forall 1 \leq i \leq n - p, \sum_{k=1}^p s_{i+k} = 0 ,$$

où p est la ‘période’ de la fluctuation saisonnière. On impose également à w_i d’avoir une moyenne nulle, la moyenne de x_i étant alors également comprise dans la tendance g_i .

Dans ce modèle, les amplitudes des variations saisonnières et de la composante irrégulière à un instant donné ne dépendent pas de la valeur de la composante tendance au même instant.

On reprend à la Figure 3.1 deux exemples déjà donnés précédemment, où des tendances et variations saisonnières sont nettement visibles.

3.1.1.2 Le modèle multiplicatif

$$x_i = g_i(1 + s_i)(1 + w_i)$$

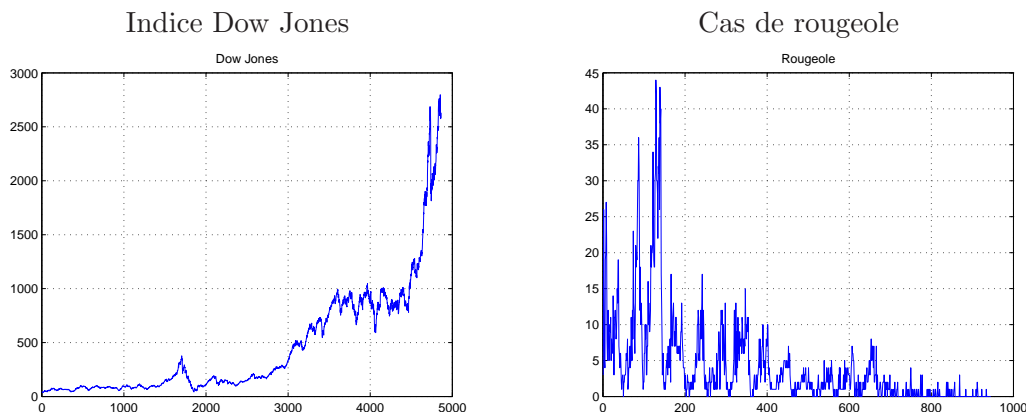


FIG. 3.1 – Tendances et variations saisonnières

où $x_i (i = 1, N)$ est la série observée, g_i la tendance, $(1 + s_i)$ le facteur saisonnier et $(1 + w_i)$ la composante aléatoire (ou irrégulière, ou fluctuation résiduelle). Cette formulation conduit aux propriétés suivantes :

- le facteur saisonnier $g_i s_i$ qui s'ajoute à la tendance est proportionnel à la valeur de celle-ci,
- la composante irrégulière $g_i(1 + s_i)w_i$ qui s'ajoute à la somme des deux termes précédents est elle-même proportionnelle à cette somme.

3.1.1.3 Les modèles hybrides

Différentes combinaisons de modèles additifs et de modèles multiplicatifs existent, qualifiées de modèles hybrides (ou mixtes), par exemple :

$$x_i = (g_i + s_i)(1 + w_i)$$

3.1.2 Les phases de traitement

Face à ces modèles, qui recouvrent une grande partie des séries temporelles rencontrées dans la pratique, on distingue deux grandes phases de traitement :

- déterminer s'il existe des tendances et / ou des facteurs saisonniers, et de quel type,
- analyser finement et au besoin éliminer ces tendances et facteurs saisonniers, pour une caractérisation ultérieure de la partie résiduelle.

Ce cours est orienté essentiellement vers les modèles additifs.

3.2 Analyse

3.2.1 Analyse visuelle

Dans beaucoup de cas, une simple analyse visuelle permet de dire s'il existe dans une série temporelle une tendance et / ou des fluctuations saisonnières (se reporter en particulier aux exemples donnés dans le chapitre 'Introduction').

3.2.2 Analyse par fonction de covariance échantillonnée

Dans le chapitre d'introduction, on a défini la fonction de covariance ou d'autocovariance, ainsi que la fonction de corrélation ou d'autocorrélation, d'un processus aléatoire *stationnaire*. Cette

fonction, associée à la moyenne (constante) du processus, définit entièrement ses caractéristiques statistiques du deuxième ordre. Rappelons qu'elle est définie de la façon suivante :

$$\begin{aligned}\gamma(\tau) &= \text{Cov}(X_t, X_{t+\tau}) \\ &= \mathbb{E}[(X_t - \mathbb{E}(X_t))(X_{t+\tau} - \mathbb{E}(X_{t+\tau}))] \\ &= \mathbb{E}(X_t X_{t+\tau}) - [\mathbb{E}(X_t)]^2\end{aligned}$$

Il n'a pas été précisé pour l'instant de quelle façon il est possible d'estimer cette fonction à partir de l'observation d'une réalisation du processus.

Un outil voisin est utilisé pour l'analyse générale de séries chronologiques *quelconques*, il s'agit de la fonction de covariance échantillonnée (ou empirique, ou estimée), définie, à partir des valeurs x_1 à x_n de la série, par (cas des séries réelles) :

$$\gamma_e(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - m_e)(x_j - m_e) \quad \text{pour } 0 \leq h < n,$$

où m_e est la moyenne empirique :

$$m_e = \frac{1}{n} \sum_{j=1}^n x_j.$$

L'utilisation du terme n au dénominateur (au lieu de $n - h$) garantit que :

- la matrice carrée Γ_e définie par $\Gamma_e(i, j) = \gamma_e(i - j)$ est définie non-négative,
- la fonction $\gamma_e(h)$ tend vers 0 lorsque h augmente.

Les propriétés asymptotiques de la fonction de covariance empirique γ_e et de la matrice Γ_e seront étudiées ultérieurement, pour l'estimation des processus *ARMA*.

On définit également la fonction de corrélation empirique, qui est la fonction de covariance normalisée par sa valeur en $h = 0$, soit :

$$\rho_e(h) = \frac{\gamma_e(h)}{\gamma_e(0)}$$

La fonction de covariance empirique non centrée est donnée par :

$$\gamma_{nc}(h) = \frac{1}{n} \sum_{j=1}^{n-h} x_{j+h} x_j \quad \text{pour } 0 \leq h < n$$

Il existe la relation suivante entre la fonction de covariance empirique non centrée et le périodogramme :

$$\sum_{|h| < n} \gamma_{nc}(h) e^{-jh\omega} = \frac{1}{n} \left| \sum_{i=1}^n x_i e^{-ji\omega} \right|^2 \quad (3.1)$$

La fonction de covariance empirique présente les propriétés suivantes, utiles pour la détection de tendances et de composantes saisonnières :

- elle décroît lentement lorsqu'une tendance est présente (les valeurs successives de la série sont fortement corrélées positivement),
- elle présente des pics périodiques marqués lorsque la série comprend une composante périodique ; ces pics sont plus visibles dans la fonction de covariance empirique que dans la série elle-même, grâce à l'effet d'intégration apporté par le calcul des coefficients de covariance, il est donc possible de détecter des périodicités "cachées" (par exemple par de fortes fluctuations aléatoires additionnelles),
- dans le cas d'une série ne présentant ni tendance ni variations saisonnières, elle ne présente pas de pics périodiques, et sa décroissance est soit :

- très rapide (processus peu corrélés),
- assez rapide, mais régulière (processus corrélés).

La fonction de corrélation empirique présente les mêmes propriétés. On démontre également que la mise en évidence de tendances par cette méthode est d'autant plus efficace que le nombre d'échantillons disponibles est plus grand.

La Figure 3.2 illustre les propriétés citées.

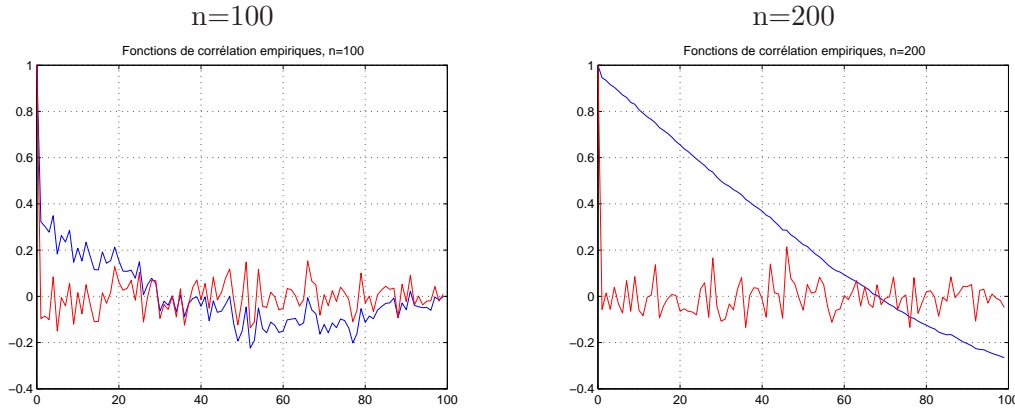


FIG. 3.2 – Corrélation empirique et tendance

Deux courbes sont représentées sur chacune de ces deux figures :

- en bleu : la fonction de corrélation empirique d'une série égale à la somme d'une tendance polynomiale du deuxième degré et d'un bruit blanc,
- en rouge : la fonction de corrélation empirique de la même série à laquelle on a retiré l'estimée de sa tendance quadratique.

La figure de gauche a été obtenue avec un échantillon de taille 100, celle de droite avec un échantillon de taille 200, avec une représentation de la fonction de corrélation sur 100 points. La comparaison de ces deux courbes nous confirme que la mise en évidence de tendances est plus efficace quand n augmente.

Exercice 3.1 Soit $x_t = a + bt$ une tendance déterministe affine (a et b sont des constantes réelles).

1. Montrer que l'autocorrélation empirique $\hat{\rho}_{x,n}(k)$ tend vers 1 pour tout k fixé lorsque n tend vers l'infini.
2. Soit $X_t = x_t + Y_t$, où Y_t est un processus aléatoire stationnaire centré. On suppose que pour tout k fixé,

$$\hat{\gamma}_{Y,n}(k) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \gamma_Y(k),$$

et :

$$\hat{\gamma}_{X,n}(k) = (\hat{\gamma}_{x,n}(k) + \hat{\gamma}_{Y,n}(k))(1 + o_P(1)),$$

quand n tend vers l'infini. Montrer que $\hat{\rho}_{X,n}(k)$ tend vers 1 en probabilité pour tout k fixé lorsque n tend vers l'infini.

3.2.3 Analyse spectrale

Dans le cadre de la recherche de tendances ou de variations saisonnières, l'analyse spectrale permet essentiellement de révéler des signaux (ou des sommes de signaux) sinusoïdaux fortement

noyés dans du bruit (composantes périodiques “cachées”). Le périodogramme, qui est le carré du module de la Transformée de Fourier Discrète, constitue la forme la plus simple d’analyse spectrale.

3.3 Traitement des tendances

On s’intéresse dans ce chapitre au cas où il n’existe pas de variations saisonnières, la série à traiter est donc de la forme suivante :

$$x_i = g_i + w_i$$

Cependant, le cas de certaines variations saisonnières pourra être traité par la méthode des moindres carrés (paragraphe 3.3.1).

3.3.1 Estimation de tendance par la méthode des moindres carrés

Cette méthode suppose que la fonction g , qui représente la tendance, peut être représentée de façon paramétrique “simple”, c’est-à-dire comportant un petit nombre de paramètres. C’est en particulier le cas des polynômes de degré faible (tendance linéaire, tendance quadratique ...). Il s’agit donc d’une méthode dite *paramétrique*.

Pour un degré de polynôme donné, on cherche les coefficients de ce polynôme qui permettent d’approcher au mieux, au sens des moindres carrés, la série initiale. Par exemple, pour une tendance supposée linéaire, on cherche les coefficients a_0 et a_1 qui minimisent l’expression :

$$S(a_0, a_1) = \sum_i (x_i - a_0 - a_1 i)^2$$

soit :

$$(\hat{a}_0, \hat{a}_1) = \operatorname{argmin}_{a_0, a_1} S(a_0, a_1)$$

Cette opération est encore appelée “régression linéaire”.

Remarque : la méthode des moindres carrés ne fait aucune hypothèse explicite sur la partie résiduelle w_i . Une connaissance plus détaillée de cette partie résiduelle, en particulier de ses caractéristiques statistiques, permet d’estimer les paramètres de la droite avec plus de précision.

La technique de régression linéaire se généralise pour un polynôme de degré d , où l’on cherche à minimiser la somme suivante :

$$S(a_0, a_1, \dots, a_d) = \sum_i (x_i - a_0 - a_1 i - \dots - a_d i^d)^2$$

c’est-à-dire à trouver le vecteur $A = (\hat{a}_0, \hat{a}_1, \dots, \hat{a}_d)$ qui minimise S .

Soit I la matrice de dimension $(n, d+1)$, définie par :

$$I_{i,j} = i^{j-1} \quad i = 1, n \quad j = 1, d+1$$

alors on démontre que :

$$A^T = (I^T I)^{-1} I^T x$$

où x est le vecteur colonne composé des x_i .

La série y correspondant au polynôme ainsi estimé est donnée par :

$$y = I A^T = H x$$

où H est la matrice chapeau (“Hat Matrix”) définie par :

$$H = I(I^T I)^{-1} I^T$$

Il se pose alors la question du choix du degré optimal du polynôme qui modélise la tendance. Si la tendance est réellement polynomiale de degré d , on souhaite que le degré choisi soit effectivement égal à d . Pour parvenir à cet objectif, une méthode consiste à calculer la tendance pour des degrés croissants, et la somme quadratique des résidus pour chacun de ces degrés. Lorsque cette somme cesse de décroître, on a atteint le degré optimal.

Pour affiner cette procédure, on peut utiliser par exemple une technique de pénalisation : ajouter à la somme quadratique des résidus un terme qui augmente avec la complexité du modèle (ici d), la fonction obtenue présente alors un minimum pour la meilleure valeur de d , i.e. le meilleur compromis entre adéquation du modèle aux données et complexité du modèle.

Nota 1 : la méthode de régression linéaire se généralise pour les fonctions linéaires par rapport aux paramètres (par exemple $a \cos(\omega i) + b \sin(\omega i)$, où ω est connu). La valeur de ω s'obtient par exemple grâce au périodogramme.

Nota 2 : cette méthode permet de prendre en compte aisément les données manquantes.

Globalement, cette technique permet de séparer la tendance et le résidu, pour une éventuelle analyse ultérieure de celui-ci.

La Figure 3.3 illustre la méthode. La courbe bleue représente la série, constituée de la somme

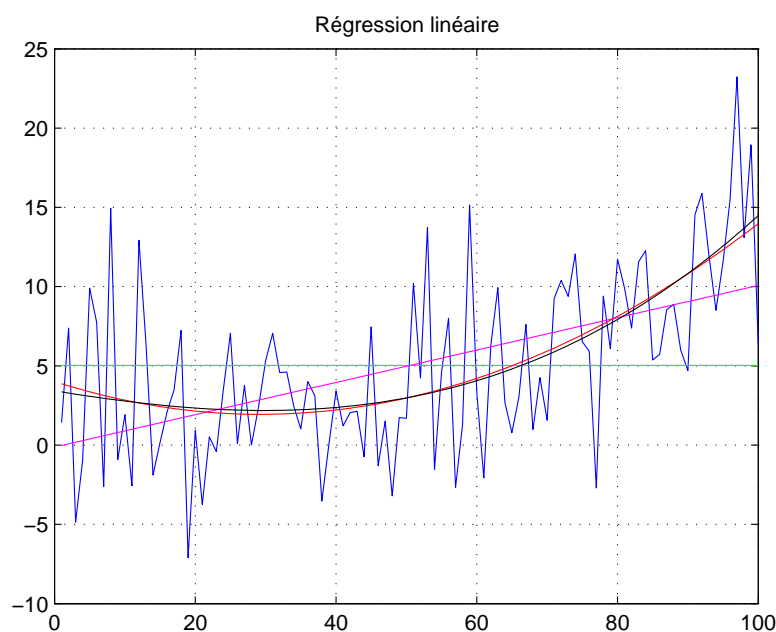


FIG. 3.3 – Une régression linéaire

d'une tendance quadratique et d'un bruit blanc. La courbe verte (horizontale) représente la régression d'ordre 0 (soit une constante égale à la moyenne des valeurs de la série), la droite magenta est la régression d'ordre 1 (droite des moindres carrés), les deux dernières courbes (rouge et noire), quasiment confondues, sont les régressions linéaires sur des polynômes de degrés 2 et 3 respectivement. Les résidus successifs sont : 3176, 2309, 1988, 1984. Il apparaît donc clairement, malgré un bruit relativement fort, que l'ordre optimal est bien 2.

Exercice 3.2 On reprend l'exemple de la tendance linéaire du paragraphe 3.3.1.

1. Trouver les formules explicites qui donnent (\hat{a}_0, \hat{a}_1) en fonction des observations x_i .
Indication : dériver l'expression $S_{lin}(a_0, a_1)$ par rapport à a_0 et a_1 et trouver le couple (\hat{a}_0, \hat{a}_1) qui annule ces deux dérivées.
2. Démontrer que cette méthode est équivalente à trouver le couple (\hat{a}_0, \hat{a}_1) le plus probable dans le cas où la séquence des résidus est une séquence de variables aléatoires gaussiennes i.i.d. de variance connue ou inconnue.
3. Que devient cette formule dans le cas où la séquence des résidus est constituée de variables aléatoires gaussiennes indépendantes, dont la variance V_i dépend de i de façon connue (moindres carrés pondérés) ?
4. Montrer que cette méthode des moindres carrés s'applique à d'autres types de tendances que les tendances polynomiales ou sinusoidales. Quelles sont les difficultés potentielles ?

Exercice 3.3 On reprend l'exemple de la tendance polynomiale du paragraphe 3.3.1.

1. Démontrer que la formule $A^T = (I^T I)^{-1} I^T x$ donne l'ensemble des valeurs $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_d$ les plus probables dans le cas où la séquence des résidus est une séquence de variables aléatoires gaussiennes i.i.d. de variance connue ou inconnue.
Indication : l'ensemble des équations $x_i = a_0 + a_1 i + \dots + a_d i^d \quad i = 1, \dots, n$ peut s'écrire de façon concise sous la forme $x = I A^T + w$ où w est le vecteur colonne composé des w_i .
2. Que devient cette formule si la séquence des résidus est corrélée de façon connue, avec la matrice de covariance Cw , $Cw_{i,j} = E(w_i w_j) \quad i, j = 1, \dots, n$?

3.3.2 Lissage par la méthode de la Moyenne Mobile

La méthode de lissage par la Moyenne Mobile est une méthode *non paramétrique* qui permet de reproduire le plus fidèlement possible la tendance, tout en éliminant le résidu. L'avantage de cette méthode est qu'elle ne fait pas d'hypothèse sur la tendance, qui peut avoir une forme quelconque.

Il existe trois sous-méthodes :

- la moyenne mobile symétrique, définie par :

$$y_i = \frac{1}{2q+1} \sum_{j=-q}^q x_{i+j}$$

où q est un nombre entier, et :

$$q+1 \leq i \leq n-q$$

Le choix de q résulte encore d'un compromis : une valeur de q trop faible ne permet pas d'extraire la tendance du résidu, une valeur trop élevée rend mal compte des *évolutions* de la tendance.

- la moyenne mobile récursive, définie, pour $0 \leq \alpha \leq 1$, par :

$$\begin{aligned} y_i &= \alpha x_i + (1 - \alpha) y_{i-1} \quad \text{pour } i > 1 \\ y_1 &= x_1 \end{aligned}$$

Elle est encore appelée "lissage exponentiel".

Le cas particulier $\alpha = 0$ conduit à $y_i = x_1$ pour tout i .

Le cas particulier $\alpha = 1$ conduit à $y_i = x_i$ pour tout i , donc à une absence totale de lissage.

– la moyenne mobile pondérée, définie par :

$$y_i = \sum_j \alpha_j x_{i-j} .$$

On peut obtenir des propriétés particulières du lissage par un choix approprié des α_i .

Exercice 3.4 Donner des conditions nécessaires sur les α_j pour que cette opération ne provoque aucune distorsion sur une tendance linéaire.

Voir exercice complémentaire 3.12 pour obtenir des propriétés plus complexes du lissage.

Le résultat obtenu dans cet exercice (trouver des conditions nécessaires sur les α_j pour que la moyenne mobile ne provoque pas de distorsion sur une tendance linéaire) se généralise pour un polynôme de degré plus élevé, voir exercice complémentaire 3.12. Quand le nombre des α_j est plus grand que le degré du polynôme plus 1, il existe une infinité de solutions ; parmi ces solutions, il en existe une qui minimise la puissance du résidu (optimisation sous contrainte, à condition de connaître les caractéristiques de la partie aléatoire du processus).

Exercice 3.5 On s'intéresse à l'effet d'une moyenne mobile sur un bruit blanc. Soit M une moyenne mobile de la forme :

$$M = \sum_{i=-m_1}^{m_2} \theta_i B^{-i} ,$$

et $(Z_t, t \in \mathbb{Z})$ un bruit blanc de moyenne nulle et variance σ^2 . On pose :

$$\forall t \in \mathbb{Z}, \quad Y_t = M Z_t .$$

1. Montrer que pour tout $t \in \mathbb{Z}$, $\mathbb{E}[Y_t] = 0$ et, pour tout $t \in \mathbb{Z}$ et tout $k \in \mathbb{Z}$,

$$\text{Cov}[Y_t, Y_{t+k}] = \sigma^2 \sum_{j \in \mathbb{Z}} \theta_j \theta_{j-k} = \gamma(k) , \quad (3.2)$$

où on a posé $\theta_j = 0$ si $j < -m_1$ ou $j > m_2$.

2. Soit $X_t = m_t + s_t + Z_t$, avec m_t une tendance quadratique et s_t une saisonnalité de période p . On suppose que M laisse invariantes les tendances quadratiques et absorbe les saisonnalités de période p . Que vaut $M X_t$? Le bruit dans la série $M X_t$ est-il un bruit blanc ? Que vaut sa variance, et sous quelle condition est-elle plus faible que σ^2 ?
3. Montrer que γ vérifie :

$$\sum_{k \in \mathbb{Z}} \gamma(k) z^k = \sigma^2 M(z) M\left(\frac{1}{z}\right) ,$$

où $M(z) = \sum_{j \in \mathbb{Z}} \theta_j z^j$ est la fonction de transfert du filtre constitué par la moyenne mobile M .

Exercice 3.6 Dans le cas de la moyenne mobile récursive, exprimer y_i en fonction de x_1, \dots, x_i et interpréter l'expression "lissage exponentiel".

3.3.3 Elimination par différentiation

Soit l'opérateur de différentiation défini par :

$$\Delta x_i = x_i - x_{i-1} = (1 - B)x_i$$

où B est l'opérateur 'backward' : $Bx_i = x_{i-1}$.

L'opérateur de différentiation peut être doublé, triplé,...

$$\Delta^2 x_i = \Delta \Delta x_i = (1 - B)^2 x_i = x_i - 2x_{i-1} + x_{i-2}$$

Attention, ne pas confondre :

$$(1 - B)^2 x_i = x_i - 2x_{i-1} + x_{i-2}$$

avec :

$$(1 - B^2)x_i = x_i - x_{i-2}$$

Exercice 3.7 Montrer l'effet de l'opérateur de différentiation sur une tendance polynomiale d'ordre k . En déduire quel opérateur il suffit d'appliquer à une série à tendance polynomiale d'ordre k pour réduire cette tendance à une constante ou à 0. Caractériser la conséquence de l'application de cette opération sur le résidu. En déduire le domaine d'application de cette méthode.

3.3.4 Tendances de formes spécifiques

Il arrive que l'analyse visuelle suggère une tendance autre que polynomiale (par exemple logarithmique ou exponentielle). Un changement de variable approprié permet en général de se ramener à une tendance linéaire.

On voit à la Figure 3.4 un exemple simple de traitement d'une tendance en $1/x$, qu'on a cherché à linéariser par la transformation non linéaire $y = \log(x)$. Sur cet exemple, la série

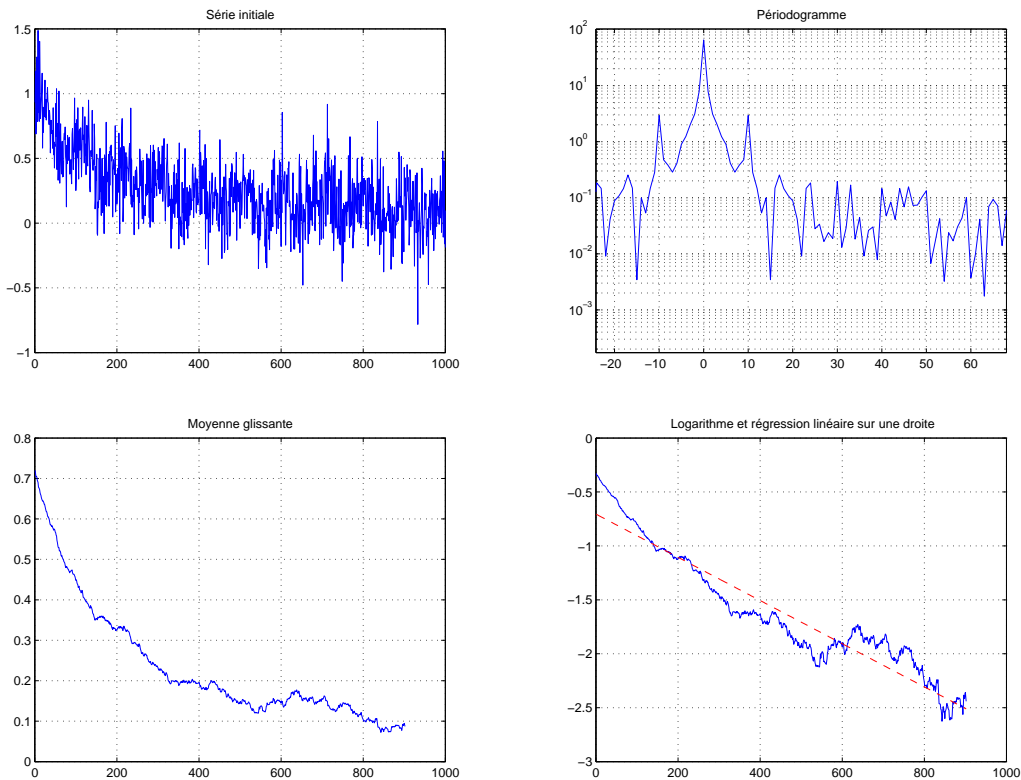


FIG. 3.4 – Une tendance en $1/x$

initiale a été synthétisée (figure en haut à gauche), elle comprend une tendance en $1/x$ plus une variation saisonnière sinusoïdale, plus un bruit. La deuxième figure (en haut à droite) montre le périodogramme qui fait clairement apparaître la variation saisonnière. On voit qu'après correction des variations saisonnières par une moyenne glissante (troisième figure) et transformation par la fonction logarithme (quatrième figure), la série peut se modéliser par la somme d'une

tendance linéaire décroissante et d'une composante aléatoire. Il faut noter que la transformation non linéaire a modifié la statistique de la partie aléatoire.

L'exemple suivant (figure 3.5) montre la suite des mêmes opérations sur un cas réel (rougeole, où l'on a retiré les cent premières valeurs de la série). Le résultat est analogue à celui du cas précédent.

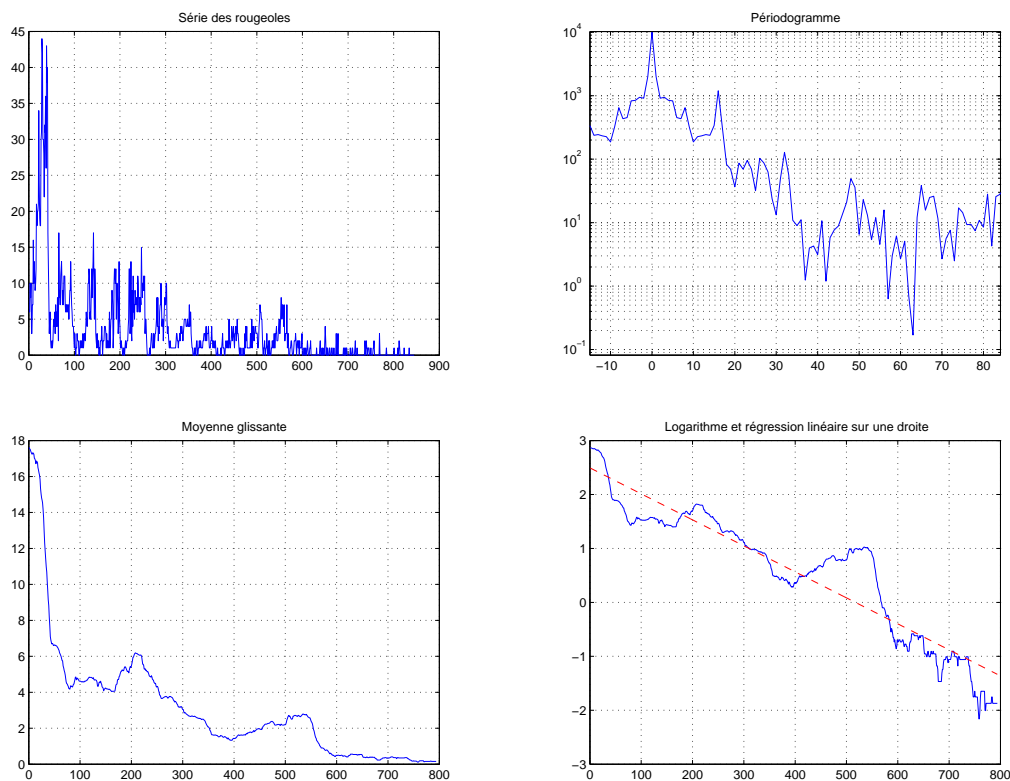


FIG. 3.5 – Un exemple réel de tendance en $1/x$

Exercice 3.8 Soit la courbe de “type exponentiel” donnée par la formule :

$$g_t = ae^{bt} + \gamma.$$

A quel système d'équations conduit la méthode des moindres carrés ? Quelle conclusion en tire-t-on ? Montrer que si l'on choisit seulement trois points équidistants en abscisse et supposés appartenir à la courbe, il est possible de déterminer algébriquement ses paramètres (méthode des trois points).

On peut rencontrer également des courbes *logistiques*, par exemple dans le cas des séries présentant une valeur “plancher” et une valeur “plafond”.

La courbe logistique est de la forme :

$$x_i = \frac{1}{ae^{bi} + c}$$

La Figure 3.6 donne un exemple de courbe logistique, pour $a = 1$, $b = -1$, $c = 1$.

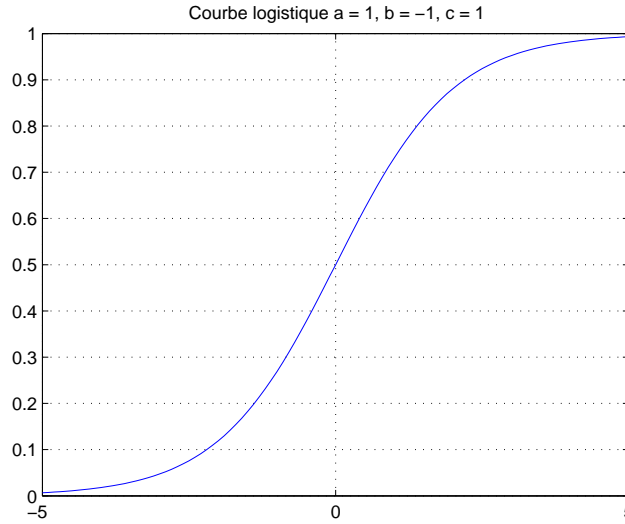


FIG. 3.6 – Un exemple de courbe logistique

3.3.5 Traitement des tendances et estimation de fonctions

D'une façon générale, le traitement des tendances peut être considéré comme un problème d'estimation fonctionnelle et plus précisément comme un problème de débruitage lorsque la série à traiter est supposée de la forme $x_i = g_i + w_i$ comme cela est le cas dans ce chapitre.

L'estimation fonctionnelle 'locale' fait appel à deux grandes classes d'estimateurs : les estimateurs à noyau et les estimateurs par projection orthogonale.

Les méthodes locales à base de noyau estiment g_i par la formule générale suivante :

$$\hat{g}_i = \sum_t [K(\frac{i-t}{\delta}) / \sum_s K(\frac{i-s}{\delta})] x_i$$

où la fonction K est un noyau régularisant (par exemple une fonction gaussienne) et δ sert à régler la largeur de la fenêtre temporelle. On peut remarquer que la méthode décrite au paragraphe 3.3.2 (Lissage par la méthode de la moyenne mobile) est un cas particulier d'estimateur fonctionnel à noyau.

Les méthodes par projection orthogonale supposent que la fonction g appartient à un espace G pour lequel on dispose d'une base orthogonale e_m , soit $g = \sum_m \gamma_m e_m$. L'estimateur de g est obtenu en déterminant des estimateurs $\hat{\gamma}_m$ de γ_m à partir des x_i et en posant $\hat{g} = \sum_m \hat{\gamma}_m e_m$.

On peut citer ici une technique de débruitage très efficace, qui utilise des ondelettes. Les techniques d'ondelettes étant développées dans un autre module, on ne donne ici que quelques éléments (voir par exemple [15]) : dans une version simplifiée, on pose $\hat{g} = \sum_k \hat{\beta}_{J,k} \phi_{J,k}$ qui n'utilise que la fonction d'échelle de niveau J . C'est donc un estimateur par projection orthogonale, mais on peut démontrer que c'est également un estimateur à noyau. L'estimation peut être complétée en utilisant des termes de détail :

$$\hat{g} = \sum_k \hat{\beta}_{J,k} \phi_{J,k} + \sum_{J \leq j \leq J_0} \sum_k \hat{\alpha}_{j,k} \psi_{j,k}$$

Un choix judicieux des coefficients $\hat{\alpha}_{j,k}$, par exemple par un seuillage des coefficients d'ondelettes du signal x_i , permet de débruiter le signal tout en conservant les irrégularités dont l'amplitude est significative par rapport à celle du bruit. Cette propriété, rendue possible par

l'effet non linéaire du seuillage, n'est pas présente dans les autres méthodes décrites, qui sont toutes linéaires.

3.4 Traitement global

On se place ici dans le cadre général où la série chronologique contient simultanément une tendance et des variations saisonnières supposées *périodiques* ; dans le cas du modèle additif, on a donc :

$$x_i = g_i + s_i + w_i$$

Pour l'exposé, les hypothèses et notations suivantes seront prises, sans nuire à la généralité :

- la période des variations saisonnières est égale à p , on a donc :

$$\forall 1 \leq i \leq n - p, s_{i+p} = s_i .$$

Par convention, on impose à s_i d'avoir une moyenne nulle, pour séparer de façon non ambiguë la tendance et le facteur saisonnier :

$$\forall 1 \leq i \leq n - p, \sum_{k=1}^p s_{i+k} = 0 .$$

- le nombre de périodes est entier et égal à P , ce qui conduit à $n = pP$, où n est le nombre total d'observations. Il en résulte que la période n° j , pour $1 \leq j \leq P$, couvre les échantillons numérotés de $(j-1)p + 1$ à jp .

Les méthodes diffèrent selon les vitesses de variation relatives de la tendance et des fluctuations saisonnières.

3.4.1 Méthode des "tendances faibles"

On suppose ici que la tendance g est suffisamment faible (ou lentement variable) pour pouvoir être assimilée à une constante pendant une période de la variation saisonnière s , c'est-à-dire p échantillons.

On estime alors successivement la tendance g pour chaque période et les variations saisonnières par les formules suivantes :

- tendance g pour chaque période n° j :

$$\hat{g}_j = \frac{1}{p} \sum_{i=(j-1)p+1}^{jp} x_i \quad j = 1, \dots, P$$

- variations saisonnières :

$$\hat{s}_k = \frac{1}{P} \sum_{j=1}^P (x_{(j-1)p+k} - \hat{g}_j) \quad k = 1, \dots, p$$

3.4.2 Estimation par la méthode de la moyenne mobile

Cette méthode est plus générale que la précédente, car elle accepte une tendance non constante pendant une période de la variation saisonnière.

Elle se décompose en trois étapes :

- une première estimation de la tendance, à structure imposée pour éliminer la composante saisonnière,
- une estimation de la composante saisonnière comme variation périodique autour de cette tendance, suivie d'une 'désaisonnalisation' de la série initiale,
- une nouvelle estimation de la tendance, par une des méthodes décrites au paragraphe précédent.

Première étape : la tendance est estimée par une moyenne mobile symétrique sur une longueur p (structure imposée) :

- si la période est impaire ($p = 2q + 1$), on applique la formule de la moyenne mobile vue précédemment :

$$\widehat{g1}_i = \frac{1}{2q+1} \sum_{j=-q}^q x_{i+j} = \frac{1}{p} \sum_{j=-q}^q x_{i+j} \quad i = q+1, \dots, n-q$$

- si la période est paire ($p = 2q$), il faut légèrement modifier la formule pour respecter la symétrie :

$$\widehat{g1}_i = \frac{1}{p} (0.5x_{i-q} + \sum_{j=-q+1}^{q-1} x_{i+j} + 0.5x_{i+q}) \quad i = q+1, \dots, n-q$$

Deuxième étape : estimation de la composante saisonnière, respectant la contrainte de moyenne nulle, suivie d'une désaisonnalisation :

$$s1_k = \frac{1}{P-2} \sum_{j=2}^{P-1} (x_{k+(j-1)p} - \widehat{g1}_{k+(j-1)p}) \quad k = 1, \dots, p$$

nota : on ne peut pas utiliser la première ni la dernière période, car la moyenne mobile a tronqué la série.

$$\widehat{s2}_k = s1_k - \frac{1}{p} \sum_{i=1}^p s1_i \quad k = 1, \dots, p$$

$$\widehat{g2}_i = x_i - \widehat{s2}_{i-pE(\frac{i-1}{p})} \quad i = 1, \dots, n$$

$g2$ contient la tendance plus le bruit.

Troisième étape : on réestime la tendance, par une des méthodes décrites précédemment au paragraphe 3.3.

La technique décrite ci-dessus estime en fait la tendance en deux étapes. Des techniques plus complexes permettent d'éliminer des composantes saisonnières (de période connue), tout en respectant sans distorsion des tendances polynomiales d'ordre donné (voir Exercice 3.12).

3.4.3 Elimination par différentiation

Exercice 3.9 Pour les questions qui suivent, on se basera sur le modèle général d'une série avec tendance de période p et fluctuation saisonnière (additifs) : $X_t = m_t + s_t + W_t$.

1. Trouver un opérateur retard, noté Δ_p , qui élimine la tendance saisonnière de période p .
2. Quel est l'effet de cet opérateur sur une tendance ?
3. Comment éliminer cette tendance modifiée ?
4. Quels sont les avantages et inconvénients de cette méthode ?

3.5 Elimination de composantes spectrales

La présence d'un signal sinusoïdal fortement noyé dans du bruit peut être décelée par une analyse spectrale, de type périodogramme. Le signal a la forme suivante :

$$x_i = ae^{j\omega i} + w_i ,$$

où a est un nombre complexe, et ω est la pulsation 'réduite'.

Dire que le rapport signal sur bruit est petit est équivalent à $a \ll \sigma$ où σ est l'écart-type de w .

Considérons l'opération de *filtrage* définie par la relation suivante :

$$y_i - \alpha e^{j\omega} y_{i-1} = \alpha(x_i - e^{j\omega} x_{i-1}) \quad 0 < \alpha < 1 \quad (3.3)$$

Cette opération permet de supprimer la composante de pulsation ω , quelle que soit son amplitude complexe a . Elle est obtenue par la succession de trois opérations :

- décaler le signal en pulsation d'une quantité $-\omega$,
- appliquer au signal un filtre qui rejette la pulsation nulle,
- redécaler le signal en pulsation d'une quantité $+\omega$.

Exercice 3.10 Montrer que l'opération définie par la relation 3.3 permet de supprimer la composante de pulsation ω , quelle que soit son amplitude complexe a .

Quel est son effet sur le bruit w ?

Quelle analogie y a-t-il avec les modèles ARMA ?

Les six courbes de la figure 3.7 montrent l'effet d'une telle opération dans le cas où la composante spectrale est une sinusoïde au lieu d'une exponentielle complexe :

$$x_i = a \sin \omega i + w_i$$

(les figures ne montrent qu'une partie des données pour une meilleure visibilité).

Dans ce cas, il faut appliquer deux fois l'opération 3.3 au signal x , une fois pour la pulsation ω , et une fois pour la pulsation $-\omega$, ce qui conduit à la formule de filtrage suivante :

$$y_i - 2\alpha \cos(\omega)y_{i-1} + \alpha^2 y_{i-2} = \alpha^2 [x_i - 2\cos(\omega)x_{i-1} + x_{i-2}]$$

- les deux premières figures montrent respectivement le signal x et son périodogramme, la première faisant apparaître la sinusoïde seule (trait noir, artificiellement décalée vers le bas pour la lisibilité de la figure) et x (en bleu),
- les deux figures suivantes montrent respectivement le signal y et son périodogramme,
- les deux dernières figures montrent pour comparaison :
 - x en bleu et y en rouge,
 - le périodogramme de x en bleu et celui de y en rouge.

On voit clairement à la Figure 3.7 la différence entre x et y dans le domaine spectral ¹(périodogrammes), mais la différence est imperceptible dans le domaine temporel.

Ces courbes ont été obtenues avec les paramètres suivants :

- nombre d'échantillons : 2048
- fréquence réduite de la sinusoïde : 0.05
- $\alpha = 0.9$

Cette technique est une alternative à celle des moindres carrés décrite au paragraphe 3.3.1 : elle est en un sens plus performante, car elle élimine totalement la composante spectrale de pulsation ω sans qu'il soit nécessaire d'en estimer l'amplitude et la phase, mais en contrepartie elle supprime également des composantes spectrales proches de la pulsation ω .

¹le périodogramme de y n'est pas rigoureusement nul pour les pulsations $\pm \omega$; ceci est dû au fait que le signal traité n'a pas une durée infinie

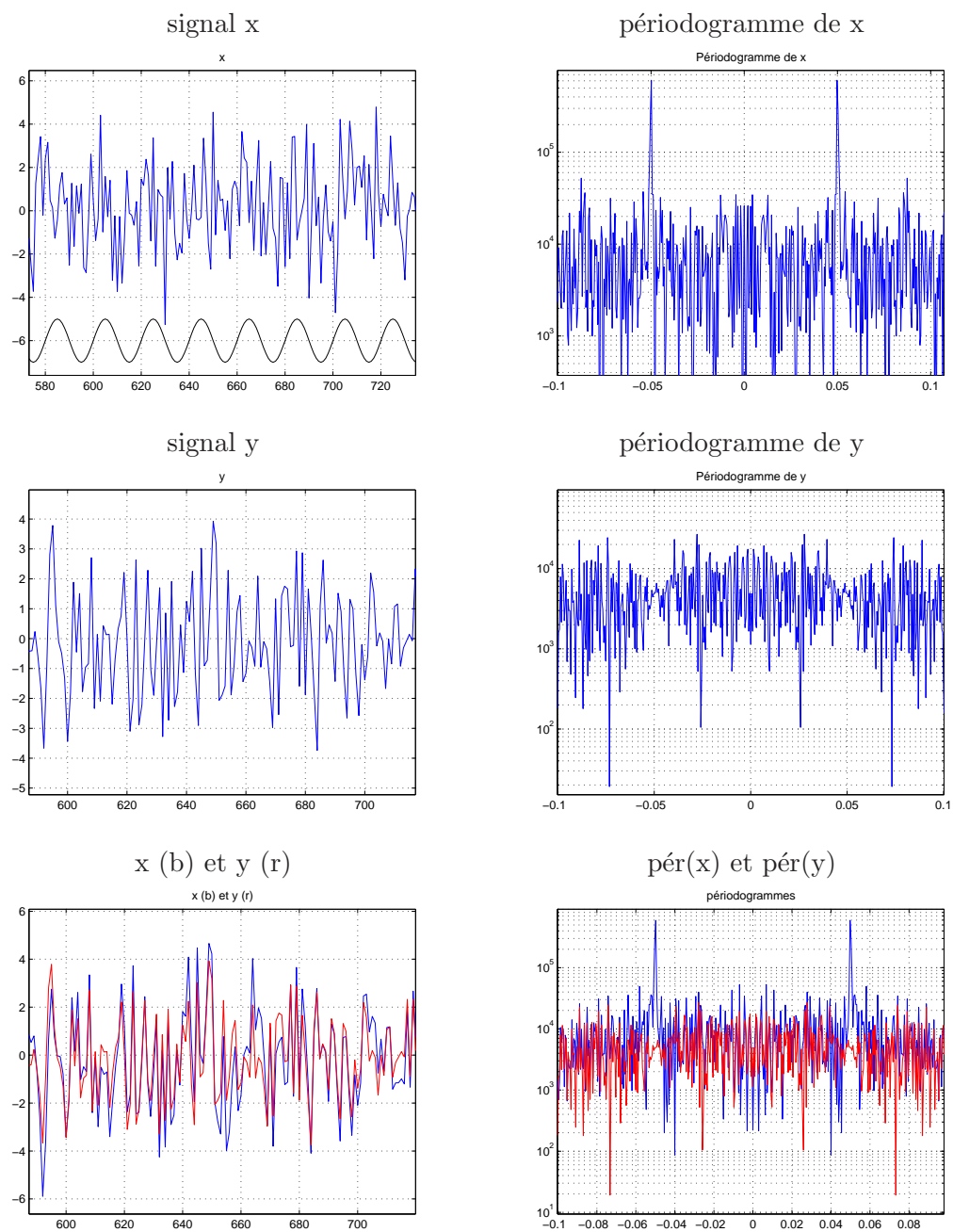


FIG. 3.7 – Filtrage d'une sinusoïde de fréquence connue

3.6 Exercices complémentaires

Exercice 3.11 Soient (m_t) est une tendance, (s_t) une composante saisonnière de période p et (Z_t) un processus stationnaire de moyenne nulle, et fonction d'autocovariance γ . Calculer l'espérance et la variance de X_t , pour tout $t \in \mathbb{Z}$, ainsi que la covariance entre X_t et X_{t+k} lorsque :

1. On considère le modèle additif :

$$\forall t \in \mathbb{Z}, \quad X_t = m_t + s_t + Z_t ,$$

2. On considère le modèle multiplicatif :

$$\forall t \in \mathbb{Z}, \quad X_t = m_t s_t + Z_t ,$$

3. On considère le modèle multiplicatif complet :

$$\forall t \in \mathbb{Z}, \quad X_t = m_t s_t (1 + Z_t) ,$$

Exercice 3.12 Construire une moyenne mobile symétrique d'ordre 5 qui élimine les composantes périodiques de période 3 et laisse passer les polynômes de degré inférieur ou égal à 2. Remarque : cette moyenne mobile doit donc s'écrire $M = a_2 B^{-2} + a_1 B^{-1} + a_0 I + a_1 B + a_2 B^2$. Un indice : une base des fonctions de période 3 et de moyenne nulle sur toute période est donnée par (s_1, s_2) , avec :

$$s_1(0) = 1, \quad s_1(1) = -1 \text{ et } s_2(0) = 1, s_2(2) = -1 .$$

Exercice 3.13 Soit M une moyenne mobile d'ordre $2q + 1$ telle que :

1. la variance du bruit soit réduite au maximum ;
2. les constantes soient conservées.

Montrer que M est la moyenne mobile arithmétique qui est de la forme suivante :

$$M = \frac{1}{2q+1} \sum_{i=-q}^q B^{-i} .$$

Remarquer également que M est symétrique et qu'elle laisse invariantes les tendances linéaires.

Chapitre 4

Prédiction linéaire - Modèles d'état - Filtrage de Kalman

De nombreuses séries chronologiques étudiées dans ce cours peuvent être représentées par un modèle d'état. Le filtrage de Kalman, fondé sur un modèle d'état, permet donc de réaliser de nombreuses opérations sur les séries chronologiques, en particulier la prédiction à un pas, nécessaire pour l'identification des processus. Les modèles d'état et le filtrage de Kalman permettent en outre de traiter aisément des cas particuliers, comme les données manquantes ou les données irrégulièrement espacées dans le temps.

4.1 Prédiction linéaire

4.1.1 Prédiction linéaire d'une variable aléatoire réelle

Soit (Y_1, \dots, Y_n, U) un vecteur aléatoire formé de $n + 1$ variables aléatoires réelles de carré intégrable. Une des questions principales de ce chapitre est de prédire U après avoir observé (Y_1, \dots, Y_n) . Nous allons y répondre dans le cadre suivant : on cherchera à prédire U comme une combinaison linéaire des Y_i (plus un terme constant), et on mesurera la qualité de la prédiction par la distance L^2 . Si on pose $Y_0 = 1$, on veut donc trouver des coefficients $(\alpha_0^*, \dots, \alpha_n^*)$ tels que, si on pose :

$$U^* = \sum_{i=0}^n \alpha_i^* Y_i ,$$

alors :

$$\mathbb{E}[(U - U^*)^2] = \min_{Y' \in \text{Vect}(Y_0, \dots, Y_n)} \mathbb{E}[(U - Y')^2] . \quad (4.1)$$

On voit donc que U^* est la projection orthogonale dans L^2 sur le sous-espace vectoriel engendré par Y_0, \dots, Y_n . La proposition suivante est un résultat classique d'algèbre linéaire.

Proposition 4.1 *Il existe une unique solution U^* à (4.1). On l'appelle le meilleur prédicteur linéaire de U sachant (Y_1, \dots, Y_n) . De plus, (4.1) est équivalente à :*

$$\forall j = 0, \dots, n, \quad \mathbb{E}(Y_j U) = \mathbb{E}(Y_j U^*) . \quad (4.2)$$

Si on cherche U^* sous la forme $\sum_{i=0}^n \alpha_i Y_i$, la proposition 4.1 nous dit qu'il faut et il suffit de résoudre les équations suivantes :

$$\forall j = 0, \dots, n, \quad \sum_{i=0}^n \alpha_i \mathbb{E}(Y_i Y_j) = \mathbb{E}(U Y_j) , \quad (4.3)$$

et on voit ainsi que les α_i ne dépendent que de la matrice de covariance du vecteur $(Y_0, Y_1, \dots, Y_n, U)$.

Remarquons aussi que lorsque (Y_1, \dots, Y_n, U) est un vecteur gaussien, le meilleur prédicteur linéaire est égal à l'espérance conditionnelle de U sachant (Y_1, \dots, Y_n) , et c'est la fonction de (Y_1, \dots, Y_n) qui approche le mieux U au sens L^2 .

4.1.2 Prédiction linéaire d'un vecteur aléatoire

On observe maintenant des vecteurs aléatoires $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$, où $\mathbf{Y}_i \in \mathbb{R}^k$ et on veut prédire un vecteur aléatoire $\mathbf{U} \in \mathbb{R}^d$, en supposant que toutes les coordonnées de ces vecteurs sont de carré intégrable. On appelle meilleur prédicteur linéaire de \mathbf{U} sachant $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ le vecteur de \mathbb{R}^d , dont la j -ème coordonnée est le meilleur prédicteur de U_j sachant les $n \times k$ coordonnées des vecteurs $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$. En notant $P(\mathbf{U}|\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ ce prédicteur linéaire, on a la proposition suivante.

Proposition 4.2 Soient $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ des vecteurs aléatoires de \mathbb{R}^k , \mathbf{U} un vecteur aléatoire de \mathbb{R}^d et V un vecteur aléatoire de \mathbb{R}^q .

– Si A est une matrice $v \times d$, alors :

$$P(A\mathbf{U}|\mathbf{Y}_1, \dots, \mathbf{Y}_n) = AP(\mathbf{U}|\mathbf{Y}_1, \dots, \mathbf{Y}_n) .$$

– En notant $M = \mathbb{E}(UY_1^T)(\mathbb{E}(Y_1Y_1^T))^{-1}$, où S^{-1} désigne une matrice telle que $SS^{-1}S = S$ (c'est une inverse généralisée), on a :

$$P(\mathbf{U}|\mathbf{Y}_1) = M\mathbf{U} ,$$

Notamment si \mathbf{V} est décorrélé de \mathbf{Y}_1 , alors :

$$P(\mathbf{V}|\mathbf{Y}_1) = 0 .$$

– Si \mathbf{V} est décorrélé de \mathbf{Y}_1 :

$$P(\mathbf{U}|\mathbf{Y}_1, \mathbf{V}) = P(\mathbf{U}|\mathbf{Y}_1) + P(\mathbf{U}|\mathbf{V}) .$$

La première partie de la proposition est triviale, la seconde partie est la conséquence directe de (4.3) et la dernière partie est un résultat classique pour la projection orthogonale sur la somme de deux sous-espaces orthogonaux.

4.1.3 Minimiser les erreurs de prédiction, maximum de vraisemblance et cadre gaussien

Plaçons-nous dans un cadre où :

- on observe $U = (U_1, \dots, U_n)$, un vecteur de n variables aléatoires réelles de carré intégrable,
- notre but est de prédire le mieux possible la valeur à venir U_{n+1} ,
- la loi de (U_1, \dots, U_{n+1}) appartient à un ensemble de lois $(\mathbb{P}_\theta)_{\theta \in \mathbb{R}^k}$.

Si on connaissait θ , le prédicteur linéaire U_{n+1}^* décrit dans la section 4.1.1 nous permettrait de prédire de manière optimale (pour un prédicteur linéaire) U_{n+1} en fonction de (U_1, \dots, U_n) . Si la loi de (U_1, \dots, U_{n+1}) est assez stationnaire, un critère naturel pour estimer θ est alors de minimiser la somme des carrés des erreurs de prédiction $\sum_{i=1}^n (U_i - U_{i,\theta}^*)^2$, où $U_{i,\theta}^*$ est le prédicteur linéaire de U_i sachant (U_1, \dots, U_{i-1}) , sous la loi \mathbb{P}_θ . Remarquons toutefois que les variables $U_i - U_{i,\theta}^*$ n'ont pas toutes la même variance, et il serait donc plus judicieux de minimiser :

$$\gamma(\theta, U) := \sum_{i=1}^n \frac{(U_i - U_{i,\theta}^*)^2}{v_{i-1}(\theta)} ,$$

où $v_{i-1}(\theta) = \mathbb{E}_\theta[(U_i - U_{i,\theta}^*)^2]$ est la variance de l'innovation $\nu_{i,\theta} = U_i - U_{i,\theta}^*$. Avec le filtre de Kalman présenté dans la section 4.4, cela nous donne une voie assez générale pour estimer un paramètre dans une optique de prédiction.

Finissons cette sous-section en remarquant l'écriture suivante.

Proposition 4.3 Notons $\Gamma_n(\theta)$ la matrice de covariance du vecteur (U_1, \dots, U_n) sous \mathbb{P}_θ :

$$\forall i, j \in \{1, \dots, n\}, \quad [\Gamma_n(\theta)]_{i,j} = \text{Cov}_\theta(U) := \mathbb{E}_\theta(U_i U_j) - \mathbb{E}_\theta(U_i) \mathbb{E}_\theta(U_j) .$$

Alors,

$$(U - \mathbb{E}_\theta(U))^T \Gamma_n^{-1}(\theta) (U - \mathbb{E}_\theta(U)) = \gamma(\theta, U) = \sum_{i=1}^n \frac{(U_i - U_{i,\theta}^*)^2}{v_{i-1}(\theta)} ,$$

où $v_{i-1}(\theta) = \mathbb{E}_\theta[(U_i - U_{i,\theta}^*)^2]$. De plus,

$$\det(\Gamma_n(\theta)) = \prod_{i=1}^n v_{i-1}(\theta) .$$

Démonstration : Sans perte de généralité, on peut supposer que les U_i sont centrés. Remarquons d'abord que les innovations $(\nu_{i,\theta})_{i=1,\dots,n}$ sont décorréliées entre elles. Ainsi, leur matrice de covariance est la matrice diagonale D telle que $D_{ii} = v_{i-1}(\theta)$. Par ailleurs, $U_{i,\theta}^*$ est une fonction linéaire de U_1, \dots, U_{i-1} , et donc il existe une matrice C triangulaire, dépendant de θ , avec des éléments diagonaux égaux à 1 et telle que $C\nu_\theta = U$ (en notant ν_θ le vecteur des innovations). Par conséquent :

$$\Gamma_n(\theta) = \text{Cov}_\theta(U) = \text{Cov}_\theta(C^{-1}\nu_\theta) = CDC^T ,$$

et ainsi,

$$U^T \Gamma_n^{-1}(\theta) U = U^T (C^T)^{-1} D^{-1} C^{-1} U = \nu_\theta^T D^{-1} \nu_\theta = \sum_{i=1}^n \frac{(U_i - U_{i,\theta}^*)^2}{v_{i-1}(\theta)} .$$

Et bien sûr, $\det(\Gamma_n(\theta)) = \det(C)^2 \det(D) = \prod_{i=1}^n v_{i-1}(\theta)$. □

L'intérêt de la Proposition 4.3 réside dans le fait que lorsque (U_1, \dots, U_n) est un vecteur de loi gaussienne \mathbb{P}_θ , la vraisemblance est exactement une fonction de $U^T \Gamma_n^{-1}(\theta) U$ et $\det(\Gamma_n(\theta))$. Cela permettra dans la section 5.6.6 d'estimer facilement par maximum de vraisemblance un ARMA gaussien.

4.2 Les modèles d'état

Les lettres en gras représentent des vecteurs-colonne.

Un modèle d'état est constitué de deux équations :

- une équation d'observation (ou de mesure) :

$$\mathbf{Y}_t = G_t \mathbf{X}_t + \mathbf{W}_t \quad t = 1, 2, \dots \quad (4.4)$$

- une équation d'état :

$$\mathbf{X}_{t+1} = F_t \mathbf{X}_t + \mathbf{V}_t \quad t = 1, 2, \dots \quad (4.5)$$

Dans ces équations :

\mathbf{Y}_t est un vecteur de dimension w

\mathbf{X}_t est un vecteur de dimension v

G_t est une séquence temporelle de matrices (w, v)

F_t est une séquence temporelle de matrices (v, v)

\mathbf{W}_t est une séquence temporelle de vecteurs aléatoires indépendants de dimension w (bruit de mesure), de moyenne nulle et de matrice de covariance $R_t(w, w)$

\mathbf{V}_t est une séquence temporelle de vecteurs aléatoires indépendants de dimension v (bruit d'état), de moyenne nulle et de matrice de covariance $Q_t(v, v)$

Les séquences \mathbf{W}_t et \mathbf{V}_t sont indépendantes entre elles, et *décorrélées* avec X_1 .

L'équation (1.3) traduit l'évolution au cours du temps du vecteur \mathbf{X} , de taille v , F_1, \dots est une séquence connue de matrices (v, v) , \mathbf{V}_t est une séquence de vecteurs aléatoires de taille v .

L'équation (1.4) traduit la mesure (ou observation) de \mathbf{X} ; \mathbf{Y} est de taille w , G_t est une matrice (w, v) , \mathbf{W}_t est le bruit de mesure à l'instant t (vecteur colonne de taille w).

Dans de nombreux cas pratiques, les matrices F_t , G_t , Q_t et R_t sont indépendantes du temps, et sont alors notées respectivement F , G , Q et R .

Il existe des formes plus générales de ces équations :

- les séquences \mathbf{W}_t et \mathbf{V}_t peuvent être corrélées,
- un terme de contrôle $H_t \mathbf{u}_t$ peut venir s'ajouter dans le terme de droite de l'équation d'état, pour imposer un changement de \mathbf{X}_{t+1} .

Ces formes ne sont pas étudiées dans ce cours ; elles sont utilisées en théorie du contrôle et pour définir les notions de contrôlabilité (capacité de passer d'un état à un autre en un nombre fini d'itérations) et d'observabilité (capacité de déterminer l'état initial à partir des mesures, en l'absence de bruit de mesure).

Par définition, une série temporelle (ou série chronologique) \mathbf{Y}_t possède une représentation par modèle d'état s'il existe pour \mathbf{Y}_t un modèle correspondant aux équations 4.4 et 4.5.

Si la série est *univariée* (cas des séries étudiées dans le cours), alors $w = 1$. La dimension v de \mathbf{X}_t est supérieure à 1 en général (\mathbf{X}_t est une série multivariée).

Une valeur de w strictement supérieure à 1 correspond à une série chronologique *multivariée*.

4.3 Stationnarité

Soient les équations d'observation et d'état suivantes :

$$\mathbf{Y}_t = G\mathbf{X}_t + \mathbf{W}_t \quad t \in \mathbb{Z} \quad (4.6)$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + \mathbf{V}_t \quad t \in \mathbb{Z} \quad (4.7)$$

avec $Q_t = Q$ et $R_t = R$.

Par rapport aux équations 4.4 et 4.5 :

- les matrices G , F , Q et R sont constantes,
- le temps est indicé par \mathbb{Z} .

L'équation d'état 4.7 est dite *stable* (ou *causale*) si toutes les valeurs propres de F sont de module strictement inférieur à 1, ou de façon équivalente si $\det(I - Fz) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| \leq 1$ (la matrice F est dite stable).

Alors l'équation d'état 4.7 a une solution stationnaire unique donnée par :

$$\mathbf{X}_t = \sum_{j=0}^{\infty} F^j \mathbf{V}_{t-j-1}$$

et la séquence d'observations correspondante \mathbf{Y}_t , donnée par :

$$\mathbf{Y}_t = \mathbf{W}_t + G \sum_{j=0}^{\infty} F^j \mathbf{V}_{t-j-1}$$

est également stationnaire.

Exercice 4.1 Calculer les valeurs propres de F dans les 3 cas de l'Exercice 1.5 page 10. Interpréter les résultats.

Exercice 4.2 Soit X_t la solution stationnaire de l'équation d'état (4.7) : $X_t = \sum_{j=0}^{\infty} F^j V_{t-j-1}$. On suppose que F n'est pas la matrice nulle. Trouver une autre solution (nécessairement non stationnaire) de l'équation (4.7).

On a vu sur des exemples qu'il est possible de représenter par des modèles d'état des séries "standard" (somme d'une tendance et d'un bruit). On sait également que, dans la réalité, une tendance peut elle-même fluctuer au cours du temps (ce qui conduit alors à utiliser des techniques de moyenne mobile de préférence à une régression globale pour estimer cette tendance). Le formalisme du modèle d'état permet, en agissant de façon ad hoc sur le bruit d'état, de modéliser de telles fluctuations ; par exemple, pour une tendance localement linéaire, on peut introduire un bruit d'état sur la composante dérivée de l'état. Lorsque l'on introduit un tel bruit d'état, il est recommandé d'analyser les propriétés statistiques du processus généré.

Exercice 4.3 A partir du premier exemple de l'Exercice 1.5 page 10, proposer une modification de la représentation par modèle d'état qui rend compte d'évolutions possibles de la tendance linéaire.

De la même façon, la représentation par modèles d'état permet de rendre compte de variations saisonnières ; là encore, il est possible de modéliser des fluctuations de ces variations en jouant sur le bruit d'état.

4.4 Le filtre de Kalman

Les formules récursives de Kalman ont pour objectif général de trouver des estimateurs linéaires optimaux (au sens du minimum de l'erreur quadratique défini au paragraphe 4.1), du vecteur d'état \mathbf{X}_k en fonction des observations $\mathbf{Y}_1, \mathbf{Y}_2, \dots$ et d'un vecteur aléatoire \mathbf{Y}_0 orthogonal à \mathbf{V}_t et \mathbf{W}_t pour tout $t \geq 1$ (on prend en général pour \mathbf{Y}_0 le vecteur $(11\dots 1)'$). On notera $P_t(\mathbf{X}_k)$ le meilleur prédicteur linéaire de \mathbf{X}_k en fonction de $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_t$.

Les définitions de prédiction, filtrage, lissage sont les suivantes :

- prédiction : estimer (au sens défini ci-dessus) \mathbf{X}_t en fonction de $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$
- filtrage : estimer \mathbf{X}_t en fonction de $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_t$
- lissage : estimer \mathbf{X}_t en fonction de $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_n$ avec $n > t$

Dans les trois cas, la récursion porte sur la prise en compte d'une observation \mathbf{Y} supplémentaire. Dans les cas de la prédiction et du filtrage, l'indice t du \mathbf{X}_t estimé augmente également à chaque itération ; dans le cas du lissage, t est fixé alors que n augmente.

Il existe différentes façons d'écrire les équations récursives de Kalman selon le but poursuivi. Nous décrirons ici seulement le système d'équations le plus couramment utilisé (filtrage de Kalman), qui inclut lui-même une phase de prédiction.

Soit $P_t(\mathbf{X}_k)$ le meilleur prédicteur linéaire de \mathbf{X}_k en fonction de $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_t$.

Le filtre de Kalman a donc pour objectif de calculer à chaque instant t la valeur de $P_t(\mathbf{X}_t)$, ainsi que la matrice de covariance de l'erreur commise sur cet estimateur.

Pour cela, le filtre opère à chaque itération t en plusieurs étapes :

- prédiction de l'état : calcul de $P_{t-1}(\mathbf{X}_t)$ et covariance associée $C_{t|t-1}$ de l'erreur de prédiction
- prédiction de la mesure : calcul de $P_{t-1}(\mathbf{Y}_t)$
- calcul du terme d'innovation, égal à l'écart entre la mesure \mathbf{Y}_t et sa valeur prédite, et de sa covariance S_t
- calcul du gain de filtrage K_t
- filtrage : calcul de $P_t(\mathbf{X}_t)$ comme combinaison linéaire de la prédiction et de l'innovation, et covariance associée $C_{t|t}$ de l'erreur d'estimation

Il est nécessaire également d'initialiser ce processus.

Le filtre de Kalman est décrit de façon détaillée dans les paragraphes suivants.

4.4.1 Equations du filtre

4.4.1.1 Initialisation

Celle-ci consiste à donner une estimée de \mathbf{X} à l'instant 0, soit $P_0(\mathbf{X}_0)$ ainsi que la matrice de covariance $C_{0|0}$. Ceci se fait à partir des informations a priori dont on dispose sur l'état à l'instant 0; en absence d'information a priori, $C_{0|0}$ sera pris très grand, ce qui a pour effet d'accorder un poids négligeable à la valeur initiale.

4.4.1.2 Formules récursives $t = 1, 2, \dots$

1. Prédiction de l'état

$$\begin{aligned} P_{t-1}(\mathbf{X}_t) &= F_{t-1}P_{t-1}(\mathbf{X}_{t-1}) \\ C_{t|t-1} &= F_{t-1}C_{t-1|t-1}F_{t-1}^T + Q_{t-1} \end{aligned}$$

2. Prédiction de la mesure

$$P_{t-1}(\mathbf{Y}_t) = G_t P_{t-1}(\mathbf{X}_t)$$

3. Calcul de l'innovation à l'instant t et de sa covariance S_t

$$\begin{aligned} \nu_t &= \mathbf{Y}_t - P_{t-1}(\mathbf{Y}_t) \\ S_t &= G_t C_{t|t-1} G_t^T + R_t \end{aligned}$$

4. Calcul du gain K_t

$$K_t = C_{t|t-1} G_t^T S_t^{-1}$$

5. Estimation de l'état courant, comme combinaison linéaire de la valeur prédite et de l'innovation

$$\begin{aligned} P_t(\mathbf{X}_t) &= P_{t-1}(\mathbf{X}_t) + K_t \nu_t \\ C_{t|t} &= C_{t|t-1} - K_t S_t K_t^T \end{aligned}$$

Nota 1

La dernière équation :

$$C_{t|t} = C_{t|t-1} - K_t S_t K_t^T$$

fait apparaître explicitement la diminution de la covariance de l'erreur d'estimation apportée par une nouvelle mesure. On peut l'écrire de façon équivalente :

$$C_{t|t} = (I - K_t G_t) C_{t|t-1} (I - K_t G_t)^T + K_t R_t K_t^T$$

issue de l'équation :

$$P_t(\mathbf{X}_t) = (I - K_t G_t) P_{t-1}(\mathbf{X}_t) + K_t \mathbf{Y}_t$$

Nota 2

Une mesure \mathbf{Y}_t de mauvaise qualité ($R_t \rightarrow \infty$) conduit à $P_t(\mathbf{X}_t) = P_{t-1}(\mathbf{X}_t)$ (la mesure \mathbf{Y}_t n'est pas prise en compte); une mesure \mathbf{Y}_t sans erreur ($R_t = 0$) conduit à $G_t P_t(\mathbf{X}_t) = Y_t$.

Ci-dessous, on donne la démonstration des formules récursives du filtre de Kalman.

Démonstration : Rappelons que V_{t-1} est décorrélé de X_{t-1} , et W_t est décorrélé de X_t . Ainsi, en utilisant la Proposition 4.2

$$P_{t-1}(\mathbf{X}_t) = P_{t-1}(F_{t-1}\mathbf{X}_t + V_t) = F_{t-1}P_{t-1}(\mathbf{X}_t) + P_{t-1}(V_t) = F_{t-1}P_{t-1}(\mathbf{X}_t) .$$

Et de même, $P_{t-1}(\mathbf{Y}_t) = G_t P_{t-1}(\mathbf{X}_t)$. Pour estimer \mathbf{X}_t à l'instant t , i.e en ayant observé $\mathbf{Y}_1, \dots, \mathbf{Y}_t$, on décompose la projection à l'aide du point (iii) de la Proposition 4.2 :

$$P(\mathbf{X}_t | \mathbf{Y}_0, \dots, \mathbf{Y}_t) = P(\mathbf{X}_t | \mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}, \nu_t) = P(\mathbf{X}_t | \mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}) + P(\mathbf{X}_t | \nu_t) .$$

C'est à dire $P_t(\mathbf{X}_t) = P_{t-1}(\mathbf{X}_t) + P(\mathbf{X}_t | \nu_t)$. En utilisant le point (ii) de la Proposition 4.2 :

$$P(\mathbf{X}_t | \nu_t) = \mathbb{E}(\mathbf{X}_t \nu_t^T) S_t^{-1} \nu_t ,$$

où $S_t = \mathbb{E}(\nu_t \nu_t^T)$ est la matrice de covariance des innovations à l'instant t . Posons $K_t = \mathbb{E}(\mathbf{X}_t \nu_t^T) S_t^{-1}$. Il reste à évaluer récursivement K_t . En écrivant $\nu_t = G_t(\mathbf{X}_t - P_{t-1}(\mathbf{X}_t)) + W_t$, et en notant que W_t est décorrélé de X_t ,

$$\mathbb{E}(\mathbf{X}_t \nu_t^T) = \mathbb{E}(\mathbf{X}_t (\mathbf{X}_t - P_{t-1}(\mathbf{X}_t))^T) G_t^T = C_{t|t-1} G_t^T ,$$

où $C_{t|t-1} = \mathbb{E}[(\mathbf{X}_t - P_{t-1}(\mathbf{X}_t))(\mathbf{X}_t - P_{t-1}(\mathbf{X}_t))^T]$. On a :

$$\begin{aligned} C_{t|t-1} &= \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^T] - \mathbb{E}[P_{t-1}(\mathbf{X}_t) P_{t-1}(\mathbf{X}_t)^T] , \\ &= F_{t-1} \mathbb{E}[\mathbf{X}_{t-1} \mathbf{X}_{t-1}^T] F_{t-1}^T + Q_{t-1} - F_{t-1} \mathbb{E}[P_{t-1}(\mathbf{X}_{t-1}) P_{t-1}(\mathbf{X}_{t-1})^T] F_{t-1}^T , \\ &= C_{t|t-1} + Q_{t-1} , \end{aligned}$$

où $C_{t|t} = \mathbb{E}[(\mathbf{X}_t - P_t(\mathbf{X}_t))(\mathbf{X}_t - P_t(\mathbf{X}_t))^T]$. De même,

$$\begin{aligned} S_t &= \mathbb{E}(\nu_t \nu_t^T) , \\ &= G_t \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^T] G_t^T + R_t - G_t \mathbb{E}[P_{t-1}(\mathbf{X}_{t-1}) P_{t-1}(\mathbf{X}_{t-1})^T] G_t^T , \\ &= G_t C_{t|t-1} G_t^T + R_t . \end{aligned}$$

Enfin,

$$\begin{aligned} C_{t|t} &= C_{t|t-1} - \mathbb{E}[(P_t(\mathbf{X}_t) - P_{t-1}(\mathbf{X}_t))(P_t(\mathbf{X}_t) - P_{t-1}(\mathbf{X}_t))^T] = C_{t|t-1} - \mathbb{E}[(K_t \nu_t)(K_t \nu_t)^T] , \\ &= C_{t|t-1} - K_t S_t K_t^T . \end{aligned}$$

□

4.4.2 Le filtre d'information

Les formules précédentes mènent en parallèle le calcul de l'estimateur et celui de sa covariance d'erreur.

Dans certains cas, seul le calcul de covariance est intéressant, en particulier si l'on s'intéresse seulement à la performance potentielle du filtre. Il n'est alors pas utile de calculer la covariance de l'innovation et le gain du filtre, seules les valeurs successives de C nous intéressent.

Soit I la matrice inverse de C , pour toutes les valeurs des indices; I est appelée matrice d'information, elle est homogène à l'inverse d'une covariance.

La formule simple suivante donne l'évolution de I :

$$I_{t|t} = I_{t|t-1} + G_t^T R_t^{-1} G_t$$

avec :

$$I_{t|t-1} = (F_{t-1}I_{t-1|t-1}F_{t-1}^T + Q_{t-1})^{-1}$$

qui se réduit à :

$$I_{t|t-1} = (F_{t-1}^T)^{-1}I_{t-1|t-1}F_{t-1}^{-1}$$

dans le cas où il n'y a pas de bruit d'état. Dans ce cas, on obtient une formule globale pour passer de $I_{t-1|t-1}$ à $I_{t|t}$:

$$I_{t|t} = (F_{t-1}^T)^{-1}I_{t-1|t-1}F_{t-1}^{-1} + G_t^T R_t^{-1} G_t$$

Cette formule se simplifie encore si l'état est constant, soit si la matrice de transition F se réduit à l'identité :

$$I_{t|t} = I_{t-1|t-1} + G_t^T R_t^{-1} G_t$$

Le terme $G_t^T R_t^{-1} G_t$ est l'information apportée par une nouvelle mesure \mathbf{Y}_t à la connaissance de \mathbf{X}_t , elle est d'autant plus grande que l'erreur de mesure est plus petite.

Cette formule, qui peut être vérifiée à partir du lemme d'inversion matricielle, exprime le caractère additif de l'information.

Le cas d'une initialisation sans connaissance a priori se traduit en prenant $I_{0|0} = (0)_{v,v}$.

La Figure 4.1 illustre le fonctionnement du filtre de Kalman : soit une série égale à la somme d'une tendance linéaire et d'un bruit blanc (premier cas de l'exercice 4.1). Cette série est représentée sur la figure de gauche. La figure de droite représente l'évolution au cours du temps de $P_t(\mathbf{Y}_t) = G_t P_t(\mathbf{X}_t)$ qui est dans ce cas une quantité scalaire (courbe continue en bleu), avec un intervalle de confiance à $\pm 1\sigma$ calculé à partir de sa variance $G_t C_{t|t} G_t^T$ (courbes pointillées en rouge).

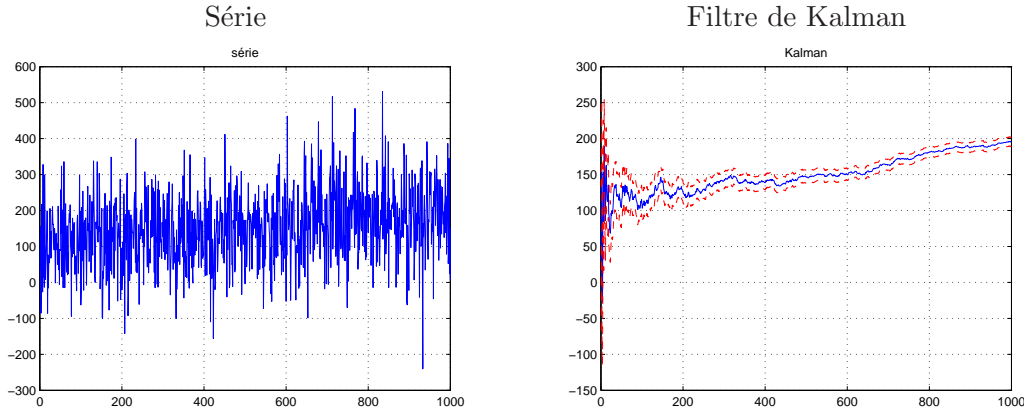


FIG. 4.1 – Un exemple de filtrage de Kalman

4.5 Observations manquantes

Le formalisme de Kalman est particulièrement bien adapté pour traiter cette situation.

Supposons qu'une observation particulière \mathbf{Y}_r ne soit pas accessible, on souhaite alors pouvoir répondre à l'une des deux questions suivantes (ou aux deux) :

- quelle est la meilleure estimée de l'état \mathbf{X} à l'instant r , et la covariance associée ?
- quelle est la meilleure estimée de l'observation manquante \mathbf{Y}_r , et la covariance associée ? (ceci correspond à une opération d'interpolation sur une série chronologique).

En absence de mesure \mathbf{Y}_r , il est visible que dans la suite des opérations de 1 à 5 décrites ci-dessus, seules les deux premières peuvent être effectuées.

On prendra donc dans ce cas :

$$\begin{aligned} P_r(\mathbf{X}_r) &= P_{r-1}(\mathbf{X}_r) = F_{r-1}P_{r-1}(\mathbf{X}_{r-1}) \\ C_{r|r} &= C_{r|r-1} = F_{r-1}C_{r-1|r-1}F_{r-1}^T + Q_{r-1} \\ P_r(\mathbf{Y}_r) &= G_rP_r(\mathbf{X}_r) \end{aligned}$$

$$\text{Cov}(P_r(\mathbf{Y}_r)) = G_rC_{r|r}G_r^T$$

Dans ces conditions, l'absence de mesure ne permet pas d'améliorer la précision d'estimation de l'état ; la présence du bruit d'état ne peut donc que venir dégrader cette précision.

L'opération décrite ci-dessus peut être répétée à chaque observation manquante.

4.6 Observations irrégulières

Le formalisme de Kalman permet encore de traiter ce cas de façon relativement aisée. On peut envisager deux méthodes différentes, dont les résultats sont équivalents.

Si les écarts entre les instants d'observation sont tous multiples d'une même période t_e , il est possible d'écrire les formules du filtre de Kalman pour cette période commune, et de traiter toutes les itérations où il n'y a pas de mesure avec la méthode des observations manquantes décrite ci-dessus.

On peut également ne traiter que les instants où une mesure est accessible. Cela signifie que les écarts de temps entre deux itérations successives ne sont pas égaux à une constante, il faut donc réécrire les formules du filtre de Kalman en tenant compte de cette particularité. Ceci est à prendre en compte dans l'équation d'état :

$$\mathbf{X}_{t+1} = F_t\mathbf{X}_t + \mathbf{V}_t$$

en adaptant les valeurs de F_t et de la covariance de \mathbf{V}_t (soit Q_t).

4.7 Filtre de Kalman étendu

Les modèles d'état et d'observation décrits ci-dessus, ainsi que les formules itératives du filtre de Kalman, sont linéaires.

Dans beaucoup d'applications, il arrive que soit l'équation d'état, soit l'équation d'observation, soit les deux, ne soient pas linéaires ; on a alors :

- une équation d'observation (ou de mesure) :

$$\mathbf{Y}_t = g_t(\mathbf{X}_t) + \mathbf{W}_t \quad t = 1, 2, \dots$$

- une équation d'état :

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t) + \mathbf{V}_t \quad t = 1, 2, \dots$$

où les fonctions f_t et/ou g_t ne sont pas linéaires.

Le filtre de Kalman étendu (ou non linéaire) reprend le même canevas que le filtre de Kalman linéaire, en 'linéarisant' les équations précédentes autour des estimées courantes. Son bon fonctionnement suppose donc que les estimées courantes sont toujours assez proches des valeurs

exactes, ce qui est loin d'être garanti dans tous les cas.

Les deux principaux problèmes que l'on peut rencontrer avec le filtre de Kalman étendu sont :

- la sensibilité à une initialisation de mauvaise qualité,
- une instabilité qui se traduit par une divergence, même dans les cas d'initialisation correcte.

Le filtre de Kalman étendu doit donc être utilisé avec de grandes précautions.

La linéarisation des équations d'état s'obtient :

- pour la manipulation du vecteur d'état ou de la mesure, en remplaçant G_t et F_t respectivement par g_t et f_t ,
- pour la manipulation des matrices de covariance, en remplaçant G_t et F_t respectivement par les matrices jacobienues des fonctions g et f (qui sont des matrices de mêmes dimensions que les matrices G et F), prises pour la valeur de \mathbf{X} qui est le meilleur estimateur de \mathbf{X}_t à l'instant où cette opération est réalisée.

Exercice 4.4 Equation d'observation non linéaire. Reprendre le premier exemple de l'Exercice 1.5 page 10. Dans cet exemple, l'état est de dimension 2 (la valeur courante d'une quantité et sa dérivée, supposée constante), et l'observation est égale à la valeur courante de la quantité plus un bruit de mesure blanc.

Supposer maintenant que, pour le même état, l'observation soit égale au carré de la valeur courante plus un bruit de mesure blanc. Ecrire les équations du filtre de Kalman étendu correspondant à ce cas.

Chapitre 5

Les processus ARMA

Les modèles *ARMA* permettent de représenter un grand nombre de processus aléatoires stationnaires. Ce chapitre détaille les principales méthodes de prédiction et d'identification de ces processus. Il est supposé à ce niveau que l'on a isolé, dans la série chronologique d'origine, sa partie aléatoire, en lui retirant tendances et facteurs saisonniers (chapitre 3).

On ne décrit pas ici les techniques liées aux modèles *ARMAX*, utilisées en particulier dans le domaine de l'automatique, où la séquence d'entrée (notée Z dans ce chapitre) n'est pas en général un bruit blanc.

5.1 Définitions et importance pratique

Ce paragraphe contient essentiellement des rappels, plus quelques notions supplémentaires.

Définition 5.1 *Le processus X_t , où $t \in \mathbb{Z}$, est un $ARMA(p, q)$ de moyenne nulle si :*

$$\begin{cases} (X_t) \text{ est stationnaire,} \\ X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \end{cases}$$

où Z_t est un bruit blanc faible, c'est-à-dire ici un processus aléatoire stationnaire de moyenne nulle avec $\gamma_Z(0) = \sigma^2$ et $\gamma_Z(h) = 0$ pour tout $h \neq 0$.

Dans le cadre de ce cours, Z_t sera par défaut un *bruit blanc fort*, c'est-à-dire que la séquence Z_t est de plus *i.i.d.*. Dans le chapitre consacré aux processus *ARMA* / *(G)ARCH*, les propriétés de Z (appelé ϵ) sont spécifiques.

Une représentation équivalente et plus concise est la suivante :

$$\Phi(B)X_t = \Theta(B)Z_t,$$

où Φ et Θ sont des polynômes de degrés respectifs p et q :

$$\begin{aligned} \Phi(z) &= 1 - \phi_1 z - \dots - \phi_p z^p, \\ \Theta(z) &= 1 + \theta_1 z + \dots + \theta_q z^q, \end{aligned}$$

et B est l'opérateur "retard" défini par :

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

L'équation qui lie X à Z est linéaire et à coefficients constants.

Ce modèle est donc linéaire au sens où, si $Z_t = Z_{1,t} + Z_{2,t}$, alors $X_t = X_{1,t} + X_{2,t}$ pour tout t , avec $\Phi(B)X_{1,t} = \Theta(B)Z_{1,t}$ et $\Phi(B)X_{2,t} = \Theta(B)Z_{2,t}$.

Le processus X_t est un $ARMA(p, q)$ de moyenne m si le processus $X_t - m$ est un $ARMA(p, q)$ de moyenne nulle.

Proposition 5.2 (fonction de covariance) *Pour tout processus aléatoire stationnaire dont la fonction de corrélation $\gamma(\cdot)$ tend vers 0 à l'infini, et pour tout entier $k > 0$, il est possible de trouver un processus ARMA dont la fonction de corrélation est égale à celle du processus aléatoire jusqu'à l'ordre k .*

Cette propriété est l'une des raisons de l'intérêt pratique des modèles ARMA; une autre raison est leur facilité de synthèse.

Proposition 5.3 (mémoire courte) *Pour un processus ARMA, la fonction de corrélation est bornée géométriquement :*

$$|\rho(k)| \leq Cr^k \quad k = 1, 2, \dots \quad C > 0 \quad 0 < r < 1$$

on dit qu'un processus ARMA est à mémoire courte (décroissance "rapide" de ρ).

Des éléments plus précis sur ce point sont donnés dans le chapitre 7 consacré aux processus à mémoire longue.

Proposition 5.4 (symétrie temporelle) *Lorsque l'on observe un processus ARMA en 'retournant' le temps, on obtient un nouveau processus qui a les mêmes propriétés statistiques; la représentation ARMA ne permet donc pas de modéliser un processus dont les décroissances sont en moyenne plus rapides que les croissances (cas des indices boursiers par exemple).*

Proposition 5.5 (ergodicité) *Un processus ARMA est un processus aléatoire stationnaire ergodique (voir [8] page 45), ce qui entraîne que pour toute application f de $\mathbb{R}^{\mathbb{Z}}$ dans \mathbb{R} mesurable et telle que $\mathbb{E}|f(X)| < \infty$, alors :*

$$\frac{1}{n} \sum_{k=1}^n f \circ B^k(X_t) = \frac{1}{n} \sum_{k=1}^n f(X_{t-k}) \rightarrow \mathbb{E}f(X) \quad \text{quand } n \rightarrow \infty \text{ et pour tout } t$$

Cette propriété justifie l'évaluation des caractéristiques du processus à partir d'une réalisation.

X_t est un processus MA d'ordre q si $\Phi(z) \equiv 1$, soit :

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

X_t est un processus AR d'ordre p si $\Theta(z) \equiv 1$, soit :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t$$

X_t est un processus MA d'ordre infini, $MA(\infty)$, si :

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

avec :

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

Un processus ARMA est dit *causal* s'il peut être écrit sous la forme d'un $MA(\infty)$, soit :

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

(X_t ne dépend que de Z_t et des valeurs *précédentes* de Z).

La causalité est une propriété non pas du processus lui-même, mais de la relation entre X et Z .

Exercice 5.1 Soit X le processus défini par :

$$X_t - \phi X_{t-1} = Z_t ,$$

où $|\phi| < 1$ et Z est un processus aléatoire stationnaire blanc de moyenne nulle et d'écart-type σ . Montrer que la seule solution stationnaire de cette équation (processus $AR(1)$) est causale par rapport à Z .

Si ce processus $ARMA$ causal est défini par $\Phi(B)X_t = \Theta(B)Z_t$, alors son écriture sous la forme d'un $MA(\infty)$ est :

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)} \quad \text{pour tout } z \text{ tel que } |z| \leq 1 \quad (5.1)$$

Si les polynômes $\Phi(\cdot)$ et $\Theta(\cdot)$ n'ont pas de zéros communs, le processus est causal si et seulement si tous les zéros de Φ ont un module strictement supérieur à 1.

Exercice 5.2 Soit (X_t) un processus stationnaire solution de l'équation :

$$X_t - \phi X_{t-1} = U_t , \text{ cordlnul} \quad (5.2)$$

avec $|\phi| > 1$ et (U_t) un bruit blanc faible de variance σ^2 .

1. Donner une expression de X_t en fonction de (U_t) . La représentation ainsi obtenue est-elle causale ?
2. Déterminer $\gamma_X(0)$ et $\gamma_X(1)$. Déterminer $\gamma_X(k)$ en fonction de $\gamma_X(k-1)$ pour tout $k \geq 1$.
3. Montrer que $V_t = X_t - \frac{1}{\phi} X_{t-1}$ est un bruit blanc faible de variance $\frac{\sigma^2}{\phi^2}$.

Un processus $ARMA$ défini par $\Phi(B)X_t = \Theta(B)Z_t$ est dit *inversible* si l'on peut écrire :

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

avec :

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

Si les polynômes $\Phi(\cdot)$ et $\Theta(\cdot)$ n'ont pas de zéros communs, le processus est inversible si et seulement si tous les zéros de Θ ont un module strictement supérieur à 1 et les π_j sont donnés par :

$$\sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)} \quad \text{pour tout } z \text{ tel que } |z| \leq 1 \quad (5.3)$$

Exercice 5.3 En reprenant l'équation 5.3, donner les valeurs des π_j pour un $ARMA(1,1)$.

5.2 Covariance, corrélation, corrélation partielle

Les techniques de base (dites "robustes") d'identification d'un processus $ARMA$ sont basées sur la connaissance de la fonction de covariance du processus. Il est donc important, pour étudier les processus $ARMA$, d'établir les liens entre leurs paramètres de définition (coefficients de Φ et Θ , variance de Z) et leur fonction de covariance.

5.2.1 Formulation générale de la fonction de covariance d'un processus ARMA causal

La façon la plus générale d'exprimer la fonction de covariance d'un processus ARMA causal consiste à utiliser sa représentation par un processus $MA(\infty)$: $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$.

La fonction de covariance $\gamma(k)$ de ce processus est donnée par :

$$\gamma_X(k) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} \quad (5.4)$$

σ_Z^2 étant la variance du bruit blanc Z_t .

Exercice 5.4 Démontrer la relation (5.4). Que devient cette relation pour un processus $MA(q)$? (distinguer selon que $|k|$ est inférieur ou supérieur à q). Faire l'application pour le processus $MA(2)$ défini par :

$$X_t = Z_t + 1.5Z_{t-1} - Z_{t-2} .$$

Quelle est la fonction de covariance d'un processus $AR(1)$?

5.2.2 Fonction de corrélation partielle

La fonction de corrélation partielle $\alpha(\cdot)$ d'un processus stationnaire de moyenne nulle est définie de la façon suivante : considérons le processus X aux instants $t, \dots, t+k$, et soit Y_t (respectivement Y_{t+k}) la projection de X_t (respectivement X_{t+k}) sur $X_{t+1}, \dots, X_{t+k-1}$. Alors on a par définition :

$$\alpha(k) = \text{Corr}[(X_t - Y_t), (X_{t+k} - Y_{t+k})]$$

La corrélation partielle entre deux échantillons est donc la corrélation entre les deux erreurs de projection aux instants t et $t+k$, pour des projections qui prennent en compte l'ensemble des échantillons intermédiaires. C'est la corrélation qui reste entre X_t et X_{t+k} , après avoir retranché de ces deux quantités la partie 'expliquée' par les échantillons intermédiaires X_{t+1} à X_{t+k-1} .

Par construction, on a : $\alpha(1) = \rho(1)$.

Proposition 5.6 $\alpha(k)$ est égal au coefficient $\phi_{k,k}$ que l'on applique à X_{t-k} quand on projette X_t sur $X_{t-k}, X_{t-k+1}, \dots, X_{t-1}$ selon la formule :

$$\widehat{X}_t = \sum_{j=1}^k \phi_{k,j} X_{t-j}$$

équivalente à la prédiction linéaire à un pas à erreur quadratique minimale (voir paragraphe 4.1.1 et algorithme de Durbin-Levinson).

Proposition 5.7 Pour tout processus stationnaire, on a :

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

Proposition 5.8 (acf d'un $MA(1)$) Pour un processus $MA(1)$, on a pour tout k :

$$\alpha(k) = (-\theta)^k \frac{\theta^2 - 1}{1 - \theta^{2(k+1)}}$$

Proposition 5.9 (pacf d'un AR) Pour un $AR(p)$, $\alpha(k) = 0$ pour $k > p$ (la fonction de corrélation partielle est à horizon fini)

En combinant l'exercice 5.4 et la Proposition 5.9, on en déduit la proposition suivante, *très utile pour l'identification des processus ARMA* :

Proposition 5.10 *Pour un processus MA, la fonction de covariance (ou de corrélation) est à horizon fini.*

Pour un processus AR, la fonction de corrélation partielle est à horizon fini.

Ces propriétés sont utilisées en particulier pour des analyses préliminaires, du type :

- le processus $ARMA$ est-il un processus MA , et si oui de quel ordre ?
- le processus $ARMA$ est-il un processus AR , et si oui de quel ordre ?

Illustrons ces propriétés :

La Figure 5.1 montre respectivement une réalisation d'un processus $AR(2)$, sa fonction de corrélation empirique et sa fonction de corrélation partielle empirique. La Figure 5.2 montre

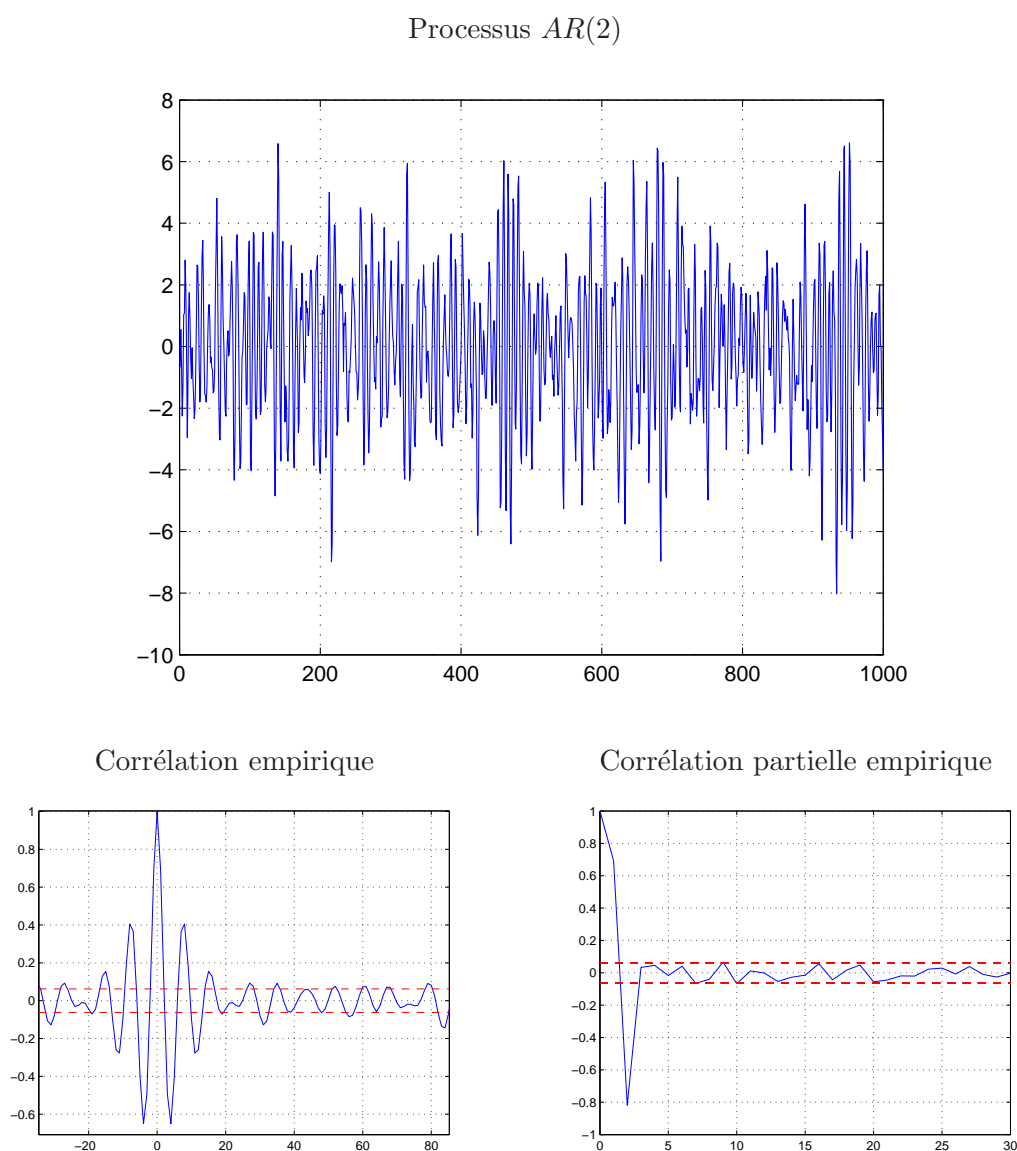


FIG. 5.1 – Propriétés d'un processus $AR(2)$

respectivement une réalisation d'un processus $MA(2)$, sa fonction de corrélation empirique et sa fonction de corrélation partielle empirique.

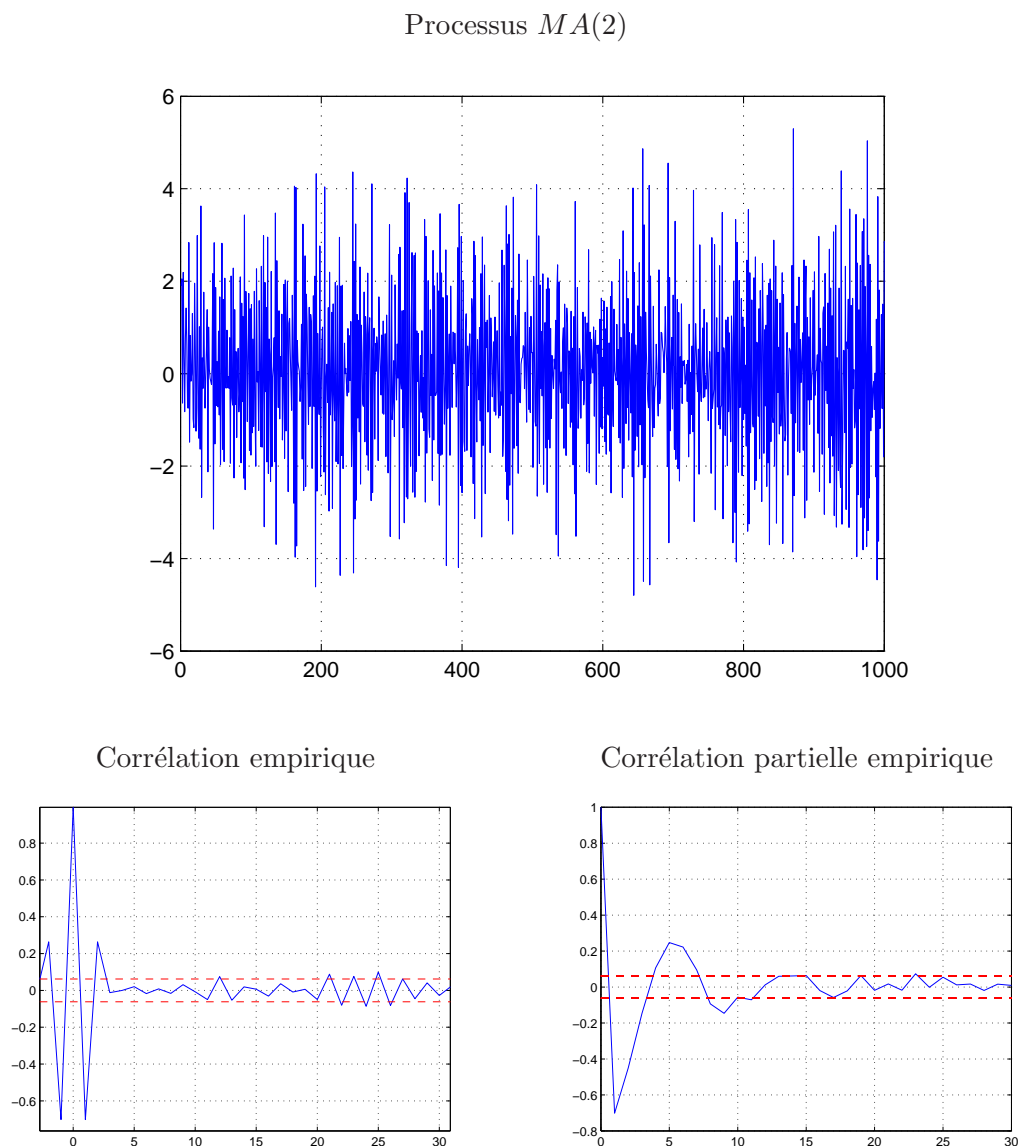


FIG. 5.2 – Propriétés d'un processus $MA(2)$

5.3 Modèle d'état pour un processus ARMA

Les processus $ARMA$ peuvent être représentés par des modèles d'état. On donne ici successivement la représentation par modèle d'état :

- d'un processus AR causal,
- d'un processus $ARMA$ causal.

5.3.1 Modèle d'état pour un $AR(p)$

Pour la cohérence avec les notations du chapitre 4 sur les modèles d'état, le processus $AR(p)$ considéré est noté Y (comme Y est ici un scalaire, il n'est pas noté en gras).

Si Y est un processus $AR(p)$, on a l'équation suivante :

$$Y_{t+1} = \phi_1 Y_t + \dots + \phi_p Y_{t-p+1} + Z_{t+1} \quad t \in \mathbb{Z}$$

Soit le vecteur d'état \mathbf{X} de taille p défini par :

$$\mathbf{X}_t = (Y_{t-p+1}, Y_{t-p+2}, \dots, Y_t)' \quad t \in \mathbb{Z}$$

alors on peut vérifier que Y_t est un processus $AR(p)$ causal si l'on prend pour matrices (constantes) G et F :

$$G = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix} \quad \text{de dimension } p$$

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & \dots & & & \\ 0 & 0 & 0 & \dots & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \dots & \phi_1 \end{bmatrix} \quad \text{de dimension } (p, p)$$

et si l'on pose $V_t = (0, \dots, Z_{t+1})'$ et $W_t = 0$.

Exercice 5.5 Considérer la matrice F du paragraphe 5.3.1 (représentation d'un processus AR par modèle d'état). Vérifier pour $p = 2$ que les valeurs propres de la matrice F sont de module inférieur à 1 si le processus est causal, c'est-à-dire si les zéros du polynôme

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

sont de module supérieur à 1.

5.3.2 Modèle d'état pour un ARMA(p,q)

Le processus $ARMA(p, q)$ considéré est encore noté Y .

Y satisfait l'équation suivante :

$$\Phi(B)Y_t = \Theta(B)Z_t \quad t \in \mathbb{Z}$$

On définit le processus U_t qui est un $AR(p)$ causal par :

$$\Phi(B)U_t = Z_t$$

soit :

$$Y_t = \Theta(B)U_t$$

Alors si l'on pose $r = \max(p, q + 1)$, on a :

$$Y_t = [\theta_{r-1} \quad \theta_{r-2} \quad \dots \quad \theta_0] \mathbf{X}_t$$

où \mathbf{X}_t est le vecteur de dimension r :

$$\mathbf{X}_t = (U_{t-r+1}, U_{t-r+2}, \dots, U_t)' \quad t \in \mathbb{Z}$$

On peut alors étendre les équations précédentes en prenant :

$$G = [\theta_{r-1} \quad \theta_{r-2} \quad \dots \quad \theta_0] \quad \text{de dimension } r$$

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & \dots & & & \\ 0 & 0 & 0 & \dots & 1 \\ \phi_r & \phi_{r-1} & \phi_{r-2} & \dots & \phi_1 \end{bmatrix} \quad \text{de dimension } (r, r)$$

et si l'on pose $V_t = (0, \dots, Z_{t+1})'$ et $W_t = 0$.

On voit que :

- G contient des zéros si $r - 1 > q$, c'est-à-dire si $p > q + 1$,
- la dernière ligne de la matrice F contient des zéros si $r > p$, c'est-à-dire si $q + 1 > p$

Cette représentation d'un processus *ARMA* par modèle d'état conduit à prendre un vecteur d'état de dimension $r = \max(p, q + 1)$.

Notons qu'il est possible d'avoir une représentation plus concise (de taille $\max(p, q)$), auquel cas W est non nul.

Cette représentation sera en particulier utilisée pour la prédiction, qui est une phase-clé de l'identification des processus.

Nota : il est nécessaire pour cela d'appliquer un filtrage de Kalman aux observations Y_t , $t = 1, \dots, n$ (voir initialisation et formules récursives au paragraphe 4.4.1). \mathbf{X}_t étant lui-même stationnaire de moyenne nulle, défini pour $t \in \mathbb{Z}$, mais observé à des instants t tels que $t = 1, \dots, n$ ($Y_t = G\mathbf{X}_t$), la partie initialisation doit être traitée de la façon suivante : prendre $P_0(\mathbf{X}_0) = (0)_2$ (vecteur nul de taille 2) et $C_{0|0} = \text{Cov}(\mathbf{X}_t) = E(\mathbf{X}_t \mathbf{X}_t')$ qui est indépendant de t , car \mathbf{X}_t est stationnaire. Comme $\mathbf{X}_t = (U_{t-r+1}, \dots, U_t)'$ où U_t est le processus $AR(p)$ causal défini par $\Phi(B)U_t = Z_t$, on obtient $C_{0|0}(i, j) = \gamma_U(|i - j|)$ pour $i, j = 1 \dots r$.

5.4 Densité spectrale d'un processus ARMA

La densité spectrale d'un processus *ARMA*(p, q) défini par l'équation :

$$\Phi(B)X_t = \Theta(B)Z_t$$

est donnée par :

$$P_X(\omega) = \frac{1}{2\pi} \sigma_Z^2 \frac{|\Theta(e^{-j\omega})|^2}{|\Phi(e^{-j\omega})|^2} \quad \omega \in [-\pi, \pi] \quad (5.5)$$

Exercice 5.6 En utilisant l'équation (5.5), calculer la densité spectrale d'un processus *AR*(1) et celle d'un processus *MA*(1).

La densité spectrale d'un processus autorégressif fait apparaître des 'pics' pour des valeurs de ω égales aux phases des zéros (complexes) de Φ . Chaque pic est d'autant plus étroit que le module du zéro correspondant de Φ est plus proche de 1.

La densité spectrale d'un processus *ARMA* étant égale au rapport de deux polynômes trigonométriques, on la qualifie de *densité spectrale rationnelle*.

5.5 Calcul de la covariance d'un processus ARMA

Ce paragraphe indique le calcul de la covariance d'un processus *ARMA* dont le modèle est *connu* (valeurs des coefficients ϕ et θ et de σ_Z^2). Il n'est utile que lorsque l'on utilise pour la prédiction la méthode alternative au filtrage de Kalman (voir ci-dessous).

5.5.1 Méthode directe

Cette méthode est basée sur les formules générales développées ci-dessus, (5.1) et (5.4), qui sont :

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)} \quad \text{pour tout } z \text{ tel que } |z| \leq 1$$

$$\gamma(k) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

La méthode directe consiste donc à calculer les coefficients ψ_j (formule (5.1)) de façon récursive et à en déduire $\gamma(k)$ (formule (5.4)).

Les formules récursives qui permettent de calculer les ψ_j sont les suivantes :

– si $p = 0$ (processus MA)

$$\psi_j = \theta_j \quad \text{pour } j = 1 \dots q \quad (5.6)$$

$$\psi_j = 0 \quad \text{pour } j \geq q + 1 \quad (5.7)$$

– si $p \geq 1$:

$$\psi_0 = 1 \quad (5.8)$$

$$\psi_j = \sum_{k=1}^{\min(p,j)} \phi_k \psi_{j-k} + \theta_j \quad \text{pour } 1 \leq j \leq q \quad (5.9)$$

$$\psi_j = \sum_{k=1}^{\min(p,j)} \phi_k \psi_{j-k} \quad \text{pour } j > q \quad (5.10)$$

Exercice 5.7 Démontrer les formules (5.6) à (5.10) à partir de l'équation (5.1).

Le calcul des $\gamma(k)$ à partir des ψ_j par la formule (5.4) faisant intervenir une somme infinie, cette méthode n'est intéressante que dans des cas particuliers où la totalité du calcul se fait de façon littérale.

Exercice 5.8 Calculer la représentation en $MA(\infty)$ d'un $ARMA(1, 1)$ causal. Puis calculer la fonction d'autocorrélation d'un $ARMA(1, 1)$.

5.5.2 Equation aux différences

A partir de la définition générale du processus $ARMA(p, q)$ causal :

$$\Phi(B)X_t = \Theta(B)Z_t$$

soit :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

on peut obtenir les équations suivantes, en multipliant cette équation par X_{t-k} pour les valeurs successives de k et en prenant les espérances mathématiques des deux termes de chaque égalité obtenue :

$$\begin{aligned} \gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) &= \sigma_Z^2 \sum_{j=k}^q \theta_j \psi_{j-k} \quad \text{pour } 0 \leq k \leq q \\ \gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) &= 0 \quad \text{pour } k \geq q+1 \end{aligned}$$

Ces équations supposent que l'on a préalablement déterminé les valeurs de ψ pour les indices de 0 à q .

Comme par ailleurs elles font appel à une récursivité d'ordre p , il faut procéder en deux temps :

- déterminer $\gamma(0), \dots, \gamma(p)$ par résolution du système linéaire constitué des $p + 1$ premières équations. Dans ces $p + 1$ premières équations, k varie de 0 à p , les indices de k à $k - p$ varient donc de $-p$ à p ; mais la propriété de symétrie de γ conduit bien à $p + 1$ équations,
- trouver les valeurs suivantes de γ de façon récursive.

Exercice 5.9 Vérifier que la solution trouvée à l'exercice 5.8 vérifie bien les équations aux différences.

Cette méthode est utilisée pour la détermination *numérique* de la covariance d'un processus *ARMA* causal dont on connaît les coefficients ϕ et θ et la variance σ_Z^2 .

On voit que globalement il est nécessaire, pour calculer la covariance d'un processus *ARMA* à partir de l'ensemble de ses paramètres supposés connus, d'appliquer *successivement* deux méthodes récursives, la première pour les valeurs des ψ , la deuxième pour les valeurs des γ .

5.6 Identification des processus ARMA

5.6.1 Définition

L'identification d'un processus *ARMA de moyenne nulle* est l'opération qui consiste à estimer au mieux les paramètres qui définissent ses propriétés *du deuxième ordre*, à savoir :

- les ordres p et q ,
- les p coefficients ϕ et les q coefficients θ ,
- l'écart-type de Z : σ_Z

Si le processus est gaussien, ses propriétés statistiques du deuxième ordre déterminent la totalité de ses propriétés statistiques.

Si l'on a affaire à un processus *ARMA* de moyenne m non nulle, il convient d'estimer préalablement m comme moyenne empirique des observations ($\hat{m} = \frac{1}{n} \sum_{t=1}^n X_t$), puis de travailler sur la série initiale à laquelle on a retiré cette moyenne empirique (par construction, la nouvelle série aura une moyenne *empirique* nulle).

L'estimation des paramètres une fois effectuée, il est nécessaire de compléter l'analyse en vérifiant que le processus considéré est bien un processus *ARMA* (avec les paramètres ainsi estimés).

5.6.2 Critères de présomption de processus ARMA

Il y a présomption de processus *ARMA* si les conditions suivantes sont satisfaites :

- le processus est stationnaire à l'analyse visuelle :
 - pas de tendance,
 - pas de saisonnalité,
 - variance constante.
- la fonction de corrélation empirique est :
 - à décroissance pas trop lente,
 - sans pics périodiques.

5.6.3 Principes d'identification

L'identification se fait selon un principe global, illustré par la figure 5.3. Décrivons ces étapes, en finissant par la plus délicate, celle de l'hypothèse sur p et q .

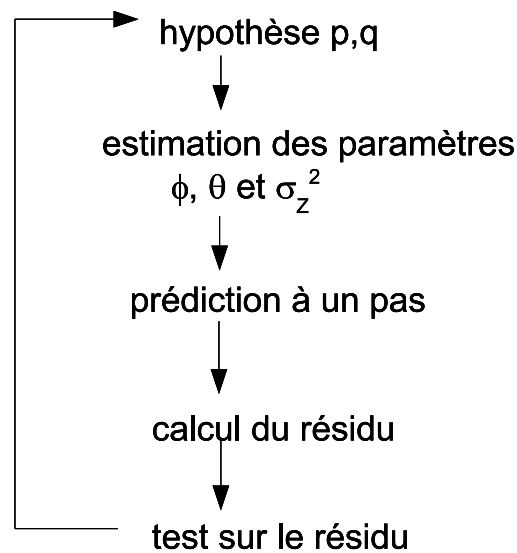


FIG. 5.3 – Principes généraux d'identification des processus ARMA

5.6.3.1 Estimation des paramètres ϕ , θ et σ_z^2

Il existe deux méthodes pour l'estimation des paramètres ϕ , θ et σ_z^2 :

- une méthode basée sur la fonction de covariance empirique, dite *robuste*, qui fait une estimation directe des paramètres ϕ , θ et σ_z^2 . Elle est décrite dans la section 5.6.5.
- une méthode d'estimation *paramétrique* basée sur le maximum de vraisemblance gaussien, applicable de façon optimale au cas des processus gaussiens. L'estimation des paramètres ϕ , θ et σ_z^2 résulte d'une itération qui inclut les calculs de la prédiction à un pas et des résidus. Cette méthode est décrite dans la section 5.6.6.

La deuxième méthode donne généralement des résultats plus précis que la première, au moins dans le cas des processus gaussiens, mais sa mise en oeuvre passe par la minimisation d'une fonction de plusieurs variables. Or il n'existe pas de résultats généraux sur la forme de cette fonction : convexité, existence de minima locaux, vitesse de variation ... si bien que le résultat obtenu dépend fortement du programme de minimisation utilisé.

Il est donc préférable de disposer de valeurs approchées des paramètres pour l'initialisation. Dans ce cas, on utilisera la procédure suivante : utiliser la méthode 1, puis :

- soit s'arrêter à ce stade en considérant que l'estimation est satisfaisante,
- soit passer à la méthode 2, qui utilise pour l'initialisation les résultats de la méthode 1, pour obtenir une estimation plus précise.

On peut également faire confiance au programme de minimisation utilisé et mettre en oeuvre directement la méthode 2 sans se soucier de l'initialisation (le logiciel R semble performant dans ces conditions).

5.6.3.2 Prédiction à un pas et calcul du résidu

On voit selon la figure 5.3 que la prédiction à un pas est une étape obligée du processus d'identification. Cette prédiction se fait de préférence par filtrage de Kalman, mais une méthode alternative existe, en utilisant l'algorithme de Durbin-Levinson ou celui des innovations, cf. section 5.6.5.4.

5.6.3.3 Test sur le résidu : validation du modèle

Au cours de la prédiction d'un processus *ARMA*, le résidu est défini comme la série temporelle des erreurs de prédiction standardisées. Lorsque le processus est estimé correctement, la série des résidus suit asymptotiquement la même loi que le processus générateur Z_t (bruit blanc faible en général). On peut donc tester la qualité de l'estimation en vérifiant, visuellement et à l'aide de tests statistiques, qu'il est vraisemblable que ces résidus proviennent d'un bruit blanc. Ces tests sont détaillés dans la section 5.6.7.

5.6.3.4 Hypothèse sur p et q

De nombreux travaux ont été consacrés à la détermination 'directe' des valeurs de p et q , voir par exemple [7] qui recense plusieurs méthodes. Citons la méthode du coin ("corner method") qui détermine p et q à partir de la corrélation empirique. Quelle que soit la méthode retenue, il est nécessaire de délimiter un sous-ensemble borné de $\mathbb{N} \times \mathbb{N}$ dans lequel chercher les couples (p, q) possibles.

Dans ce cours, on conseille d'employer la procédure suivante.

1. On trouve un ordre M_1 tel qu'un $MA(M_1)$ s'ajuste correctement à la série, et l'ordre M_2 tel qu'un $AR(M_2)$ s'ajuste correctement à la série. Cela revient à faire les hypothèses $(p, q) = (M_1, 0)$ et $(p, q) = (0, M_2)$ dans le schéma ci-dessus, et suppose donc que ces modèles auront passé l'étape de validation (test sur le résidu). Les détails sont donnés dans la section 5.6.4. On pose alors $M = \min\{M_1, M_2\}$.
2. On cherche ensuite à réduire le nombre des paramètres en trouvant un modèle $ARMA(p, q)$ s'ajustant correctement à la série et tel que $p + q < M$. Plusieurs options sont possibles, par exemple :
 - recherche en conservant les modèles passant le test de validation et ayant un nombre de paramètres minimal,
 - recherche en conservant le(s) modèle(s) minimisant un critère d'ajustement tel le critère d'Akaike (cf. section 5.6.8).

Ces recherches seront idéalement exhaustives dans l'ensemble des couples (p, q) tels que $p + q < M$, mais si le temps de calcul est prohibitif, on peut être amené à adopter une stratégie de recherche non exhaustive.

Il est à noter que quelle que soit la stratégie utilisée, le ou les modèles conservés devront faire l'objet d'une validation (cf. section 5.6.7). Notamment, le fait qu'un modèle minimise un critère d'ajustement ne dispense pas d'examiner les résidus.

5.6.4 Détermination de l'ordre d'un MA et d'un AR

5.6.4.1 Propriétés asymptotiques de l'ACF et de la PACF

Pour un processus aléatoire réel de moyenne nulle, la fonction de covariance empirique est définie par :

$$\gamma_e(h) = \frac{1}{n} \sum_{j=1}^{n-h} x_{j+h} x_j \quad \text{pour } 0 \leq h < n$$

On définit également la fonction de corrélation empirique, qui est la covariance normalisée par sa valeur en $h = 0$, soit :

$$\rho_e(h) = \frac{\gamma_e(h)}{\gamma_e(0)}$$

L'utilisation du terme n au dénominateur (au lieu de $n - h$) dans la définition de $\gamma_e(\cdot)$ garantit que :

- la matrice carrée Γ_e définie par :

$$\Gamma_e(i, j) = \gamma_e(i - j)$$

est définie non-négative,

- la fonction $\gamma_e(h)$ tend vers 0 lorsque h augmente.

En résumé, la fonction $\gamma_e(h)$ a les propriétés de la covariance d'un processus aléatoire stationnaire, quelle que soit la séquence x_t .

Sous certaines conditions (qui sont vérifiées en particulier par les processus *ARMA*), les $\gamma_e(h)$ pour $h = 0, \dots, n - 1$ sont des estimateurs de $\gamma(h)$ qui sont asymptotiquement sans biais et qui suivent asymptotiquement une loi de distribution conjointe normale dont la matrice de covariance décroît en $\frac{1}{n}$ (en particulier, la variance de chaque estimateur décroît en $\frac{1}{n}$) (formule de Bartlett).

Les mêmes propriétés s'appliquent à la fonction de corrélation empirique $\rho_e(h)$, pour laquelle s'applique le théorème suivant :

si x_i peut s'écrire sous la forme $x_i = \sum_{j=-\infty}^{\infty} \psi_j Z_{i-j}$ avec $\sum_j |\psi_j| < \infty$ et $E Z_t^4 < \infty$, alors pour tout $h \in \{1, 2, \dots\}$, $\vec{\rho}_e(h)$ est un vecteur asymptotiquement gaussien de moyenne $\vec{\rho}(h)$ et de matrice de covariance $n^{-1}W$, avec :

$$\begin{aligned} \vec{\rho}_e(h)' &= [\rho_e(1), \rho_e(2), \dots, \rho_e(h)] \\ \vec{\rho}(h)' &= [\rho(1), \rho(2), \dots, \rho(h)] \end{aligned}$$

et W est la matrice de covariance dont l'élément (i, j) est donné par la formule de Bartlett :

$$\begin{aligned} w_{i,j} &= \sum_{k=-\infty}^{\infty} \{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) \\ &\quad - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \} \end{aligned}$$

Les deux cas particuliers suivants de ce théorème sont utilisés dans la suite :

1. cas d'un bruit blanc :

- la fonction de corrélation vaut :

$$\begin{aligned} \rho(0) &= 1 \\ \rho(h) &= 0 \quad \text{pour } h \geq 1 \end{aligned}$$

- pour la fonction de corrélation empirique :

$$\begin{aligned} \rho_e(0) &= 1 \text{ (par construction)} \\ \text{pour } h \geq 1, \rho_e(h) &\text{ suit une loi normale de moyenne } 0 \\ &\text{et de variance } \frac{1}{n}. \end{aligned}$$

Il en résulte que la fonction de corrélation empirique, pour $h \geq 1$, est comprise entre $-1.96/\sqrt{n}$ et $1.96/\sqrt{n}$ avec une probabilité de 95% (test de "blancheur" de Bartlett).

2. cas d'un processus $MA(q)$:

- la fonction de corrélation vaut 0 pour $h > q$,
- pour $h > q$, la fonction de corrélation empirique suit une loi normale de moyenne 0 et de variance $\frac{1}{n}(1 + 2\rho^2(1) + \dots + 2\rho^2(q))$.

Cette propriété est utilisée pour une estimation préliminaire de l'ordre q .

Dans la pratique, la fonction de covariance empirique est utilisable pour $n \geq 50$ et $h \leq \frac{n}{4}$.

Exercice 5.10 Dans le cas d'un bruit blanc, calculer la moyenne de $\hat{\gamma}_n(0)$ et $\hat{\gamma}_n(1)$ en fonction de n . Quelle condition supplémentaire est nécessaire pour que $\text{Var}(\hat{\gamma}_n(0))$ existe ? Que vaut $\text{Var}(\hat{\gamma}_n(0))$ dans ce cas ?

Exercice 5.11 Soit X un processus $MA(q)$ de moyenne nulle défini à partir d'un bruit blanc fort possédant un moment d'ordre 4. Montrer que pour $h > q$, $\rho_X(h)$ est nul. Montrer aussi que pour $h > q$, $\hat{\rho}_n(h)$ suit asymptotiquement une loi normale de moyenne 0 et de variance $n^{-1}(1 + 2\rho_X^2(1) + \dots + 2\rho_X^2(q))$. On utilisera les notions et résultats suivants concernant le théorème central limite pour des variables dont la dépendance est à portée finie.

Définition 5.11 Soit m un entier naturel. Un processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire au sens strict est dit m -dépendant si pour tout t , les deux ensembles de variables aléatoires $\{X_j, j \leq t\}$ et $\{X_j, j \geq t + m + 1\}$ sont indépendants.

Théorème 5.12 Soit X un processus réel stationnaire au sens strict, m -dépendant et de moyenne μ . Alors,

$$\frac{\sum_{i=1}^n X_i - \mu}{\sqrt{\text{Var}(\sum_{i=1}^n X_i)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

5.6.4.2 Application à la détermination de l'ordre d'un MA

Une première estimation de l'ordre M d'un $MA(M)$ peut être faite en utilisant la propriété de la fonction de corrélation empirique énoncée ci-dessus :

1. la fonction de corrélation vaut 0 pour $h > M$
2. pour $h > M$, la fonction de corrélation empirique est asymptotiquement gaussienne de moyenne 0 et de variance $\frac{1}{n}(1 + 2\rho^2(1) + \dots + 2\rho^2(M))$.

Comme les valeurs de ρ ne sont pas connues, mais estimées, la procédure suivante est mise en oeuvre :

- calculer la fonction de corrélation empirique $\rho_e(h)$,
- pour m croissant, tracer les limites : $\pm l(m) = \pm \frac{1.96}{\sqrt{n}} \sqrt{1 + 2\rho_e^2(1) + \dots + 2\rho_e^2(m)}$
- on prendra pour estimateur de M la première valeur de m telle que $\rho_e(h)$ soit compris entre ces deux limites pour tout $h > m$ avec une probabilité de 95%. Notons M_1 cet estimateur.

La Figure 5.4 illustre cette procédure (cas d'un processus $MA(2)$). Pour confirmer que ce choix de M_1 est correct, il conviendra d'estimer les paramètres d'un $MA(M_1)$ par une des méthodes des sections 5.6.5 ou 5.6.6, et de valider le modèle à l'aide de la section 5.6.7. Si le modèle ne passe pas les tests de validation, on essaiera d'augmenter M_1 .

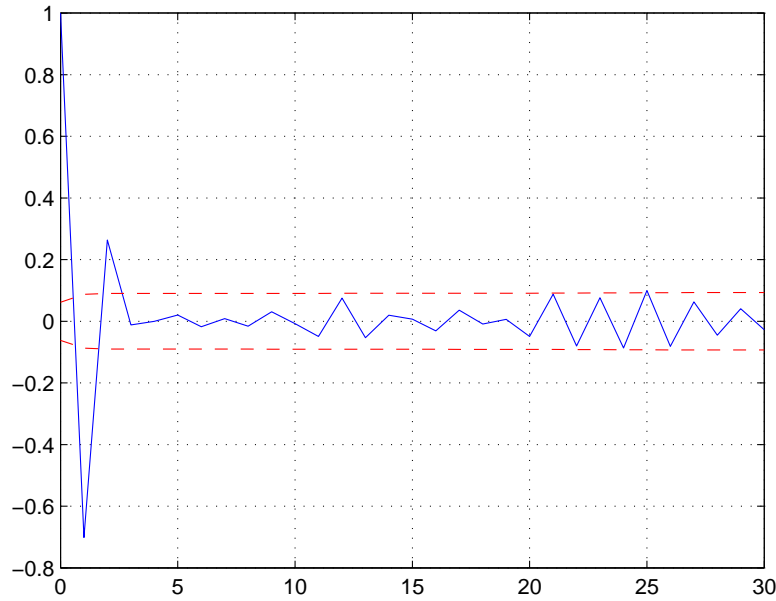
5.6.4.3 Application à la détermination de l'ordre d'un AR

Une première estimation de l'ordre M d'un $AR(M)$ peut être faite en utilisant la propriété de la fonction d'autocorrélation partielle empirique énoncée au corollaire 5.16 de l'exercice 5.11 :

1. la fonction d'autocorrélation partielle vaut 0 pour $h > M$
2. pour $h > M$, la fonction d'autocorrélation partielle empirique est asymptotiquement gaussienne de moyenne 0 et de variance $\frac{1}{n}$.

La procédure suivante est mise en oeuvre :

- calculer la fonction de corrélation empirique $\rho_e(h)$,

FIG. 5.4 – Intervalle de confiance pour un processus MA

- tracer les limites : $\pm \frac{1.96}{\sqrt{n}}$
- on prendra pour estimateur de M la première valeur de m telle que $\rho_e(h)$ soit compris entre ces deux limites pour tout $h > m$ avec une probabilité de 95%. Notons M_2 cet estimateur.

Pour confirmer que ce choix de M_2 est correct, il conviendra d'estimer les paramètres d'un $AR(M_2)$ par une des méthodes des sections 5.6.5 ou 5.6.6, et de valider le modèle à l'aide de la section 5.6.7. Si le modèle ne passe pas les tests de validation, on essaiera d'augmenter M_2 .

5.6.5 Identification “robuste” d'un processus ARMA

On suppose ici que l'hypothèse sur l'ordre (p, q) a déjà été faite (cf. section 5.6.3.4). La Figure 5.5 présente un schéma général pour l'estimation robuste d'un processus $ARMA$ causal, appelé méthode de Box-Jenkins. Dans la suite, on décrit en détail l'estimation des coefficients de l' $ARMA$ lorsque p et q sont imposés.

5.6.5.1 Cas d'un MA : l'algorithme des innovations

Cet algorithme permet d'estimer les coefficients θ d'un processus MA . Il peut être utilisé de plusieurs façons :

- cas d'un processus MA d'ordre connu : calcul des coefficients θ ,
- cas d'un processus $ARMA$, qu'on cherche dans un premier temps à identifier à un MA , comme décrit au début de ce paragraphe. Dans ce cas, les coefficients trouvés par l'algorithme des innovations sont les valeurs de ψ , puis on utilisera l'algorithme suivant pour en déduire ϕ et θ .

Détermination des valeurs de θ (ou ψ)

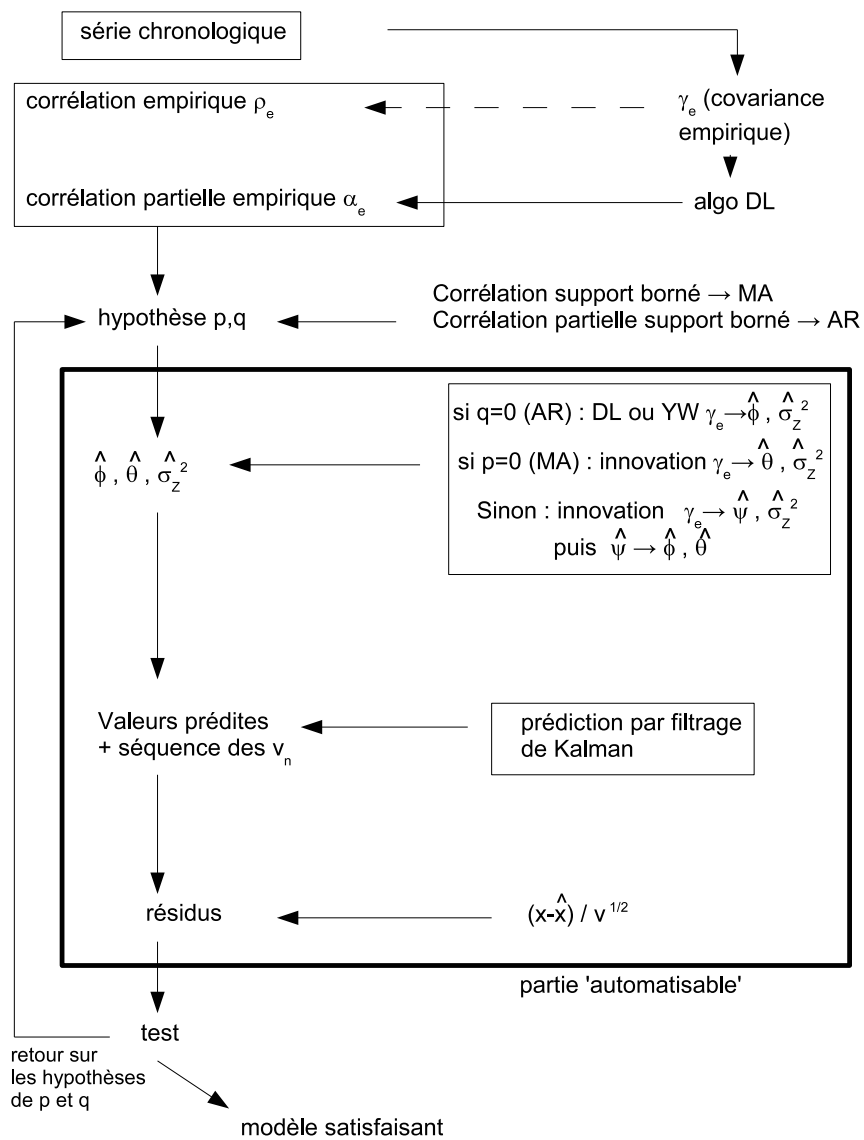


FIG. 5.5 – Méthode de Box-Jenkins

Initialisation

$$\begin{aligned}\hat{v}_0 &= \gamma_e(0) \\ \hat{\theta}_{1,1} &= \frac{1}{\hat{v}_0} \gamma_e(1) \\ \hat{v}_1 &= \gamma_e(0) - \hat{\theta}_{1,1}^2 \hat{v}_0\end{aligned}$$

Formules récursives pour $m \geq 2$:

$$\begin{aligned}\hat{\theta}_{m,m} &= \frac{1}{\hat{v}_0} \gamma_e(m) \\ \hat{\theta}_{m,m-k} &= \frac{1}{\hat{v}_k} \left[\gamma_e(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_j \right] \quad \forall k \in \{1, \dots, m-1\} \\ \hat{v}_m &= \gamma_e(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{v}_j\end{aligned}$$

Les estimées de θ pour l'itération m sont données par le vecteur :

$$\hat{\theta}_m = (\hat{\theta}_{m,1}, \dots, \hat{\theta}_{m,m})'$$

et l'estimée de σ_Z^2 pour l'itération m est donnée par :

$$\hat{\sigma}_Z^2 = \hat{v}_m$$

Cet algorithme doit fonctionner jusqu'à une valeur de m suffisante pour que les coefficients estimés se "stabilisent".

5.6.5.2 Cas d'un AR

L'algorithme de Durbin-Levinson Les entrées et sorties sont décrites ci-dessous selon l'utilisation de l'algorithme.

Initialisation

$$\begin{aligned}\phi_{1,1} &= \frac{\gamma(1)}{\gamma(0)} \\ v_1 &= \gamma(0)(1 - \phi_{1,1}^2)\end{aligned}$$

Formules récursives pour $m \geq 2$

$$\begin{aligned}\phi_{m,m} &= \frac{1}{v_{m-1}} [\gamma(m) - \sum_{j=1}^{m-1} \phi_{m-1,j} \gamma(m-j)] \\ \phi_{m,k} &= \phi_{m-1,k} - \phi_{m,m} \phi_{m-1,m-k} \quad k = 1, m-1 \\ v_m &= v_{m-1}(1 - \phi_{m,m}^2)\end{aligned}$$

Dans le processus d'identification, cet algorithme peut être utilisé à deux niveaux :

- calcul de la fonction de corrélation partielle empirique : l'entrée est la fonction de covariance empirique $\gamma_e(\cdot)$, la sortie est la séquence des $\phi_{m,m}$,
- estimation du processus s'il est censé être un AR : c'est le même algorithme, on utilise en plus la sortie v_m comme estimateur de la variance de Z . Dans ce cas, on trouve les coefficients de l'AR(p) dans la ligne n° p (par ordre croissant d'indice). Les lignes suivantes (à partir de $m = p + 1$) contiennent les mêmes coefficients pour les p premiers, puis des zéros jusqu'à m ; la valeur de v_m reste constante à partir de $m = p$.

On verra par la suite que, si l'on utilise une méthode alternative au filtre de Kalman pour la prédiction à un pas, l'algorithme de Durbin Levinson est également utilisé à ce niveau.

Algorithme de Yule Walker Cet algorithme permet de calculer les mêmes coefficients $\phi_{m,k}$ que l'algorithme de Durbin Levinson, mais de façon non récursive à chaque itération m .

Soient (à chaque itération m) :

- Φ_m le vecteur (à déterminer) composé des coefficients $\phi_{m,k}$ pour $k = 1, m$,
- Γ_m la matrice (m, m) définie par $\Gamma_m = [\gamma_e(i - j)]_{i,j=1}^m$
- γ_m le vecteur constitué des valeurs de la fonction de covariance de 1 à m : $\gamma_m = (\gamma_e(1), \dots, \gamma_e(m))'$

On a alors les relations suivantes :

$$\Gamma_m \Phi_m = \gamma_m \quad (5.11)$$

$$v_m = \gamma_e(0) - \Phi_m' \gamma_m \quad (5.12)$$

Exercice 5.12 Démontrer les relations 5.11 et 5.12 de l'algorithme de Yule Walker.

On voit que la détermination des coefficients $\phi_{m,k}$ nécessite d'inverser une matrice de dimension m . L'algorithme de Yule Walker est donc utilisé essentiellement pour déterminer les coefficients d'un processus auto-régressif d'ordre p connu ; le vecteur Φ_p contient précisément les coefficients cherchés.

5.6.5.3 Cas général d'un $ARMA(p, q)$

Un processus $ARMA$ causal peut être écrit sous la forme d'un MA d'ordre infini :

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

avec

$$\begin{aligned} \psi_0 &= 1 \\ \psi_j &= \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i} \end{aligned}$$

où par convention $\psi_0 = 1$. Les ψ_j sont en nombre infini, mais sont liés aux coefficients ϕ et θ , dont le nombre est $p + q$, donc la connaissance des $p + q$ premières valeurs de ψ (ψ_1 à ψ_{p+q}) entraîne en général celle de toutes les autres (ψ_{p+q+1} à ψ_{∞}) et surtout suffit pour calculer les coefficients ϕ et θ .

On suppose qu'on a déjà déterminé l'ordre M d'un processus MA d'ordre fini qui modélise correctement le processus $ARMA$, cf. section 5.6.4.2, mais cet ordre sera confirmé par l'algorithme des innovations. On a généralement $M \geq p + q$.

Pour la détermination des coefficients $\hat{\psi}_j$ et $\hat{\sigma}^2$, on pourra utiliser l'algorithme des 'innovations' présenté plus loin en calculant les $\theta_{m,m-k}$ suffisamment loin pour confirmer la valeur de l'ordre M qui satisfait les critères suivants : pour m suffisamment grand, les valeurs de $\hat{\theta}_{m,j}$ sont stables pour tout $j \leq M$ et petites pour $j > M$ et les valeurs de \hat{v}_m sont stables. On pose alors $\hat{\psi}_j$ = valeur stabilisée de $\hat{\theta}_{m,j}$ pour $j = 1, \dots, M$ et $\hat{\sigma}^2$ = valeur stabilisée de \hat{v}_m .

Il s'agit ensuite d'estimer ϕ et θ ($p + q$ valeurs) à partir des $\hat{\psi}_j$ pour $j = 1 \dots p + q$ (voir ci-dessous). On peut montrer que $\hat{\Phi}$ (le vecteur estimé des coefficients ϕ) satisfait l'équation matricielle :

$$\hat{\Psi} \hat{\Phi} = \hat{\psi} \quad (5.13)$$

où $\widehat{\Psi}$ est la matrice (p, p) dont les éléments sont :

$$\widehat{\Psi}(i, j) = \begin{cases} \widehat{\psi}_{q+i-j} & \text{si } q+i-j \geq 0 \\ \widehat{\Psi}(i, j) = 0 & \text{si } q+i-j < 0 \end{cases}$$

et $\widehat{\psi}$ est le vecteur colonne (p) :

$$\widehat{\psi} = (\widehat{\psi}_{q+1}, \dots, \widehat{\psi}_{q+p})'$$

Les valeurs de $\widehat{\theta}$ sont ensuite calculées au moyen des formules :

$$\widehat{\theta}_j = \widehat{\psi}_j - \sum_{i=1}^{\min(j,p)} \widehat{\phi}_i \widehat{\psi}_{j-i} \quad j = 1, \dots, q.$$

Exercice 5.13 Démontrer l'équation 5.13.

5.6.5.4 Détails pour la prédiction

Dans le filtre de Kalman, le résidu est égal à l'innovation divisée par son écart-type :

$$r_t = \frac{\nu_t}{\sqrt{S_t}}$$

S est un scalaire positif, puisque la série est univariée. Par ailleurs, l'initialisation du filtre de Kalman est décrite à la fin du paragraphe 5.3.2.

Dans le schéma de la Figure 5.5, on utilise un filtrage de Kalman pour la prédiction. Il existe une alternative présentée dans le schéma de la figure 5.6, où la prédiction se fait en calculant la fonction de covariance *estimée* $\widehat{\gamma}(\cdot)$ en appliquant l'équation aux différences aux coefficients $\widehat{\phi}$, $\widehat{\theta}$, $\widehat{\sigma}_Z^2$ et $\widehat{\psi}$, et en faisant la prédiction par l'algorithme de Durbin Levinson ou celui de Yule Walker ou encore l'algorithme des innovations. Avec cette alternative, on peut donc utiliser pour la prédiction :

- soit l'algorithme de Durbin Levinson (ou son équivalent non récursif l'algorithme de Yule Walker), auquel cas la fonction d'entrée est $\widehat{\gamma}(\cdot)$, la sortie est l'ensemble des coefficients $\phi_{m,k}$ et la variance de l'erreur de prédiction est donnée par v_m . La formule qui donne la valeur de X prédite à l'instant $n+1$ à partir des valeurs de X_1 à X_n est :

$$\widehat{X}_{n+1} = \sum_{j=1}^n \phi_{n,j} X_{n+1-j}$$

Exercice 5.14 Que donnent les formules de Durbin-Levinson dans le cas d'un $AR(1)$?

Propriété : pour un processus $AR(p)$ de coefficients ϕ_1, \dots, ϕ_p , la prédiction de X_{n+1} à partir de X_1 à X_n se réduit à :

$$\widehat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p} \quad \text{pour } n \geq p$$

- soit l'algorithme des innovations par la formule récursive suivante :

$$\widehat{X}_{n+1} = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \widehat{X}_{n+1-j})$$

qui nécessite de calculer les coefficients θ jusqu'à l'itération n .

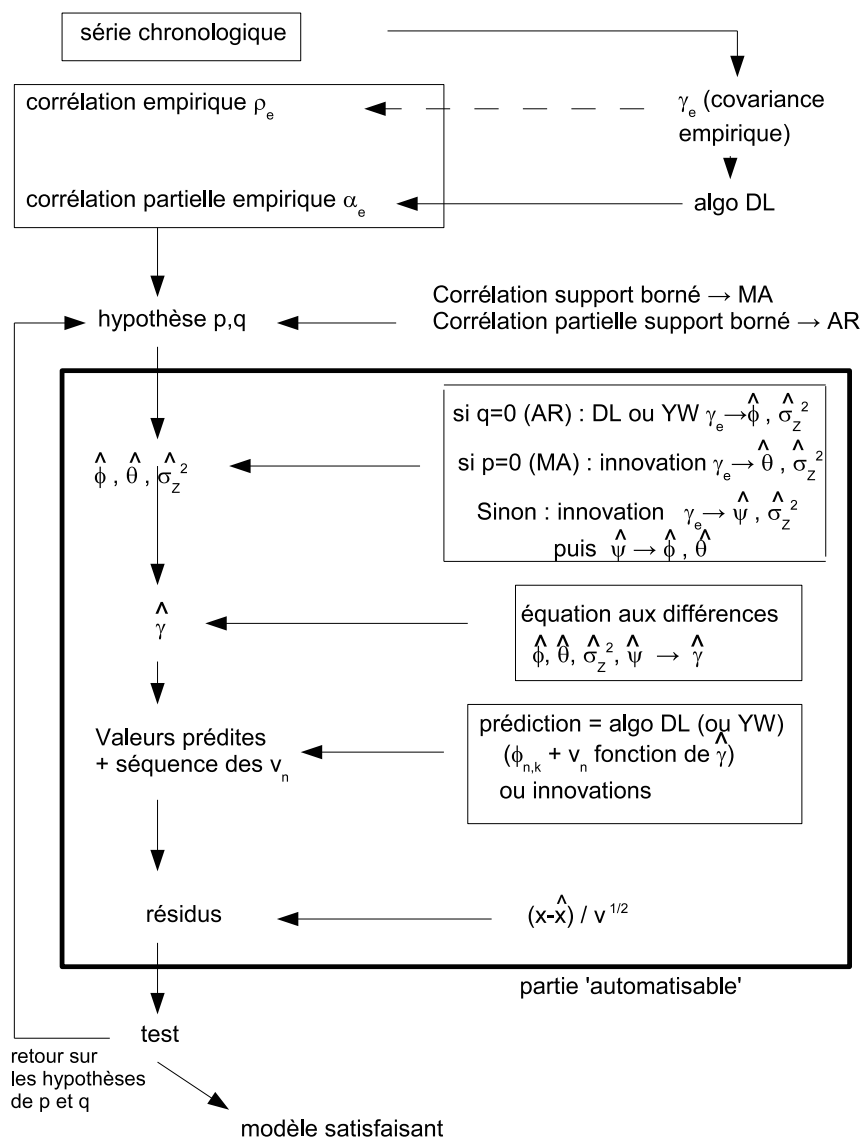


FIG. 5.6 – Méthode de Box-Jenkins, alternative pour la prédiction

L'utilisation de ces algorithmes pour la prédiction à h pas (en vue de prévision) est traitée en fin de chapitre.

Enfin, il existe une alternative à ce schéma : remplacer la phase de prédiction et de calcul des résidus par une reconstruction directe de Z à partir des *estimées* des polynômes Φ et Θ , ce qui suppose que le modèle estimé est inversible. Cette méthode est une approximation pour le calcul des résidus (voir [5] paragraphe 8.11).

5.6.6 Identification paramétrique d'un processus ARMA

La méthode exposée précédemment (identification "robuste", paragraphe 5.6.5) est basée entièrement sur la fonction de covariance empirique. Lorsque les ordres du modèle *ARMA* (p et q) sont supposés connus, l'estimation du processus consiste à trouver $r = p + q + 1$ paramètres (ϕ , θ et σ_Z^2). Une méthode directe d'estimation de ces paramètres consiste à trouver l'ensemble des r paramètres les plus vraisemblables, compte tenu des observations $(X_1 \dots X_n)$. Pour cela, la technique classique du maximum de vraisemblance est applicable : calculer la vraisemblance L (likelihood) des observations $(X_1 \dots X_n)$ en fonction des r paramètres (ϕ , θ et σ_Z^2) et trouver l'ensemble des paramètres qui maximise L .

Proposition 5.13 *La vraisemblance L d'un processus ARMA gaussien, de moyenne nulle, de paramètres $(\phi, \theta, \sigma_Z^2)$, la vraisemblance L est donnée par la formule suivante :*

$$L(\phi, \theta, \sigma_Z^2) = (2\pi\sigma_Z^2)^{-n/2} (r_0 \dots r_{n-1})^{-1/2} \exp\left(-\frac{1}{2\sigma_Z^2} \sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right)$$

où $\hat{X}_{j+1} = \hat{X}_{j+1}(\phi, \theta, \sigma_Z^2)$ est la prédiction de X_{j+1} sachant X_1, \dots, X_j , dans le modèle ARMA donné par les paramètres $(\phi, \theta, \sigma_Z^2)$, et les r_j sont les variances des erreurs de prédiction, divisées par la variance de Z :

$$r_j = r_j(\phi, \theta, \sigma_Z^2) = \frac{1}{\sigma_Z^2} v_j(\phi, \theta, \sigma_Z^2) \quad \text{où} \quad v_j(\phi, \theta, \sigma_Z^2) = E(X_{j+1} - \hat{X}_{j+1})^2.$$

La maximisation de L par rapport aux r paramètres conduit aux équations suivantes :

$$\hat{\sigma}_Z^2 = \frac{1}{n} S(\hat{\phi}, \hat{\theta}) \tag{5.14}$$

$$S(\hat{\phi}, \hat{\theta}) = \sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \tag{5.15}$$

où $\hat{\phi}$ et $\hat{\theta}$ sont les valeurs de ϕ et θ qui minimisent :

$$l(\phi, \theta) = \log\left(\frac{1}{n} S(\phi, \theta)\right) + \frac{1}{n} \sum_{j=1}^n \log r_{j-1}, \tag{5.16}$$

où "log" désigne le logarithme népérien. La fonction $l(\phi, \theta)$ est appelée *vraisemblance réduite*.

Démonstration : La forme de L est une conséquence directe de la Proposition 4.3. On peut voir sur les équations du filtre de Kalman, par exemple, que :

$$\hat{X}_{j+1}(\phi, \theta, \sigma_Z^2) = \hat{X}_{j+1}(\phi, \theta, 1) \quad \text{et} \quad r_j(\phi, \theta, \sigma_Z^2) = r_j(\phi, \theta, 1),$$

Ainsi, $\sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$ ne dépend pas de σ_Z^2 . En dérivant la log-vraisemblance en σ_Z^2 , on obtient :

$$\frac{\partial \log L}{\partial \sigma^2}(\phi, \theta, \sigma_Z^2) = -\frac{n}{2\sigma_Z^2} + \frac{1}{2\sigma_Z^4} \sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}.$$

Donc, à ϕ et θ fixé, le maximum est atteint pour $\sigma^2(\phi, \theta) = \frac{1}{n} S(\phi, \theta)$. Il reste à maximiser en ϕ, θ la fonction $\log(L(\phi, \theta, \sigma^2(\phi, \theta)))$, qui vaut :

$$\log(L(\phi, \theta, \sigma^2(\phi, \theta))) = \text{Constante} - \frac{n}{2} \log[\sigma^2(\phi, \theta)] - \frac{1}{2} \sum_{j=1}^n \log r_{j-1} = \text{Constante} - \frac{n}{2} l(\phi, \theta),$$

d'où le résultat. \square

Exercice 5.15 Dédurre des formules 5.14 à 5.16 les estimateurs de ϕ et σ^2 pour un processus AR(1).

Une alternative consiste à minimiser $S(\phi, \theta) = \sum_j \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$ par rapport à ϕ et θ (méthode des moindres carrés pondérés).

Ces formules montrent que le calcul de la vraisemblance nécessite ceux de la prédiction à un pas et des résidus, comme indiqué précédemment.

Le schéma d'identification est alors donné par la figure 5.7. Comme dans la méthode "robuste", la prédiction se fait de préférence par filtrage de Kalman, ou par la méthode alternative décrite au paragraphe 5.6.5.4.

Cette technique paramétrique donne de façon *optimale* (au sens du maximum de vraisemblance) les paramètres d'un modèle *ARMA gaussien* dont on connaît les ordres. Comme dans les techniques basées sur la fonction de covariance empirique, l'identification du modèle passe également par la recherche du couple (p, q) qui conduit au meilleur résultat.

La façon d'utiliser l'une ou l'autre des deux méthodes (méthode robuste" ou méthode paramétrique) ou les deux de façon complémentaire est décrite au paragraphe 5.6.3.

5.6.7 Validation d'un modèle : critères de contrôle sur le résidu

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

$$\widehat{W}_t = \frac{X_t - \widehat{X}_t(\widehat{\phi}, \widehat{\theta})}{\sqrt{r_{t-1}(\widehat{\phi}, \widehat{\theta})}}$$

Lorsque le processus est estimé correctement, la série des résidus suit asymptotiquement à peu près la même loi que le processus générateur Z_t (bruit blanc faible en général). On peut donc tester la qualité de l'estimation en observant la fonction de corrélation empirique $\widehat{\rho}(h)$ de \widehat{W}_t . Celle-ci vaut 1 pour $h = 0$, et pour $h \geq 1$, elle est comprise entre $\frac{-1.96}{\sqrt{n}}$ et $\frac{1.96}{\sqrt{n}}$ avec une probabilité de 95%. C'est le *test de "blancheur" de Bartlett*. Lorsque l'estimation du processus a été obtenue par maximisation de la vraisemblance, les intervalles de confiance peuvent être affinés (c'est-à-dire diminués) pour les h petits. L'application de ce critère permet de dire si le modèle *ARMA* choisi (avec ses ordres et ses paramètres) modélise correctement le processus. Il est nécessaire pour le calculer de mettre en oeuvre une prédiction du processus, comme décrit dans l'approche générale.

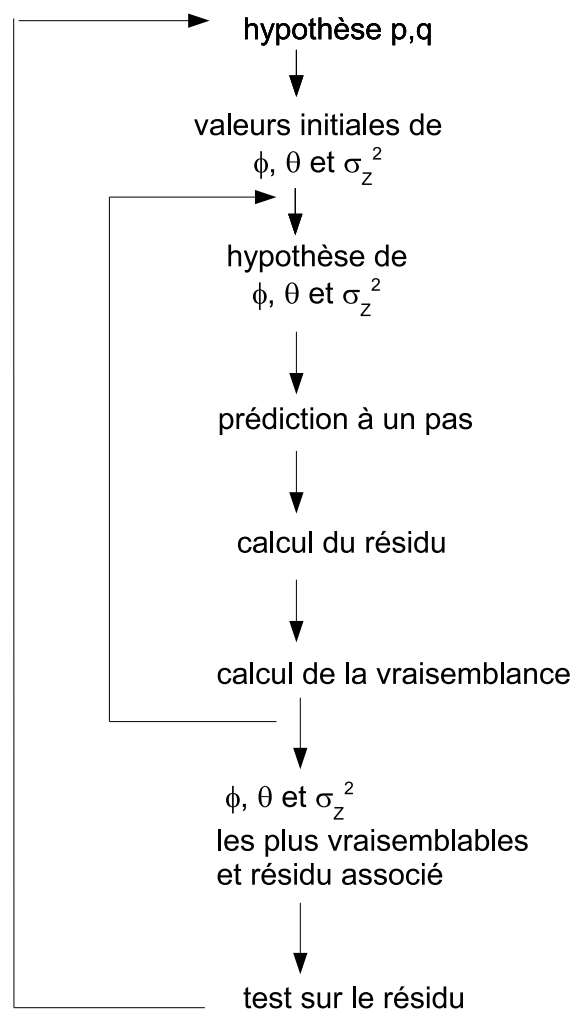


FIG. 5.7 – Schéma pour l'identification paramétrique

Les tests “portmanteau” (désigne une grosse valise en anglais, un “fourre-tout”) portent sur une statistique unique bâtie à partir de la corrélation empirique des résidus. On utilisera le *test de Ljung-Box*. Si on note $\hat{\rho}$ l'autocorrélation empirique des résidus de l'ajustement à un $ARMA(p, q)$, alors, pour h assez grand, mais petit devant n ,

$$Q = n(n+2) \sum_{j=1}^h \hat{\rho}^2(j)/(n-j) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(h - \text{df}),$$

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

5.6.8 Méthodes de sélection

On décrit ici des méthodes permettant de *comparer des modèles ARMA entre eux* :

- le critère AICC (Akaike Information Corrected Criterion) qui permet, par une méthode de pénalisation, d'éviter le phénomène d'overfitting,
- une méthode générale de sélection en “retenant” une partie des données pour choisir parmi plusieurs modèles.

5.6.8.1 Les critères AICC et AIC

Rien n'empêche théoriquement d'estimer un processus $ARMA(p, q)$ en prenant des ordres plus élevés ; mais ceci entraîne une augmentation de l'erreur d'estimation des paramètres, et un risque accru de trouver un modèle qui est trop proche de la réalisation (ou trajectoire) particulière observée (phénomène d'overfitting). Le critère AICC (Akaike Information Corrected Criterion) a pour objectif de trouver un compromis entre la vraisemblance des observations par rapport au modèle et la complexité de celui-ci. Il est donné par :

$$AICC(\phi, \theta) = -2 \log L(\phi, \theta, n^{-1}S(\phi, \theta)) + 2(p + q + 1) \frac{n}{n - p - q - 2}$$

où :

- $L(\phi, \theta, \sigma^2)$ est la vraisemblance des données pour un modèle $ARMA(p, q)$ gaussien avec les paramètres (ϕ, θ, σ^2) ,
- $S(\phi, \theta)$ est la somme des carrés des résidus.

Pour n donné, le deuxième terme de la formule qui donne AICC augmente avec la complexité du modèle, c'est-à-dire avec p et q ; c'est un terme de 'pénalisation'. On choisit alors l'ensemble $(p, q, \phi, \theta, \sigma^2)$ qui minimise ce critère. La sélection des ordres par minimisation de AICC est asymptotiquement efficace pour les modèles autorégressifs.

Le critère $AIC(\phi, \theta) = -2 \log L(\phi, \theta, n^{-1}S(\phi, \theta)) + 2(p + q + 1)$ est équivalent au critère AICC quand $n \rightarrow \infty$, mais a tendance à surestimer p pour les modèles autorégressifs.

Dans le cas d'un processus *AR en régime établi*, on a $r_j = 1$ pour tout j , soit $L(\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2}\sigma^2 S}$, où S est la somme des carrés des erreurs de prédiction. Alors $L(\phi, \theta, n^{-1}S) = (2\pi n^{-1}S)^{-n/2} e^{-\frac{n}{2}}$ et $-2 \log L = n \log \frac{2\pi}{n} + n + n \log S$. On cherchera donc à minimiser $n \log S + 2(p + q + 1) \frac{n}{n - p - q - 2}$ pour l'AICC ou $n \log S + 2(p + q + 1)$ pour l'AIC.

La figure 5.8 donne un exemple d'utilisation de ce critère pour un $AR(2)$. Certains processus $ARMA(p, q)$ ont de nombreux coefficients nuls, le nombre m de coefficients non nuls est donc significativement inférieur à $p + q$ (cas des processus saisonniers). Dans ce cas, le critère d'Akaike corrigé devient :

$$AICC(\phi, \theta) = -2 \log L(\phi, \theta, n^{-1}S(\phi, \theta)) + 2(m + 1) \frac{n}{n - m - 2}$$

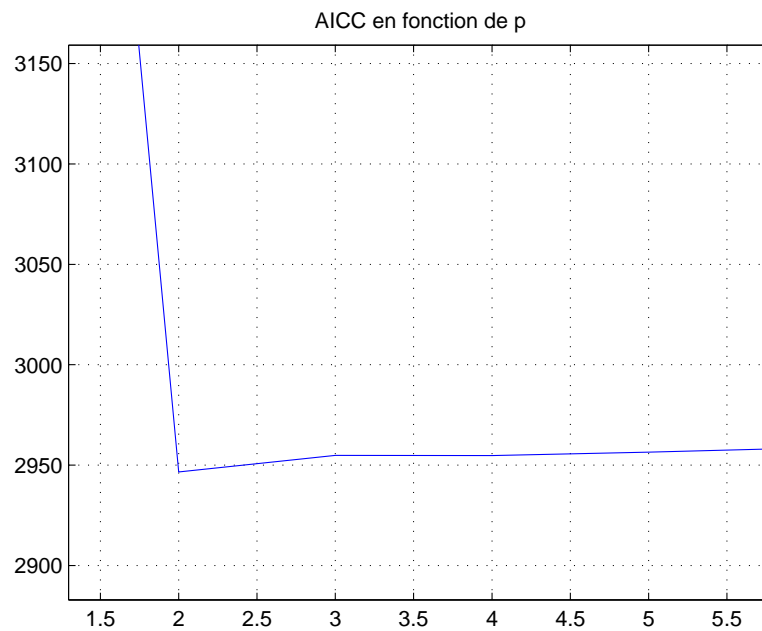


FIG. 5.8 – Le critère AICC appliqué à un AR(2)

Remarque : ce critère peut être utilisé comme base pour une recherche automatique des valeurs de p et q .

5.6.8.2 Sélection et évaluation sur une sous-partie des données

Lorsque la quantité de données le permet, on peut diviser la série temporelle en deux ou trois parties :

- La première est l'échantillon d'estimation, c'est sur celui-ci qu'on effectue la première phase d'estimation des paramètres pour un ordre (p, q) donné. Idéalement, on essaiera d'ajuster plusieurs modèles.
- La deuxième est l'échantillon de sélection, on l'utilisera pour discriminer différents modèles proposés dans la première phase (par exemple différents ordres), et n'en choisir qu'un seul, par exemple selon le critère du maximum de vraisemblance, ou sur un critère d'erreur de prédiction minimale.
- Eventuellement, une dernière partie est l'échantillon de test, qui sert à évaluer les performances du modèle choisi.

Quelle proportion garder dans chaque échantillon ? Il n'y a pas de réponse rigoureuse, mais un choix typique est 50% des valeurs pour l'estimation, et 25% pour la validation et le test.

5.7 Estimation spectrale des processus ARMA

Les processus *ARMA* sont un exemple de processus qui se modélisent entièrement grâce à un petit nombre de paramètres ($r = p + q + 1$), l'estimation paramétrique est donc une bonne façon d'estimer leur densité spectrale.

La méthode à utiliser est alors la suivante :

- estimation du processus par les méthodes temporelles étudiées précédemment (détermination des ordres p et q , calcul des valeurs des ϕ et θ et de σ_Z^2),

- calcul direct du spectre par utilisation des formules établies au paragraphe 5.4 (spectre dit 'rationnel').

Cette technique présente par rapport au périodogramme l'avantage de conduire à un spectre estimé qui est bien lissé, mais cependant affecté des erreurs de l'estimation temporelle.

5.8 Prédiction linéaire à h pas

Les techniques de prédiction à un pas étudiées précédemment (par filtrage de Kalman, ou comme combinaison linéaire des X_i ou par l'algorithme des innovations) s'étendent à la prédiction à h pas. Il s'agit de prédire la valeur de X_{n+h} à partir des valeurs de X_1, \dots, X_n . Quand h augmente pour n fixé, \hat{X}_{n+h} tend vers 0 et la variance de l'erreur d'estimation tend vers σ_X^2 .

5.8.1 Filtrage de Kalman

On utilisera une méthode analogue à celle qui a été présentée pour traiter le cas des données manquantes au chapitre 4, paragraphe 4.5 : appliquer les équations complètes de Kalman (opérations 1 à 5) jusqu'à l'instant n , puis seulement les deux premières opérations pour les instants de $n+1$ à $n+h$.

5.8.2 Combinaison linéaire des X_i

On cherche un prédicteur de la forme :

$$\hat{X}_{n+h} = \phi_{n,1}^{(h)} X_n + \dots + \phi_{n,n}^{(h)} X_1$$

Il faut donc trouver les valeurs des coefficients $\phi_{n,i}^{(h)}$ pour $i = 1, n$.

La résolution est analogue à celle qui a été décrite dans la méthode de Yule Walker (paragraphe 5.6.5.2).

Si l'on pose :

$$\begin{aligned} \Phi_n^{(h)} &= (\phi_{n,1}^{(h)}, \dots, \phi_{n,n}^{(h)})' \\ \gamma_n^{(h)} &= (\gamma_e(h), \dots, \gamma_e(n+h-1))' \end{aligned}$$

alors on a l'égalité suivante :

$$\Gamma_n \Phi_n^{(h)} = \gamma_n^{(h)}$$

avec $\Gamma_n(i, j) = [\gamma_e(i-j)]_{i,j=1}^n$ (Γ_n ne dépend pas de h).

5.8.3 La méthode des innovations

La formule suivante s'applique :

$$\hat{X}_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

5.9 Exercices complémentaires

Exercice 5.16 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus aléatoire tel que $\sup_t \mathbb{E}|X_t| < \infty$. Montrer que si $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, alors la série :

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

converge absolument avec probabilité 1. Si, de plus, $\sup_t \mathbb{E}|X_t|^2 < \infty$, montrer que la série converge en moyenne quadratique vers la même limite. Montrer l'égalité (1) du cours sur les ARMA.

Exercice 5.17 Fonction génératrice de l'autocovariance. Soit X donné par :

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j}, \quad (Z_t) \sim WN(0, \sigma^2).$$

On suppose que la série $\sum_{j \in \mathbb{Z}} |\psi_j| z^j$ converge pour $r^{-1} < |z| < r$, pour un certain $r > 1$. Montrer que la fonction génératrice de l'autocovariance de X , $G_X(z) = \sum_{j \in \mathbb{Z}} \gamma_X(j) z^j$ s'écrit :

$$G_X(z) = \sigma^2 \psi(z) \bar{\psi}(z^{-1}), \quad r^{-1} < |z| < r.$$

En déduire que pour un processus $ARMA(p, q)$ de la forme $\phi(B)X_t = \theta(B)Z_t$, avec θ n'ayant aucune racine sur le cercle unité, et ϕ ayant ses racines en dehors du disque unité, il existe $r > 1$ tel que :

$$G_X(z) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}.$$

Exercice 5.18 Calculer la fonction d'autocorrélation du processus Y défini par :

$$Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12}, \quad (Z_t) \sim WN(0, \sigma^2).$$

Exercice 5.19 Soient X et Y deux processus stationnaires vérifiant :

$$X_t - \alpha X_{t-1} = W_t, \quad (W_t) \sim WN(0, \sigma^2),$$

et

$$Y_t - \alpha Y_{t-1} = X_t + Z_t \quad (Z_t) \sim WN(0, \sigma^2),$$

avec $|\alpha| < 1$ et (W_t) et (Z_t) non corrélés. Calculer la densité spectrale de Y .

Exercice 5.20 Soit X un processus stationnaire vérifiant :

$$X_t - 0.99X_{t-3} = W_t, \quad (W_t) \sim WN(0, \sigma^2).$$

Calculer la densité spectrale de X et donner l'allure de son graphe. Celui-ci suggère-t-il que les trajectoires de X vont présenter des oscillations ? Si oui, quelle sera la période approximative d'oscillation ?

Exercice 5.21 Soient X un $ARMA(p_1, q_1)$ et Y un $ARMA(p_2, q_2)$ décorré de X . On pose $T = X + Y$. Montrer que T est un $ARMA(p_3, q_3)$, avec :

$$p_3 = p_1 + p_2, \quad q_3 = \max\{p_1 + q_2, q_1 + p_2\}.$$

Pour la première question, on pourra utiliser le résultat suivant (pour une preuve, voir par exemple [2] Théorèmes 4.4 et 4.6 pages 79–82).

Théorème 5.14 (Fejer-Riesz) Soit F une fraction rationnelle complexe. Si $f(\lambda) = F(e^{-i\lambda})$ est réelle, positive, et intégrable sur $[0, 2\pi]$, alors il existe une fraction irréductible $\frac{Q}{P}$ avec P sans racine de module 1 telle que :

$$F(z) = \left| \frac{Q}{P}(z) \right|^2 \quad \forall z \in \mathbb{C}, |z| = 1.$$

Si F est à coefficients réels, alors on peut prendre $\frac{Q}{P}$ à coefficients réels. De plus, la fraction rationnelle $\frac{Q}{P}$ est unique à une constante multiplicative près (de module 1) si on impose que Q n'a que des racines de module ≥ 1 et P des racines de module > 1 .

Exercice 5.22

Théorème 5.15 Soit X un $AR(p)$ stationnaire de moyenne nulle. Soit $\phi_m = (\phi_{m1}, \dots, \phi_{mm})$ le vecteur des coefficients du meilleur prédicteur linéaire $\phi'_m \mathbf{X}_m$ de X_{m+1} basé sur $\mathbf{X}_m = (X_m, \dots, X_1)'$. Soit Γ_m^{-1} la matrice des covariances de X , $(\gamma(i-j))_{1 \leq i, j \leq m}$. Enfin, soit $\hat{\phi}_m$ l'estimateur de Yule-Walker de ϕ_m . Alors,

$$n^{1/2}(\hat{\phi}_m - \phi_m) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_m^{-1}).$$

Corollaire 5.16 Si $m > p$,

$$n^{1/2} \hat{\phi}_{mm} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Montrer le Corollaire 5.16 à partir du Théorème 5.15.

Exercice 5.23 Soient X et Z deux variables aléatoires indépendantes et de carré intégrable. On suppose que Z est centrée et que la variance de X n'est pas nulle. Soit $Y = X^2 + Z$. Calculer $\mathbb{E}(Y|X)$, l'espérance conditionnelle de Y sachant X . Puis calculer Y^* , le meilleur prédicteur linéaire de Y sachant X . Lorsque X et Z sont des normales centrées réduites, montrer que l'erreur quadratique de prédiction linéaire de Y sachant X est strictement supérieure à l'erreur de prédiction par l'espérance conditionnelle.

Rappel : Y^* doit être de la forme $a + bX$ avec :

$$\mathbb{E}(a + bX) = \mathbb{E}(Y) \text{ et } \mathbb{E}((a + bX)X) = \mathbb{E}(YX).$$

Exercice 5.24 Soit X un $MA(1)$:

$$X_t = W_t - \alpha W_{t-1}, \quad (W_t) \sim WN(0, \sigma^2),$$

avec $|\alpha| < 1$. Montrer que le meilleur prédicteur de X_{t+1} en moyenne quadratique sachant $(X_{t-k})_{k \geq 0}$ est :

$$\hat{X}_{t+1} = - \sum_{j=1}^{\infty} \alpha^j X_{t+1-j}.$$

Quelle est l'erreur quadratique moyenne de prédiction ?

Exercice 5.25 Soit X un $AR(p)$ défini par $\phi(B)X_t = Z_t$. Calculer le meilleur prédicteur \hat{X}_{t+1} de X_{t+1} connaissant X_t, \dots, X_{t+1-p} , puis le meilleur prédicteur \hat{X}_{t+h} de X_{t+h} connaissant X_t, \dots, X_{t+1-p} . Calculer l'erreur quadratique moyenne de prédiction.

Chapitre 6

Processus ARIMA et SARIMA - Processus à corrélation périodique

Les processus *ARIMA* et *SARIMA* sont des processus aléatoires non stationnaires qui présentent des “tendances aléatoires” et/ou des “variations saisonnières aléatoires”. Ce chapitre en donne les définitions, quelques exemples, ainsi que les principes de leur identification. On indique également la représentation des processus *ARIMA* par modèle d’état.

6.1 Les processus ARIMA

6.1.1 Définitions, exemples, propriétés

Définition 6.1 Pour d entier tel que $d \geq 1$, le processus X_t est un $ARIMA(p, d, q)$ si le processus $Y_t = (1 - B)^d X_t$ est un processus $ARMA(p, q)$ de moyenne nulle. On dit aussi que le processus $ARIMA(p, d, q)$ est un processus $ARMA(p, q)$ de moyenne nulle “intégré” d fois.

X_t satisfait donc une équation de la forme :

$$\Phi^*(B)X_t \equiv \Phi(B)(1 - B)^d X_t = \Theta(B)Z_t \quad (6.1)$$

Φ et Θ sont des polynômes de degrés respectifs p et q , et B est l’opérateur retard :

$$\begin{aligned} B^k X_t &= X_{t-k} \\ \Phi(z) &= 1 - \phi_1 z - \dots - \phi_p z^p \\ \Theta(z) &= 1 + \theta_1 z + \dots + \theta_q z^q \end{aligned}$$

Remarque : si $\nabla^d X_t$ est un $ARMA(p, q)$ de moyenne non nulle m , alors

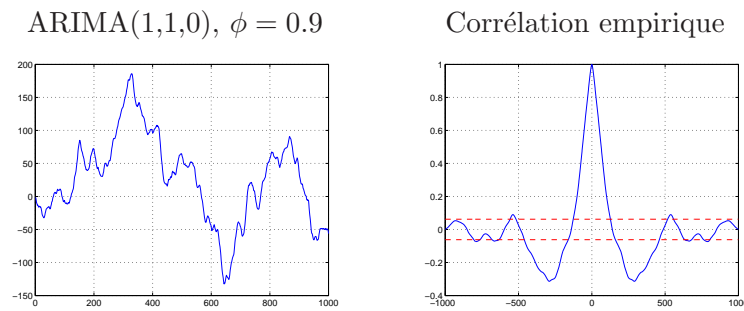
$$X_t = m \frac{t^d}{d!} + \tilde{X}_t$$

où \tilde{X}_t est un $ARIMA(p, d, q)$.

Exercice 6.1 Calculer l’évolution au cours du temps de la variance d’un processus $ARIMA(0, 1, 0)$, puis celle d’un processus $ARIMA(0, 2, 0)$.

Les processus *ARIMA* ne sont pas stationnaires.

La Figure 6.1 représente une réalisation d’un processus $ARIMA(1, 1, 0)$ avec $\phi = 0.9$ et sa fonction de corrélation empirique. Une telle fonction de corrélation empirique (à décroissance lente) est un bon indicateur du fait que l’on a affaire à un processus *ARIMA* (ou qu’il existe une tendance dans la série).

FIG. 6.1 – Un $ARIMA(1, 1, 0)$ avec $\phi = 0.9$ et sa fonction de corrélation empirique

La Figure 6.2 montre la corrélation empirique et la corrélation partielle empirique du processus obtenu en différenciant une fois le processus d'origine. La simple observation de ces quatre figures suggère fortement que le processus considéré est un $ARIMA(1, 1, 0)$.

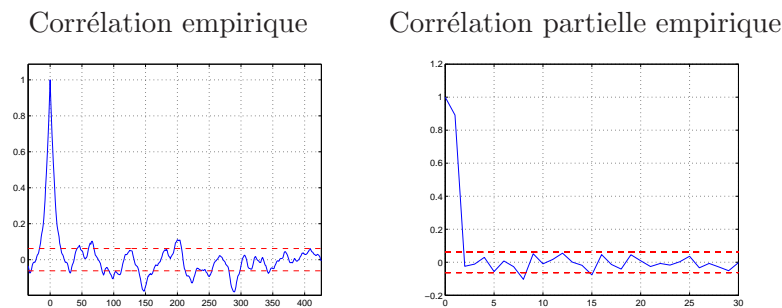


FIG. 6.2 – Corrélations empiriques du processus de la Figure 6.1 différencié une fois

La Figure 6.3 montre une réalisation d'un processus $ARIMA(0, 1, 1)$ avec $\theta = -0.8$, sa corrélation empirique, la corrélation empirique du processus différencié une fois et sa corrélation partielle empirique. La simple observation de ces quatre figures suggère fortement que le processus considéré est un $ARIMA(0, 1, 1)$

Exercice 6.2 Soit X un $ARIMA(p, d, q)$ tel que :

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t,$$

où Z est un bruit blanc de variance σ^2 . Montrer que cette équation est aussi satisfaite par le processus $W_t = X_t + A_0 + A_1 t + \dots + A_{d-1} t^{d-1}$ où les A_i sont des variables aléatoires.

Cet exercice montre que, bien que les processus $ARIMA$ ne contiennent pas de tendances au sens défini précédemment (i.e. des tendances déterministes), les modèles $ARIMA$ permettent de rendre compte de tendances polynomiales.

6.1.2 Prédiction des processus ARIMA

Exercice 6.3 Soit un processus $ARIMA(p, 2, 0)$:

$$(1 - B)^2 X_t = Y_t$$

où Y_t est un processus $AR(p)$ de paramètres connus.

Ecrire les équations de prédiction à un pas de X : prédiction de X_{t+1} en fonction de X_0 à X_t .

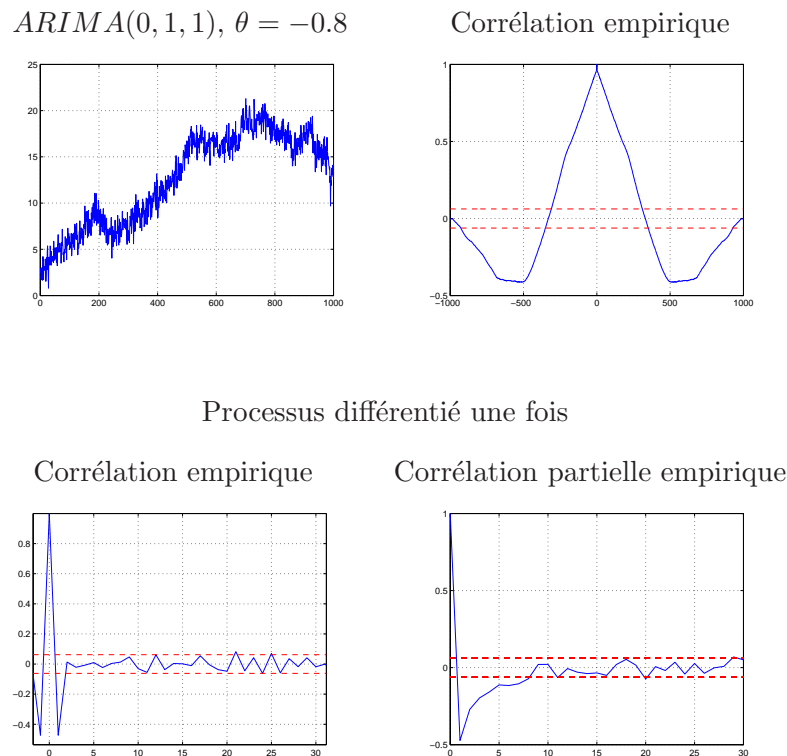


FIG. 6.3 – $ARIMA(0, 1, 1)$ avec $\theta = -0.8$, son ACF, l'ACF du processus différencié une fois et sa PACF

6.1.3 Principes d'identification

Les principes d'identification des processus $ARIMA$ sont les suivants :

- la 'présomption' de processus $ARIMA$ résulte de la conjonction :
 - de l'observation visuelle de la série,
 - du fait que la corrélation empirique décroît très lentement. En effet, si la corrélation empirique décroît lentement, on peut avoir affaire soit à un processus $ARIMA$, soit à une série avec tendance. L'observation visuelle peut permettre de lever le doute.
- on opère ensuite sur ce processus des opérations de différenciation successives, jusqu'à obtenir un processus stationnaire (sa corrélation empirique doit décroître assez rapidement); le nombre minimum d'opérations nécessaires donne la valeur de d ,
- le processus obtenu est enfin identifié à un processus $ARMA$.

Nota

Il peut arriver qu'une série chronologique contienne en même temps une tendance polynomiale et un processus $ARIMA$. Dans ce cas, l'identification sera une 'hybridation' entre la technique qui vient d'être décrite et une des techniques décrites au chapitre 3. Par exemple, si la série est la somme d'un polynôme de degré deux et d'une marche au hasard (un bruit blanc intégré une fois), l'identification pourra se faire *par exemple* en deux étapes :

- différencier une fois; le résultat est alors la somme d'une fonction linéaire et d'un bruit blanc,
- estimer les paramètres de la fonction obtenue par une estimation paramétrique.

Mais on peut aussi différencier une deuxième fois, le processus obtenu sera alors un processus $MA(1)$ avec $\theta = -1$ de moyenne *non nulle* et sera identifié comme tel en suivant les méthodes décrites au chapitre 5.

Nota sur les processus ARFIMA

La définition d'un processus *ARIMA* peut s'étendre au cas où d n'est pas entier ; on parle alors de processus *ARFIMA* (F pour Fractional). Ce point est développé au chapitre 7 consacré aux processus à mémoire longue.

6.1.4 Représentation par modèle d'état

Considérons l'équation 6.1 qui définit le processus $ARIMA(p, d, q)$ X :

$$\Phi^*(B)X_t \equiv \Phi(B)(1-B)^d X_t = \Theta(B)Z_t$$

et en particulier la relation :

$$\Phi^*(B)X_t = \Theta(B)Z_t$$

Cette relation ne définit pas un processus *ARMA*, car X n'est pas stationnaire (Φ^* a d pôles égaux à 1), mais elle est formellement identique à la relation générale qui définit un processus $ARMA(p+d, q)$.

On peut donc trouver une représentation par modèle d'état (se reporter au chapitre 5), en prenant $r = \max(p+d, q+1)$, et en déterminant les coefficients ϕ^* par développement de $\Phi^*(B) = \Phi(B)(1-B)^d$.

Par exemple, soit un $ARIMA(1, 1, 1)$ et soit ϕ le coefficient de la partie $AR(1)$. Alors le modèle d'état sera celui d'un $ARMA(2, 1)$, soit $r = 2$, avec :

$$1 - \phi_1^*B - \phi_2^*B^2 = (1 - \phi B)(1 - B)$$

ce qui donne :

$$\phi_1^* = 1 + \phi \quad \text{et} \quad \phi_2^* = -\phi.$$

6.2 Les processus SARIMA**6.2.1 Définitions, exemples, propriétés**

Les modèles *SARIMA* (S pour Seasonal) permettent de rendre compte des variations saisonnières dans la série considérée, ces variations pouvant elles-mêmes présenter un caractère aléatoire.

La Figure 6.4 représente des parties de réalisations de divers niveaux de processus *SARIMA*, générés à partir d'une même séquence de bruit blanc. On trouve successivement (de gauche à droite puis de haut en bas) :

- un processus *ARMA* présentant une composante saisonnière de période 20,
- le même processus intégré une fois, à la période de l'échantillon,
- le processus initial intégré une fois, à la période de la saison,
- le processus a été intégré une fois à la période de l'échantillon, et une fois à la période de la saison.

Exercice 6.4 Donner une interprétation des courbes de la Figure 6.4. Indiquer quels types de variations temporelles un modèle *SARIMA* est capable de représenter. Indiquer des types de variations qu'un modèle *SARIMA* ne peut pas représenter. Que faudrait-il faire dans ces cas-là ?

Définition 6.2 Le processus X_t est appelé $SARIMA(p, d, q) \times (P, D, Q)_s$ si le processus différencié :

$$Y_t = (1-B)^d(1-B^s)^D X_t \tag{6.2}$$

est un processus *ARMA* tel que :

$$\Phi(B)\Phi_s(B^s)Y_t = \Theta(B)\Theta_s(B^s)Z_t \tag{6.3}$$

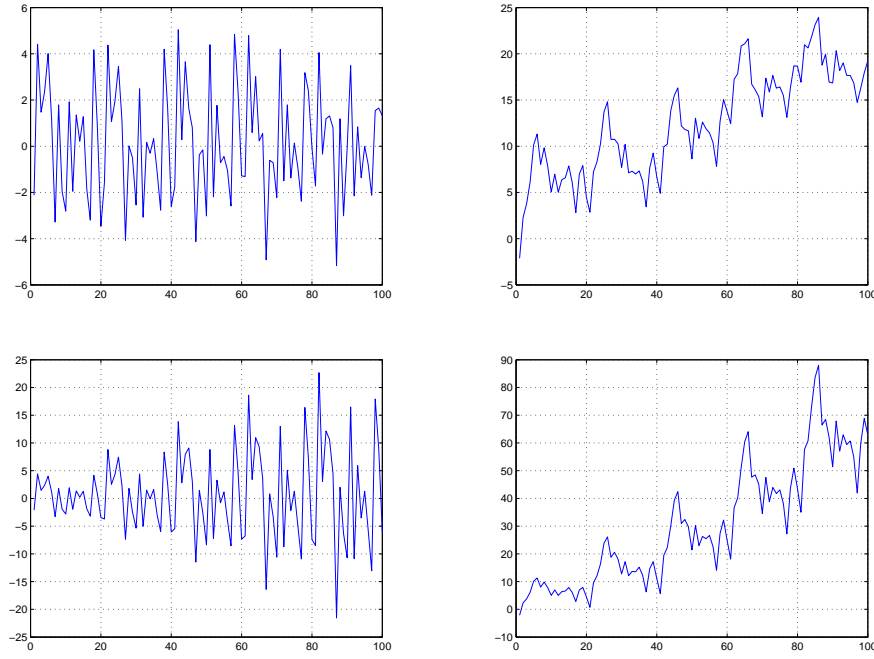


FIG. 6.4 – De l'ARMA au SARIMA

Φ et Θ sont des polynômes de degrés respectifs p et q :

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad (6.4)$$

$$\Theta(z) = 1 + \theta_1 z - \dots - \theta_q z^q \quad (6.5)$$

Φ_s et Θ_s sont des polynômes de degrés respectifs P et Q :

$$\Phi_s(z) = 1 - \phi_{s,1} z - \dots - \phi_{s,P} z^P \quad (6.6)$$

$$\Theta_s(z) = 1 + \theta_{s,1} z - \dots - \theta_{s,Q} z^Q \quad (6.7)$$

Exercice 6.5 Indiquer les étapes successives qui permettent de synthétiser un processus SARIMA tel que défini dans les équations 6.2 à 6.7.

Interpréter les quatre courbes de la Figure 6.4 par rapport à cette définition.

6.2.2 Principes d'identification

Dans un processus SARIMA, il y a interaction entre les deux modèles (l'un qui décrit les évolutions à la période de l'échantillon, c'est-à-dire entre deux saisons successives, l'autre qui décrit les évolutions à la période de la saison). Cette interaction rend complexe l'interprétation de la covariance empirique dans le cas général.

On indique ici quelques principes élémentaires pour une identification simple, dans l'hypothèse où il y a un 'découplage' entre les deux modèles ; en d'autres termes la fonction de corrélation déterminée par les polynômes $\Phi(z)$ et $\Theta(z)$ atteint des valeurs suffisamment petites pour $t = s/2$.

On peut alors utiliser la procédure suivante :

- trouver les ordres d et D tels que le processus $Y_t = (1 - B)^d (1 - B^s)^D X_t$ soit stationnaire,
- utiliser la corrélation empirique et la corrélation partielle empirique pour les valeurs multiples de s pour en déduire des valeurs préliminaires des ordres P et Q et les coefficients correspondants,

- utiliser la corrélation empirique et la corrélation partielle empirique pour les valeurs jusqu'à $s/2$ pour en déduire les ordres p et q et les coefficients correspondants,
- les techniques d'ajustement et de contrôle décrites dans le chapitre 5 sont applicables.

Lorsqu'il n'y a pas ce découplage entre les deux modèles, une méthode plus complexe peut être envisagée (en supposant s connu) :

- trouver les ordres d et D tels que le processus $Y_t = (1 - B)^d(1 - B^s)^D X_t$ soit stationnaire,
- faire une hypothèse sur les valeurs de p, q, P, Q ,
- développer la formule 6.3 du processus ARMA supposé obtenu après différentiation :

$$\Phi(B)\Phi_s(B^s)Y_t = \Theta(B)\Theta_s(B^s)Z_t$$

devient :

$$\Phi_g(B)Y_t = \Theta_g(B)Z_t$$

où l'indice 'g' signifie 'global',

- estimer globalement les valeurs des paramètres par l'une des méthodes décrites au chapitre 5,
- et revenir sur l'hypothèse de départ si nécessaire.

Les coefficients de Φ , Φ_s , Θ , Θ_s seront calculés à partir des coefficients de Φ_g et Θ_g obtenus.

Nota

La notion de processus *SARIMA* peut s'étendre à plusieurs périodes différentes pour un même processus. Par exemple, l'observation des températures en un lieu donné fait apparaître :

- des variations journalières,
- des variations annuelles.

Il est nécessaire de disposer de séries de grande taille pour analyser ces processus. Par exemple, observer une température pendant trente ans avec une période de une heure conduit à une taille de l'ordre de 260 000.

6.3 Les processus à corrélation périodique

Les processus à corrélation périodique (ou PCRPP : Periodically Correlated Random Processes) se rencontrent dans de nombreuses applications. La référence [11] donne une bibliographie importante sur les processus cyclostationnaires, dont font partie les PCRPP.

Ces processus *ne sont pas stationnaires* : leur moyenne et leur fonction de covariance sont périodiques. Un exemple de PCRPP est le produit d'une fonction périodique par un processus stationnaire.

Si la période est un multiple de la période d'échantillonnage, il est souvent possible de trouver un modèle de type PARMA (Periodic-ARMA) : un PARMA (p, q, s) , où s est la période (saisonnalité) est donné par la relation :

$$X_{ks+m} = \sum_{i=1}^p \phi_i^{(m)} X_{ks+m-i} + \sum_{j=0}^q \theta_j^{(m)} Z_{ks+m-j} \quad m = 1, s$$

ce qui signifie qu'il y a s relations de récurrence différentes en général.

On peut trouver un modèle équivalent, qui est un *ARMA vectoriel* (donc stationnaire) de taille s . L'écriture des formules générales pour le passage d'un PARMA à un ARMA vectoriel étant assez complexe, nous l'indiquons ici dans un cas simple (PARMA(1, 0, 2)) donné par :

$$\begin{aligned} X_{2k+1} &= \phi^{(1)} X_{2k} + Z_{2k+1} \\ X_{2k+2} &= \phi^{(2)} X_{2k+1} + Z_{2k+2} \end{aligned}$$

Pour l'ARMA vectoriel de dimension 2, nous avons :

$$Y_{k+1} = HY_k + U_{k+1}$$

où $Y_{k+1} = (X_{2k+1}, X_{2k+2})'$, H est une matrice 2,2 dont les coefficients sont à préciser, et la suite U_1, U_2, \dots est une suite de vecteurs aléatoires indépendants dont la matrice de covariance C_U est à préciser.

La difficulté provient de ce que Y_k ne dépend que de X_{2k-1} et X_{2k} , il faut donc 'propager' l'équation qui donne X_{2k+2} de la façon suivante :

$$\begin{aligned} X_{2k+2} &= \phi^{(2)} X_{2k+1} + Z_{2k+2} \\ &= \phi^{(2)} (\phi^{(1)} X_{2k} + Z_{2k+1}) + Z_{2k+2} \\ &= \phi^{(1)} \phi^{(2)} X_{2k} + \phi^{(2)} Z_{2k+1} + Z_{2k+2} \end{aligned}$$

On voit ainsi que :

$$H = \begin{bmatrix} 0 & \phi^{(1)} \\ 0 & \phi^{(1)} \phi^{(2)} \end{bmatrix}$$

et

$$U_{k+1} = \begin{bmatrix} Z_{2k+1} \\ \phi^{(2)} Z_{2k+1} + Z_{2k+2} \end{bmatrix}$$

La séquence des U_k est bien indépendante, et sa matrice de covariance vaut :

$$C_U = \sigma_Z^2 \begin{bmatrix} 1 & \phi^{(2)} \\ \phi^{(2)} & 1 + (\phi^{(2)})^2 \end{bmatrix}$$

On voit donc qu'un processus PARMA a une représentation par modèle d'état et qu'on peut utiliser le filtrage de Kalman pour le prédire, donc pour en estimer les paramètres. Il faut noter que, dans la représentation précédente, la valeur de s est supposée connue.

Chapitre 7

Processus à mémoire longue

On rencontre des processus à mémoire longue dans un certain nombre de domaines : métrologie (dérive des oscillateurs, bruit de capteurs inertiels, ...), météorologie (vitesses de vent), économie, etc.

Les processus à mémoire longue ont une définition précise (voir paragraphe 7.2.1). Certains peuvent être représentés par un modèle de type *ARFIMA* (*ARMA* intégré avec un ordre fractionnaire) ; seuls ces modèles sont étudiés dans ce chapitre.

Rappel : dans l'étude et la modélisation des séries chronologiques, on analyse la partie 'bruit' après avoir éliminé le plus complètement possible la partie 'déterministe' du signal. Les éléments qui suivent, en particulier la modélisation des processus à mémoire longue, supposent que ce travail préliminaire a été fait correctement ; par exemple, la présence de composantes périodiques, même faibles, peut dégrader considérablement les résultats obtenus.

7.1 Quelques rappels

7.1.1 Décroissance de la fonction de corrélation des processus *ARMA*

Les processus stationnaires *ARMA* sont dits 'à mémoire courte', car leur fonction de corrélation est bornée de la façon suivante :

$$|\rho(k)| \leq Cr^k, \quad k = 1, 2, \dots, \quad |r| < 1$$

(décroissance exponentielle de la fonction de corrélation)

ce qui entraîne que la somme des $\rho(k)$ est absolument convergente :

$$\sum |\rho(k)| < \infty$$

Eléments de démonstration :

pour un processus *ARMA*(p, q), à partir d'un certain rang, la fonction de covariance est donnée par :

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \dots + \phi_p \gamma(k-p) = \sum_{j=1}^p \phi_j \gamma(k-j)$$

alors les $\gamma(k)$ peuvent s'écrire :

$$\gamma(k) = \sum_{i=1}^p \alpha_i u_i^k$$

où les u_i sont les pôles du polynôme $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$, et si tous les pôles sont deux à deux distincts.

Le processus *ARMA* étant supposé causal, les pôles u_i ont tous un module strictement inférieur à 1. Dans le cas où un pôle particulier u_l a un module r strictement supérieur à tous les autres, alors :

$$|\gamma(k)| \sim \alpha_l r^k \quad \text{quand } k \rightarrow \infty$$

Ce cas ne se produit que si u_l est réel ; dans le cas où deux pôles complexes conjugués ont un module strictement supérieur à tous les autres, cette équivalence n'est plus valable, mais la suite du raisonnement reste juste.

On peut ensuite en déduire que :

$$|\rho(k)| \leq Cr^k, \quad k = 1, 2, \dots,$$

où $C > 0$ et $0 < r < 1$

□

Conséquence sur la variance de la moyenne échantillonnée

La moyenne échantillonnée (ou empirique) d'un processus X est définie par :

$$\hat{m}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Pour un processus *ARMA* de moyenne nulle, la moyenne échantillonnée a une moyenne nulle et une variance qui décroît en $1/n$ quand n tend vers l'infini :

$$E(\hat{m}_n) = 0 \quad \text{Var}(\hat{m}_n) \sim \frac{K}{n} \quad n \rightarrow \infty$$

Éléments de démonstration :

De l'expression

$$\hat{m}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

on déduit :

$$\text{Var}(\hat{m}_n) = \frac{1}{n^2} E\left[\left(\sum_{i=1}^n X_i\right)^2\right]$$

puisque $E(\hat{m}_n) = 0$.

En développant le carré $(\sum_{i=1}^n X_i)^2$ et en prenant la somme des espérances mathématiques des divers termes, on obtient :

$$\text{Var}(\hat{m}_n) = \frac{1}{n^2} [n\gamma(0) + 2 \sum_{p=1}^{n-1} (n-p)\gamma(p)]$$

Pour simplifier la démonstration, prenons dans un premier temps le cas d'un *AR*(1) pour lequel $\gamma(p) = \gamma(0)\phi^p$ avec $\gamma(0) = \sigma_X^2$ (variance de X).

On déduit :

$$\frac{n\text{Var}(\hat{m}_n)}{\sigma_X^2} = 1 + 2 \sum_{p=1}^{n-1} \left[\left(1 - \frac{p}{n}\right) \phi^p \right]$$

soit :

$$\frac{n\text{Var}(\hat{m}_n)}{\sigma_X^2} = 1 + 2 \sum_{p=1}^{n-1} \phi^p - \frac{2}{n} \sum_{p=1}^{n-1} p\phi^p$$

quand $n \rightarrow \infty$,

$$\frac{n \text{Var}(\hat{m}_n)}{\sigma_X^2} \rightarrow 1 + \frac{2\phi}{1-\phi}$$

qu'on peut encore écrire :

$$\text{Var}(\hat{m}_n) = \frac{\sigma_X^2}{n} \left(1 + \frac{2\phi}{1-\phi}\right) + o\left(\frac{1}{n}\right)$$

□

Si X est un bruit blanc ($\phi = 0$), on retrouve le résultat connu :

$$\text{Var}(\hat{m}_n) = \frac{\sigma_X^2}{n}$$

Noter que la corrélation du processus n'entraîne pas nécessairement une augmentation de la variance asymptotique de la moyenne échantillonnée. Dans le cas d'un $AR(1)$ par exemple, ϕ négatif entraîne une diminution de cette variance. Quand $\phi \rightarrow -1$, alors $1 + \frac{2\phi}{1-\phi} \rightarrow 0$, donc $\text{Var}(\hat{m}_n)$ également.

De façon générale, on démontre que $n \text{Var}(\hat{m}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h)$ quand $n \rightarrow \infty$ si $\sum_h |\gamma(h)| < \infty$.

7.1.2 Densité spectrale des processus ARMA

Pour un processus aléatoire stationnaire, la densité spectrale de puissance est définie par :

$$P_X(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) e^{-jk\omega}$$

Dans le cas d'un processus $ARMA$, la fonction γ est absolument intégrable, et $P_X(0)$ est finie :

$$P_X(0) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k)$$

On peut également calculer la dérivée de la densité spectrale par rapport à ω :

$$P'_X(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) (-jk) e^{-jk\omega}$$

soit :

$$P'_X(0) = \frac{-j}{2\pi} \sum_{k \in \mathbb{Z}} k \gamma_X(k)$$

Si le processus $ARMA$ est réel, sa densité spectrale est paire, et l'on a alors :

$$P'_X(0) = 0$$

En conclusion, pour un processus $ARMA$ réel, la densité spectrale en 0 existe et sa dérivée est nulle.

7.1.3 Les processus ARIMA

Ces processus correspondent à l'intégration d'ordre d (entier) d'un processus $ARMA$ et *ne sont pas stationnaires*. Le cas le plus simple est la 'marche au hasard' où X est défini par $(1 - B)X_t = Z_t$. Le polynôme $\Phi(z) = 1 - z$ a un zéro égal à 1, donc situé sur le cercle unité. On peut également envisager une intégration avec deux pôles (ou zéros) complexes conjugués situés sur le cercle unité, qui conduit à un processus de type 'marche au hasard modulée en fréquence'.

7.2 Les processus à mémoire longue

7.2.1 Définition des processus à mémoire longue

Un processus aléatoire est dit à mémoire longue s'il est stationnaire et si sa fonction d'auto-corrélation $\rho(k)$ est telle que :

$$\rho(k) \sim Ck^{2d-1} \quad \text{quand } k \rightarrow \infty$$

où $C \neq 0$ et $d < 0.5$.

On distingue :

- les processus à “mémoire intermédiaire”, tels que $d < 0$, soit $\sum_{k \in \mathbb{Z}} |\rho(k)| < \infty$
- les processus à “mémoire longue”, tels que $d > 0$, soit $\sum_{k \in \mathbb{Z}} |\rho(k)| = \infty$

Les processus à mémoire longue se rencontrent dans des phénomènes physiques (hydrologie, météorologie,...) ainsi qu'en économie.

7.2.2 Bruit blanc intégré avec un ordre fractionnaire : $ARIMA(0, d, 0)$

Le processus X_t pour $t \in \mathbb{Z}$ est un $ARIMA(0, d, 0)$ avec $d \in (-0.5, 0.5)$ si X_t est stationnaire et s'il satisfait l'équation aux différences :

$$\nabla^d X_t = Z_t \quad \text{avec } Z_t \sim iid(0, \sigma^2)$$

On a :

$$\nabla^d = (1 - B)^d = 1 - dB + \frac{d(d-1)}{2}B^2 - \dots$$

X_t est appelé bruit blanc intégré avec un ordre fractionnaire ou $ARFIMA(0, d, 0)$.

La comparaison avec un processus $ARMA$ porte essentiellement sur les propriétés de la fonction de corrélation et celles de la densité spectrale, qui sont liées.

7.2.2.1 Synthèse d'un processus $ARFIMA(0, d, 0)$

Nous allons étudier comment obtenir dans la pratique X_t à partir de Z_t , et inversement. La synthèse est utile :

- pour des besoins d'étude expérimentale des propriétés du processus, en connaissant ses paramètres de façon certaine,
- pour les phases d'identification et de vérification de modèles.

Il existe deux classes de méthodes : méthode temporelle, méthode spectrale.

Méthode temporelle

Soit $\nabla^d X_t = (1 - B)^d X_t = Z_t$

qui peut s'écrire de façon équivalente : $X_t = (1 - B)^{-d} Z_t$

On voit donc que l'on peut construire X_t à partir de Z_t comme un processus $AR(\infty)$ ou comme un processus $MA\infty$.

Plus précisément, soient $c_k(d)$ les coefficients du développement en série infinie, lorsqu'il existe, de $(1 + x)^d$, soit :

$$(1 + x)^d = \sum_{k=0}^{\infty} c_k(d) x^k$$

Les c_k sont donnés par $c_0(d) = 1$, et $c_k(d) = c_{k-1}(d) \frac{d-k+1}{k}$
 Alors la représentation “ $AR(\infty)$ ” de X_t est donnée par :

$$\sum_{k=0}^{\infty} c_k(d)(-1)^k X_{t-k} = Z_t$$

ou encore :

$$X_t = \sum_{k=1}^{\infty} -c_k(d)(-1)^k X_{t-k} + Z_t$$

et la représentation “ $MA(\infty)$ ” de X_t est donnée par :

$$X_t = \sum_{k=0}^{\infty} c_k(-d)(-1)^k Z_{t-k}$$

En appliquant la même méthode, on peut “reconstruire” Z_t à partir de X_t .

Dans la pratique, les séries sont calculées avec un nombre fini de termes $K = \min(t-1, K_0)$, où K_0 est fixé a priori. Le terme “ $t-1$ ” est applicable dans le cas où le temps est indicé par N^* (la première valeur de t est égale à 1).

Méthode spectrale

Soit à nouveau $X_t = (1 - B)^{-d} Z_t$.

Soit $f_g(\omega)$ la transformée de Fourier de la fonction g_t .

On a alors $f_X(\omega) = (1 - e^{-j\omega})^{-d} f_Z(\omega)$.

A partir de la séquence Z_t , on obtiendra donc X_t par la suite de trois opérations :

- calcul de la transformée de Fourier de Z_t , soit $f_Z(\omega)$,
- multiplication de $f_Z(\omega)$ par $(1 - e^{-j\omega})^{-d}$ pour obtenir $f_X(\omega)$,
- calcul de la transformée de Fourier inverse de $f_X(\omega)$ pour obtenir X_t .

Il faut exclure du calcul la valeur $\omega = 0$, en particulier pour $d > 0$, cas où la densité spectrale de X_t n’est pas définie pour $\omega = 0$. On peut par exemple remplacer la valeur en $\omega = 0$ par la moyenne des deux valeurs adjacentes.

La “reconstruction” de Z_t à partir de X_t fait appel au même principe.

7.2.2.2 Fonction de corrélation

Le processus X_t , $ARIMA(0, d, 0)$ avec $d \in (-0.5, 0.5)$, a les propriétés d’un processus à mémoire longue :

$$\rho(k) \sim Ck^{2d-1} \quad \text{quand } k \rightarrow \infty$$

avec :

$$C = \frac{\Gamma(1-d)}{\Gamma(d)}$$

où $\Gamma(\cdot)$ est la fonction gamma définie par : $\Gamma(x) = \begin{cases} \int_0^{\infty} t^{x-1} e^{-t} dt & \text{pour } x > 0 \\ \infty & \text{pour } x = 0 \\ x^{-1} \Gamma(1+x), & \text{pour } x < 0 \end{cases}$

Rappel : pour les valeurs de x entières et au moins égales à 1, on a $\Gamma(x) = (x-1)!$

Conséquence sur la variance de la moyenne échantillonnée

Pour un processus $ARIMA(0, d, 0)$, la moyenne échantillonnée a une moyenne nulle et une variance qui décroît quand n tend vers l’infini :

- en n^{2d-1} si d est positif (mémoire longue)

$$E(\hat{m}_n) = 0 \quad \text{Var}(\hat{m}_n) \sim K_1 n^{2d-1} \quad n \rightarrow \infty$$

– en $1/n$ si d est négatif (mémoire intermédiaire)

$$E(\widehat{m}_n) = 0 \quad \text{Var}(\widehat{m}_n) \sim \frac{K_2}{n} \quad n \rightarrow \infty$$

Dans ce cas où d est négatif, $\sum_h |\rho(h)| < \infty$ et $n\text{Var}(\widehat{m}_n) \rightarrow \sum_h \gamma(h)$ (voir fin du paragraphe 7.1.1), ce qui est conforme au résultat énoncé.

Éléments de démonstration :

Reprenons l'expression de la variance de la moyenne échantillonnée en fonction de la fonction de covariance du processus :

$$\text{Var}(\widehat{m}_n) = \frac{1}{n^2} [n\gamma(0) + 2 \sum_{p=1}^{n-1} (n-p)\gamma(p)]$$

On a vu précédemment que, pour un $ARIMA(0, d, 0)$:

$$\rho(k) \sim Ck^{2d-1} \quad \text{soit } \gamma(k) \sim \gamma(0)Ck^{2d-1} \quad \text{quand } k \rightarrow \infty$$

Pour simplifier la démonstration, on suppose pour la suite que :

$$\gamma(k) = \gamma(0)Ck^{2d-1} \quad \text{pour tout } k$$

On peut alors écrire :

$$\begin{aligned} \frac{\text{Var}(\widehat{m}_n)}{\gamma(0)} &= \frac{1}{n} + \frac{2C}{n} \sum_{p=1}^{n-1} (1 - \frac{p}{n}) p^{2d-1} \\ \frac{\text{Var}(\widehat{m}_n)}{\gamma(0)} &= \frac{1}{n} + \frac{2C}{n} \sum_{p=1}^{n-1} p^{2d-1} - \frac{2C}{n^2} \sum_{p=1}^{n-1} p^{2d} \end{aligned}$$

On a par ailleurs :

$$\sum_{p=1}^{n-1} p^{2d-1} \sim \frac{n^{2d}}{2d} \quad \text{quand } n \rightarrow \infty$$

$$\sum_{p=1}^{n-1} p^{2d} \sim \frac{n^{2d+1}}{2d+1} \quad \text{quand } n \rightarrow \infty$$

Soit :

$$\frac{1}{n} \sum_{p=1}^{n-1} p^{2d-1} \sim \frac{n^{2d-1}}{2d} \quad \text{quand } n \rightarrow \infty$$

$$\frac{1}{n^2} \sum_{p=1}^{n-1} p^{2d} \sim \frac{n^{2d-1}}{2d+1} \quad \text{quand } n \rightarrow \infty$$

Ainsi le premier terme de $\text{Var}(\widehat{m}_n)/\gamma(0)$ a une décroissance en n^{-1} , et la somme des deux autres termes a une décroissance en n^{2d-1} .

Une démonstration rigoureuse nécessite de démontrer que, dans cette somme, le terme qui vient multiplier n^{2d-1} n'est pas nul.

Donc $\text{Var}(\widehat{m}_n)/\gamma(0)$ a une décroissance en n^{-1} si $d < 0$ et en n^{2d-1} si $d > 0$. \square

7.2.2.3 Densité spectrale au voisinage de $\omega = 0$

On a la propriété suivante :

$$P_X(\omega) \sim \frac{\sigma_Z^2}{2\pi} |\omega|^{-2d} \quad \text{quand } \omega \rightarrow 0$$

Démonstration :

Selon la définition du processus X_t , on a :

$$X_t = \nabla^{-d} Z_t = (1 - B)^{-d} Z_t$$

La densité spectrale de X_t s'écrit donc :

$$P_X(\omega) = \frac{\sigma_Z^2}{2\pi} |1 - e^{-j\omega}|^{-2d} = \frac{\sigma_Z^2}{2\pi} |2 \sin(\frac{\omega}{2})|^{-2d} \sim \frac{\sigma_Z^2}{2\pi} |\omega|^{-2d} \quad \text{quand } \omega \rightarrow 0 \quad \square$$

Ainsi, la densité spectrale de X_t au voisinage de $\omega = 0$ tend vers l'infini si $d > 0$ (mémoire longue) et vers 0 si $d < 0$ (mémoire intermédiaire), et l'on a la relation :

$$\ln(P_X(\omega)) = A - 2d \ln(|\omega|) \quad \text{avec } A = \ln\left(\frac{\sigma_Z^2}{2\pi}\right)$$

7.2.2.4 Intégration fractionnaire autour d'une pulsation non nulle

De même que pour les processus *ARIMA* (non stationnaires), on peut construire des processus *ARFIMA* (stationnaires, i.e. avec un ordre d tel que $d \in (-0.5, 0.5)$) en considérant deux zéros conjugués sur le cercle unité : $z_1 = e^{j\omega_0}$ et $z_2 = e^{-j\omega_0}$. L'opérateur $\nabla = 1 - B$ est alors remplacé par $\nabla_{\omega_0} = (1 - e^{-j\omega_0} B)(1 - e^{j\omega_0} B) = 1 - 2 \cos \omega_0 B + B^2$

Si l'on appelle X_{ω_0} le processus défini par $\nabla_{\omega_0}^d X_{\omega_0,t} = Z_t$, on obtient :

$$P_{X_{\omega_0}}(\omega) = \frac{\sigma_Z^2}{2\pi} |(1 - e^{-j\omega_0} e^{-j\omega})(1 - e^{j\omega_0} e^{-j\omega})|^{-2d}$$

on peut calculer que :

$$|(1 - e^{-j\omega_0} e^{-j\omega})(1 - e^{j\omega_0} e^{-j\omega})|^2 = 4(\cos(\omega) - \cos(\omega_0))^2$$

quand ω est proche de ω_0 , soit $\omega = \omega_0 + d\omega$, alors :

$$|(1 - e^{-j\omega_0} e^{-j\omega})(1 - e^{j\omega_0} e^{-j\omega})|^2 \sim 4 \sin^2(\omega_0) (d\omega)^2 \quad \text{quand } d\omega \rightarrow 0$$

et

$$P_{X_{\omega_0}}(\omega) \sim \frac{\sigma_Z^2}{2\pi} \frac{1}{(4 \sin^2(\omega_0))^d} (d\omega)^{-2d} \quad \text{quand } d\omega \rightarrow 0$$

Cette propriété est également vraie quand ω est proche de $-\omega_0$. Le processus X_{ω_0} présente donc, du point de vue spectral, le même comportement autour des pulsations ω_0 et $-\omega_0$ que le processus X_t (*ARFIMA*(0, d , 0)) défini précédemment par $\nabla^d X_t = Z_t$ autour de la pulsation 0.

7.2.3 Les processus ARFIMA (p,d,q)

Les processus que l'on vient d'étudier sont limités à deux paramètres : d et σ_Z^2 , et ne permettent donc pas de modéliser des processus variés du point de vue de la fonction de covariance. Pour élargir ces possibilités, on introduit les processus *ARIMA*(p,d,q) :

le processus X_t est dit $ARIMA(p, d, q)$ ou $ARFIMA(p, d, q)$ avec $d \in (-0.5, 0.5)$ s'il est stationnaire et s'il satisfait l'équation :

$$\phi(B)\nabla^d X_t = \theta(B)Z_t$$

où Z_t est un bruit *i.i.d.* et ϕ et θ sont des polynômes de degrés respectifs p et q .

$U_t = \nabla^d X_t$ est un processus $ARMA$.

X_t peut être considéré comme un processus $ARMA$ prenant en entrée un bruit blanc intégré avec un ordre fractionnaire.

La figure 7.1 montre deux exemples de processus $ARFIMA$ de longueur $n = 1024$, le premier est un bruit blanc de variance 4 intégré avec un ordre fractionnaire $d = 0.4$, le deuxième est un $AR(1)$ obtenu à partir de la même séquence de bruit blanc avec $\phi = 0.9$ et intégré avec un ordre fractionnaire $d = 0.4$.

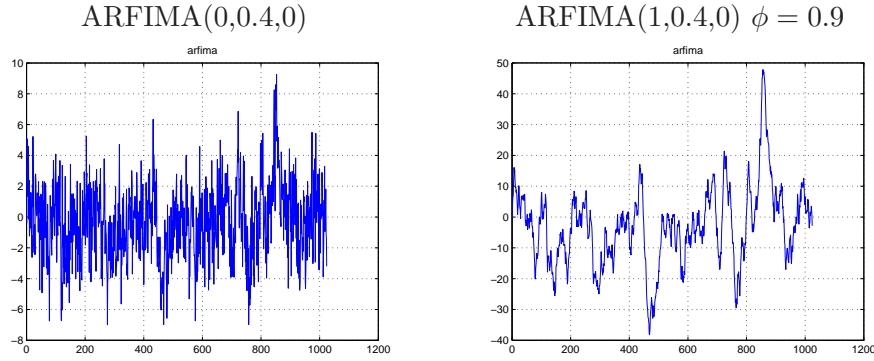


FIG. 7.1 – Deux exemples de processus ARFIMA

La fonction de corrélation d'un $ARFIMA(p, d, q)$ présente le même comportement asymptotique que celle d'un $ARFIMA(0, d, 0)$:

$$\rho(k) \sim Ck^{2d-1} \quad \text{quand } k \rightarrow \infty$$

La densité spectrale au voisinage de $\omega = 0$ d'un $ARFIMA(p, d, q)$ a la même forme, à un coefficient multiplicatif près, que celle d'un $ARFIMA(0, d, 0)$:

$$P_X(\omega) = \frac{\sigma_Z^2}{2\pi} \frac{|\theta(e^{-j\omega})|^2}{|\phi(e^{-j\omega})|^2} |1 - e^{-j\omega}|^{-2d} \sim \frac{\sigma_Z^2}{2\pi} \left(\frac{\theta(1)}{\phi(1)}\right)^2 |\omega|^{-2d} \quad \text{quand } \omega \rightarrow 0$$

avec :

$$\theta(1) = 1 + \theta_1 + \dots + \theta_q \quad \text{et} \quad \phi(1) = 1 - \phi_1 + \dots - \phi_p$$

On trouve une relation identique à celle trouvée pour un $ARIMA(0, d, 0)$ au voisinage de $\omega = 0$:

$$\ln(P_X(\omega)) = A - 2d \ln(|\omega|) \quad \text{avec } A = \ln\left(\frac{\sigma_Z^2}{2\pi} \left(\frac{\theta(1)}{\phi(1)}\right)^2\right)$$

La pente du log-périodogramme au voisinage de $\omega = 0$ n'est pas influencée par la partie $ARMA$ de l' $ARFIMA$.

7.2.4 Quelques principes d'identification

L'identification d'un $ARIMA(p, d, q)$ consiste à estimer les paramètres qui le définissent :

$$d, p, q, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \text{ et } \sigma_Z^2$$

7.2.4.1 Identification par la technique du maximum de vraisemblance

La technique du maximum de vraisemblance conduit à des équations voisines de celles vues pour les *ARMA* :
soit \mathbf{b} le vecteur :

$$\mathbf{b} = (d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

alors :

$$\hat{\sigma}_Z^2 = n^{-1} S(\hat{\mathbf{b}})$$

avec :

$$S(\hat{\mathbf{b}}) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

où $\hat{\mathbf{b}}$ est la valeur de \mathbf{b} qui minimise :

$$l(\mathbf{b}) = \ln(n^{-1} S(\mathbf{b})) + n^{-1} \sum_{j=1}^n r_{j-1}$$

Il existe une approximation de $l(\mathbf{b})$:

$$l_a(\mathbf{b}) = \ln\left(\frac{1}{n} \sum_j \frac{I_n(\omega_j)}{P(\omega_j; \mathbf{b})}\right)$$

où $I_n(\cdot)$ est le périodogramme, $P(\cdot; \mathbf{b})$ est la densité spectrale multipliée par $2\pi/\sigma^2$, et la somme inclut toutes les pulsations non nulles $\omega_j = 2\pi j/n$ de l'intervalle $(-\pi, \pi]$

Une approximation plus précise de $l(\mathbf{b})$ est donnée par :

$$l_b(\mathbf{b}) = l_a(\mathbf{b}) + n^{-1} \sum_j \ln(P(\omega_j; \mathbf{b}))$$

Après avoir estimé \mathbf{b} , soit $\hat{\mathbf{b}}$, on estime σ^2 par la formule :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_j \frac{I_n(\omega_j)}{P(\omega_j; \hat{\mathbf{b}})}$$

En utilisant cette méthode (ici l'approximation $l_a(\mathbf{b})$), on trouve dans le deuxième cas décrit ci-dessus (*ARFIMA*(1, d , 0)) :

$$\begin{aligned} \hat{\phi} &= 0,9058 & (\text{pour } \phi = 0,9) \\ \hat{d} &= 0,3724 & (\text{pour } d = 0,4) \\ \widehat{\sigma_z^2} &= 3,9888 & (\text{pour } \sigma_z^2 = 4) \end{aligned}$$

Bien sûr, cette estimation est faite en supposant que l'on sait que l'on a affaire à un *ARFIMA*(1, d , 0).

7.2.4.2 Utilisation du log-périodogramme

On peut estimer successivement d , puis les paramètres du processus *ARMA*.
Pour estimer d , on utilise la propriété du log-périodogramme au voisinage de $\omega = 0$ vue à la fin du paragraphe 2.3 :

$$\ln(P_X(\omega)) = A - 2d \ln(\omega)$$

On estimera d à partir du périodogramme. Plus précisément :

$$\hat{d} = - \frac{\sum_{i=1}^m (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^m (x_i - \bar{x})^2}$$

avec :

$$Y_i = \ln(I_n(\omega_i)) \quad \text{et} \quad x_i = \ln(|1 - e^{-j\omega_i}|^2),$$

et m est suffisamment petit pour que ω_m soit proche de zéro. Cette contrainte est nécessaire pour deux raisons :

- d’une part pour la raison que la propriété $\ln(P_X(\omega)) = A - 2d \ln(\omega)$ n’est valable qu’au voisinage de $\omega = 0$,
- et d’autre part parce que la présence d’une composante *ARMA* modifie la forme de la densité spectrale du processus quand ω s’éloigne de 0.

La figure 7.2 montre, pour les deux exemples présentés précédemment, l’évolution de \hat{d} en fonction de m , avec en ligne horizontale la valeur de d (0.4). Ces figures indiquent clairement les limites de la méthode.

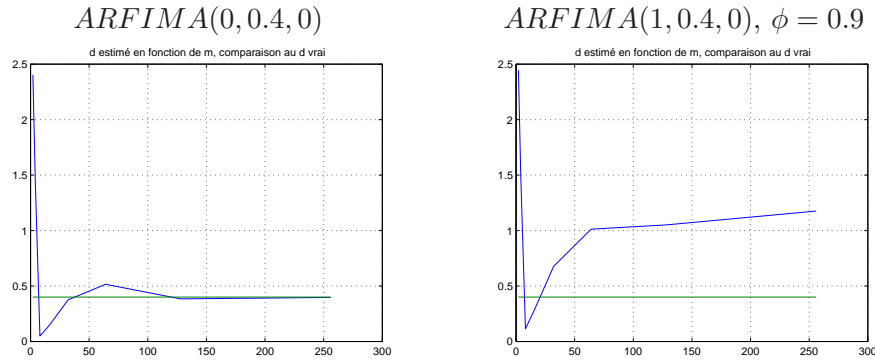


FIG. 7.2 – Evolution de \hat{d} en fonction de m

Ayant estimé d , on peut ‘reconstituer’ le processus *ARMA* :

$$U_t = \nabla^d X_t$$

Pour éviter l’opération explicite ∇^d qui contient une infinité de termes, on remarque que :

$$f_U(\omega) = (1 - e^{-j\omega})^d f_X(\omega)$$

où f est la transformée de Fourier.

On estime la transformée de Fourier de U par :

$$\hat{f}_U(\omega) = (1 - e^{-j\omega})^{\hat{d}} \hat{f}_X(\omega)$$

et on estime U_t par la transformée de Fourier inverse :

$$\tilde{U}_t = n^{-1/2} \sum_i e^{j\omega_i t} \hat{f}_U(\omega_i)$$

en excluant $\omega_i = 0$ (en particulier pour $d > 0$, cas où la densité spectrale de X n’existe pas pour $\omega = 0$).

On estime enfin les paramètres du processus *ARMA* U_t par les techniques étudiées précédemment.

7.2.4.3 Utilisation de la moyenne empirique, variance d'Allan

Si d est positif, la variance de la moyenne échantillonnée décroît en n^{2d-1} .

$$\text{Var}(\hat{m}_n) \sim K_1 n^{2d-1} \quad n \rightarrow \infty$$

Soit :

$$\ln(\text{Var}(\hat{m}_n)) = \ln(K_1) + (2d - 1) \ln(n) \quad n \rightarrow \infty$$

Si l'on représente sur un graphique $\ln(\widehat{\text{Var}}(\hat{m}_n))$ en fonction de $\ln(n)$, la pente (négative) de la droite obtenue donne la valeur de $2d - 1$.

Cette opération est illustrée par la figure 7.3 : la première courbe indique en trait plein la veme (variance empirique de la moyenne échantillonnée) pour un arfima construit avec un bruit blanc et $d=0.4$ et en pointillés la veme pour un bruit blanc de même variance que l'arfima. La deuxième courbe donne les mêmes éléments en remplaçant 'bruit blanc' par 'ar(1)' avec $\phi = 0.9$. Ici encore, on voit (deuxième courbe) que la présence d'un effet ARMA rend l'interprétation

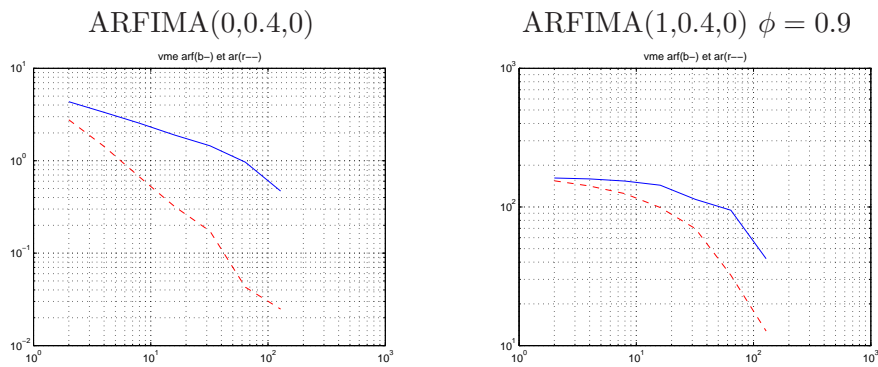


FIG. 7.3 – Variance de la moyenne échantillonnée

difficile, cela fonctionnerait peut-être mieux avec un échantillon plus grand. Au contraire, la première courbe montre une pente voisine de -0.38, soit d évalué à 0.31.

Une méthode voisine, très utilisée en métrologie, est la méthode de la variance d'Allan (on pourra se reporter à [1], article 'fondateur sur cette méthode) :

$$\sigma_X^2(\tau) = E[h_X(t, \tau)]^2$$

avec :

$$h_X(t, \tau) = \frac{1}{\sqrt{2}} [\bar{X}_\tau(t + \tau) - \bar{X}_\tau(t)]$$

et :

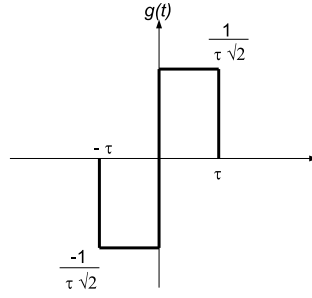
$$\bar{X}_\tau(t) = \frac{1}{\tau} \int_{t-\tau}^t X(u) du$$

Dans la pratique, on calcule :

$$\hat{\sigma}_X^2(\tau) = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} [\bar{X}_\tau(t + k\tau) - \bar{X}_\tau(t + (k-1)\tau)]^2$$

Pour une réalisation $X(t)$ du processus X_t , la fonction $h_X(t, \tau)$ est le produit de convolution de $X(t)$ avec la fonction :

$$g(t) = \frac{1}{\tau\sqrt{2}} (-\mathbf{1}_{[-\tau, 0]} + \mathbf{1}_{[0, \tau]})$$

FIG. 7.4 – Graphe de la fonction g

où :

$$\mathbf{1}_{[a,b]} = 1 \quad \text{pour } t \in [a,b] \quad \text{et } \mathbf{1}_{[a,b]} = 0 \quad \text{sinon}$$

La fonction $g(t)$ est représentée à la Figure 7.4.

La transformée de Fourier de $g(t)$ est donnée par la formule :

$$\hat{g}(f) = \int_{-\infty}^{+\infty} g(t)e^{-j2\pi ft} dt = \frac{1}{\tau\sqrt{2}} \left[\int_0^{\tau} e^{-j2\pi ft} dt - \int_{-\tau}^0 e^{-j2\pi ft} dt \right]$$

soit :

$$\hat{g}(f) = j\sqrt{2} \frac{\sin^2(\pi\tau f)}{\pi\tau f}$$

La densité spectrale du processus $h_X(t, \tau) = h_{X,\tau}$ est égale à :

$$P_{h_{X,\tau}}(f) = P_X(f) |\hat{g}(f)|^2$$

et la variance d'Allan en fonction de τ vaut :

$$\hat{\sigma}_X^2(\tau) = \int_{-\infty}^{\infty} P_{h_{X,\tau}}(f) df = \int_0^{\infty} P_X(f) \frac{2 \sin^4(\pi\tau f)}{(\pi\tau f)^2} df$$

On obtient :

$$\hat{\sigma}_X^2(\tau) = K\tau^\mu \quad \tau \rightarrow \infty$$

avec :

- $\mu = -1$ si la densité spectrale de X est constante (bruit blanc)
- $\mu = 0$ si la densité spectrale de X varie en f^{-1} (bruit de scintillation)
- $\mu = 1$ si la densité spectrale de X varie en f^{-2} (marche au hasard)

On constate que cette méthode est voisine de celle de la variance de la moyenne échantillonnée, puisque la densité spectrale (dsp) d'un *ARFIMA* varie en ω^{-2d} au voisinage de $\omega = 0$ et que $Var(\hat{m}_n)$ varie en n^{2d-1} :

- pour une dsp constante : $d = 0 \rightarrow Var(\hat{m}_n)$ varie en n^{-1} ,
- pour une dsp en f^{-1} : $d = 0.5 \rightarrow Var(\hat{m}_n)$ varie en n^0 ,

- pour une dsp en f^{-2} : $d = 1 \rightarrow \text{Var}(\hat{m}_n)$ varie en n^1 .

Les deux derniers cas ($d = 0.5$, $d = 1$) sont particuliers par rapport à ce qui a été écrit plus haut, car un ARFIMA est stationnaire pour $d < 0.5$ ¹, mais les résultats énoncés s'appliquent quand même dans ces cas : $d = 0.5$ correspond au cas limite du bruit en $1/f$, très connu dans le monde de la physique (bruit de scintillation ou flicker noise), $d = 1$ correspond à la marche au hasard.

On remarquera que la méthode de la variance d'Allan correspond à un cas particulier de l'analyse en ondelettes, puisque la fonction g est égale à l'ondelette de Haar.

La figure 7.5 illustre un résultat obtenu avec l'ondelette de Haar : la courbe continue donne la variance des détails successifs d'une décomposition en ondelettes de Haar pour un bruit blanc intégré avec $d=0.4$, la courbe en pointillés donne le même résultat avec un bruit blanc. On voit

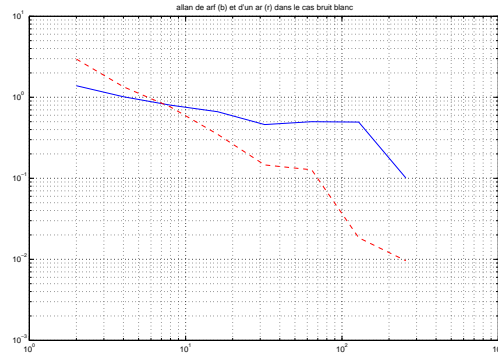


FIG. 7.5 – Analyse par une ondelette de Haar

clairement une pente de -1 pour le bruit blanc, et une pente voisine de -0.4 pour le bruit blanc intégré, soit une valeur de d estimée à 0.3.

¹donc dans les cas $d = 0.5$ et $d = 1$, la densité spectrale n'est pas définie, mais il est possible d'observer le périodogramme autour de la pulsation nulle. On parle donc de dsp en f^{-1} ou en f^{-2} de manière abusive

Chapitre 8

Processus GARCH

Les processus $ARMA/(G)ARCH$ sont des processus $ARMA$ dont la séquence de bruit d'entrée présente des propriétés particulières : c'est une séquence de variables aléatoires non corrélées, mais dépendantes (bruit blanc faible). Ces processus sont largement rencontrés en économie / finance. On trouvera des indications complémentaires dans [10] (article “fondateur”), [17], [9], [3].

8.1 Rappel sur les séries chronologiques

Dans le cours sur les séries chronologiques, on a modélisé la partie bruit par un processus $ARMA$ ou $ARIMA$ (avec un ordre d'intégration entier ou fractionnaire) ou encore $SARIMA$. On rappelle (chapitre 5, paragraphe 5.1) que le processus X_t est dit $ARMA(p, q)$ si l'on peut écrire :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

où Z_t est en général un *bruit blanc faible*, mais on a précisé que par défaut Z_t est un *bruit blanc fort*, soit :

$$Z_t \sim iid(0, \sigma_Z^2)$$

La séquence Z_t est une suite de variables aléatoires, non nécessairement gaussiennes, indépendantes et identiquement distribuées, de moyenne 0 et de variance σ_Z^2 .

On dit que X_t est “*piloté par une séquence i.i.d*” ou encore “*i.i.d. driven*”. Il en est de même lorsque l'on a affaire à un processus $ARIMA$ ou $SARIMA$.

8.2 Processus GARCH - Processus ARMA/GARCH

Dans les séries que nous allons étudier, les processus $GARCH$ (Generalized Autoregressive Conditionnaly Heteroskedastic) viennent en ‘remplacement’ du processus Z_t défini précédemment ; ces processus seront notés ϵ_t .

Attention : les définitions qui suivent peuvent varier selon les auteurs.

Définition 8.1 *Le processus ϵ_t est dit $ARCH(p)$ si :*

$$\epsilon_t = \sigma_t \eta_t ,$$

où :

$$\begin{aligned} \eta_t &\sim i.i.d \mathcal{N}(0, 1) && (\text{bruit blanc fort}) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 , \end{aligned}$$

avec $\alpha_0 > 0$ et $\alpha_i \geq 0$ pour $i > 0$ (conditions dites de 'régularité').

Le modèle ARCH n'est pas linéaire au sens défini précédemment (voir chapitre sur les processus ARMA, page 51). Il en est de même pour les processus GARCH en général.

Les valeurs successives de ϵ_t sont **décorrélées mais dépendantes**. La décorrélation se démontre aisément de la façon suivante : pour $h > 0$,

$$\mathbb{E}(\epsilon_t \epsilon_{t+h}) = \mathbb{E}(\sigma_t \eta_t \sigma_{t+h} \eta_{t+h}) = \mathbb{E}(\sigma_t \eta_t \sigma_{t+h}) \mathbb{E}(\eta_{t+h}) = 0 = \mathbb{E}(\epsilon_t) \mathbb{E}(\epsilon_{t+h})$$

car η_{t+h} ne dépend pas des trois autres variables présentes dans cette expression.

La dépendance de la séquence ϵ_t se vérifie par la relation $\rho_{\epsilon^2}(1) = \alpha_1$ (voir démonstration page 108) : il y a corrélation entre les valeurs successives de ϵ^2 .

La distribution *conditionnelle* de ϵ_t (i.e. connaissant les valeurs précédentes) est gaussienne, de variance σ_t^2 variable en fonction des valeurs précédentes de ϵ_t (CH = "Conditionnally Heteroskedastic") ; on dit également que la 'volatilité' est variable.

Les séries ϵ_t et σ_t sont des processus aléatoires dont on va étudier la stationnarité (celle de ϵ_t entraînant celle de σ_t) dans le cas d'un ARCH(1).

Les moments d'ordre impair de ϵ_t sont nuls, par symétrie. On démontre par ailleurs que pour un ARCH(1) avec $\alpha_0 > 0$ et $\alpha_1 \geq 0$, alors le moment d'ordre $2r$ de ϵ_t existe si et seulement si

$$\alpha_1^r \prod_{j=1}^r (2j-1) < 1$$

On a en effet :

$$\epsilon_t^2 = (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \eta_t^2$$

soit :

$$\epsilon_t^{2r} = \left(\sum_{i=0}^r C_r^i \alpha_0^{r-i} \alpha_1^{2i} \epsilon_{t-1}^{2i} \right) \eta_t^{2r}$$

Si le moment d'ordre $2r$ de ϵ existe, alors :

$$m_{2r_\epsilon} = \sum_{i=0}^r (C_r^i \alpha_0^{r-i} \alpha_1^{2i} m_{2i_\epsilon}) m_{2r_\eta}$$

soit :

$$[1 - \alpha_1^r m_{2r_\eta}] m_{2r_\epsilon} = m_{2r_\eta} \sum_{i=0}^{r-1} C_r^i \alpha_0^{r-i} \alpha_1^{2i} m_{2i_\epsilon}$$

η_t étant *i.i.d.* $\mathcal{N}(0, 1)$, on a :

$$m_{2r_\eta} = \prod_{j=1}^r (2j-1)$$

d'où la propriété. Ainsi le moment d'ordre 2 (variance) existe si $\alpha_1 < 1$, et la variance de la distribution *non conditionnelle* de ϵ_t est $m_{2_\epsilon} = V_\epsilon = \frac{\alpha_0}{1-\alpha_1}$. Dans ce cas, ϵ_t est stationnaire au sens des séries chronologiques (stationnarité de la covariance).

Le moment d'ordre 4 existe si $3\alpha_1^2 < 1$, soit $\alpha_1 < \frac{1}{\sqrt{3}}$ et le moment d'ordre 4 vaut :

$$m_{4_\epsilon} = 3m_{2_\epsilon}^2 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3m_{2_\epsilon}^2$$

Ainsi le processus ϵ_t présente des "queues de distribution" de sa distribution *non conditionnelle* plus importantes que celles de la distribution *conditionnelle* normale, pour laquelle on a

$m_4 = 3m_2^2$. Cette propriété est illustrée sur des exemples au paragraphe suivant. Remarquons que les $(\epsilon_t)_{t \in \mathbb{Z}}$ ne sont donc pas indépendants (mais ils sont bien décorrélés).

Ce résultat prouve également que ϵ_t dépend du 'passé' $(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$.

Dans le cas d'un $ARCH(p)$, il existe également des conditions pour la stationnarité de la covariance, non développées dans ce papier.

Le processus ϵ_t est dit $GARCH(p, q)$ si :

$$\epsilon_t = \sigma_t \eta_t ,$$

où :

$$\begin{aligned} \eta_t &\sim i.i.d \mathcal{N}(0, 1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \end{aligned}$$

avec $\alpha_0 > 0$, $\alpha_i \geq 0$ pour $i > 0$ et $\beta_j \geq 0$ pour $j > 0$.

Dans ce cas, la variance de ϵ_t dépend en plus des variances des ϵ précédents (σ_{t-j}^2)

Classiquement, pour les processus $GARCH$ rencontrés dans le domaine de la finance, la variable aléatoire η_t peut suivre, au lieu d'une loi gaussienne, une loi de Student¹ à ν degrés de liberté et de variance 1 qui présente des "queues de distribution" plus importantes que la loi gaussienne ; la variable ϵ_t suivra une loi *conditionnelle* du même type.

Les processus $ARMA/GARCH$ sont construits de la même façon que les processus $ARMA$, la séquence Z_t étant remplacée par une séquence de type ϵ_t telle que décrite ci-dessus.

La partie ARMA peut être vue comme définissant la "fonction de moyenne conditionnelle" du processus ARMA/GARCH. Soit par exemple X_t un processus $AR(1)/GARCH$ défini par :

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

et soit $\mu_t = \mathbb{E}(X_t | F_{t-1})$ où F_{t-1} représente l'ensemble des informations accessibles à l'instant $t-1$, et $V_t = \text{Var}(X_t | F_{t-1})$, alors on voit que l'on a dans ce cas $\mu_t = \phi X_{t-1}$ et $V_t = \sigma_t^2$.

Exercice 8.1 Soit X_t un processus $AR(1)$:

$$X_t = \phi X_{t-1} + Z_t$$

où Z_t est *i.i.d.* $\mathcal{N}(0, \sigma_Z^2)$.

Quelle est la loi conditionnelle de X_t , connaissant les valeurs précédentes de X (X_{t-1}, X_{t-2}, \dots) ?

Quelle est la loi non conditionnelle de X_t ?

¹soit une séquence de variables $z_t \sim iid \mathcal{N}(m, \sigma^2)$. Si l'on considère la moyenne échantillonnée sur un échantillon de taille n , soit \hat{m}_n , et l'écart-type échantillonné $\hat{\sigma}_n$, alors la variable $u = \frac{\hat{m}_n - m}{\hat{\sigma}_n} \sqrt{n-1}$ suit une loi de Student à $\nu = n-1$ degrés de liberté : $pdf(u) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu)\sqrt{\pi}\Gamma(\frac{\nu}{2})}} (1 + \frac{u^2}{\nu})^{-\frac{\nu+1}{2}}$, où pdf (Probability Density Function) désigne la loi de distribution de la variable. Cette loi conduit à une variance égale à $\frac{\nu}{\nu-2}$. Pour que η_t soit normalisée ($\text{Var}(\eta_t) = 1$), on prendra $\eta_t = \frac{u}{\sqrt{\frac{\nu}{\nu-2}}}$ et on obtient : $pdf(\eta_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\sqrt{\pi}\Gamma(\frac{\nu}{2})}} (1 + \frac{\eta_t^2}{\nu-2})^{-\frac{\nu+1}{2}}$

Mêmes questions pour un $ARCH(1)$:

$$\begin{aligned}\epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2\end{aligned}$$

Mêmes questions pour un $AR(1)/ARCH(1)$:

$$\begin{aligned}X_t &= \phi X_{t-1} + Z_t \\ \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2\end{aligned}$$

Exercice 8.2 Soit ϵ_t un processus $ARCH(1)$:

$$\begin{aligned}\epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2\end{aligned}$$

où η_t est *i.i.d.* de moyenne nulle.

Trouver une condition nécessaire sur les moments de η_t pour que ϵ_t soit gaussien : $\epsilon_t \sim \mathcal{N}(0, 1)$.

Indication : utiliser la formule du cours qui donne les moments de ϵ en fonction de ceux de η .

Dans les processus ARMA / GARCH, on dit généralement que la partie ARMA décrit la fonction de moyenne conditionnelle, et que la partie GARCH décrit la fonction de variance conditionnelle (voir [17]).

8.3 Propriétés des processus ARMA/GARCH

Les modèles $(G)ARCH$ favorisent le 'regroupement' de fortes variations sur certaines périodes (apparition de corrélations empiriques lorsque la variance est élevée). Ces modèles sont rencontrés par exemple en économie, où les périodes de forte variation des cours (volatilité élevée) sont souvent groupées (phases d'incertitude économique, tension sur les marchés, ...), contrastant avec des périodes calmes. Un alea qui se produit à un instant donné a des conséquences sur la suite du processus, en particulier sa volatilité.

La figure 8.1 permet de comparer deux processus $AR(1)$ avec $\phi = 0.9$, le premier piloté par un bruit *iid*, le deuxième piloté par un bruit blanc $ARCH(1)$ avec $\alpha_1 = 0.9$ (ce processus est appelé $AR(1)/ARCH(1)$).

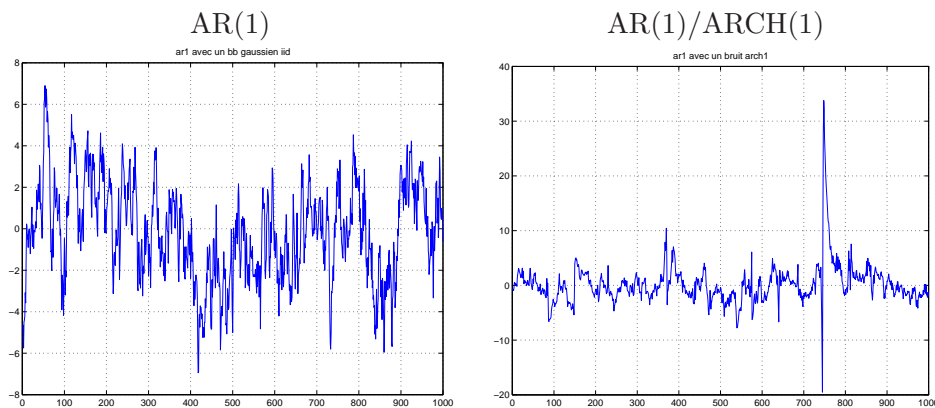


FIG. 8.1 – L'effet "GARCH"

Une série ϵ_t de type *ARCH* est une collection de variables aléatoires décorrélées et *conditionnellement* gaussiennes. Cependant, si l'on considère la distribution *non conditionnelle* du processus stationnaire ϵ_t , celle-ci n'est pas gaussienne, comme indiqué précédemment. La figure suivante (8.2) permet de comparer les distributions empiriques des deux mêmes processus que ci-dessus.

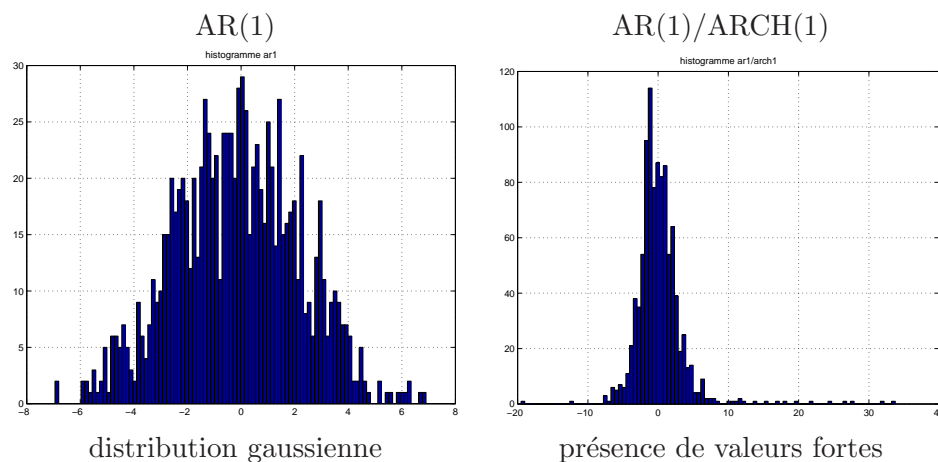


FIG. 8.2 – L'effet "GARCH" sur la densité de probabilité

La fonction de corrélation d'un *ARMA*/*GARCH* est la même que celle d'un *ARMA* de mêmes paramètres piloté par un bruit *iid*. La figure 8.3 montre les corrélations empiriques des deux processus déjà étudiés, on constate qu'elles se ressemblent.

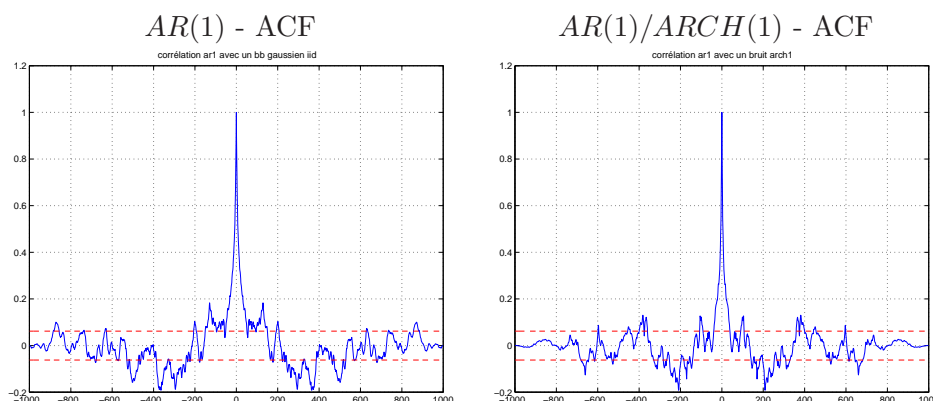


FIG. 8.3 – Fonctions de corrélation

Une 'loupe' sur ces fonctions (figure 8.4) montre plus clairement leur ressemblance au voisinage de 0.

On représente également (figure 8.5) les fonctions de corrélation partielle empirique (PACF) des deux processus.

On retrouve bien pour le processus *AR(1)/ARCH(1)* la PACF d'un processus autorégressif, avec une valeur juste pour $h = 1$, mais la valeur en $h = 2$ n'est pas très petite.

8.4 Principes d'identification

L'identification se compose des quatre étapes suivantes :

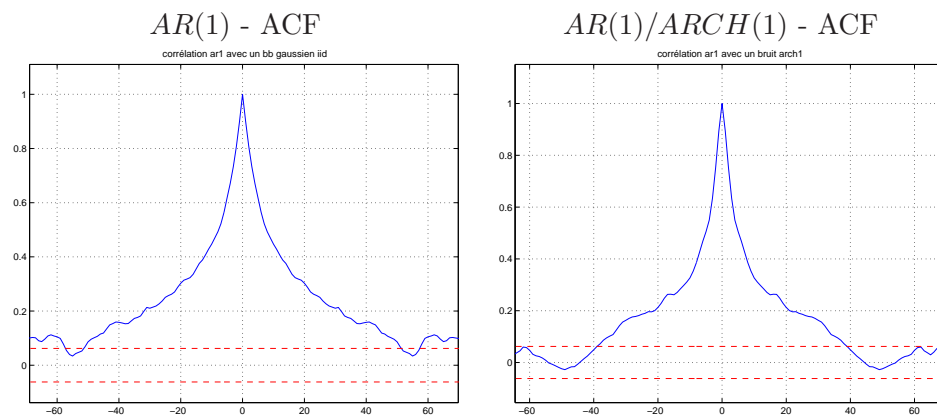


FIG. 8.4 – Fonctions de corrélation 'zoomées'

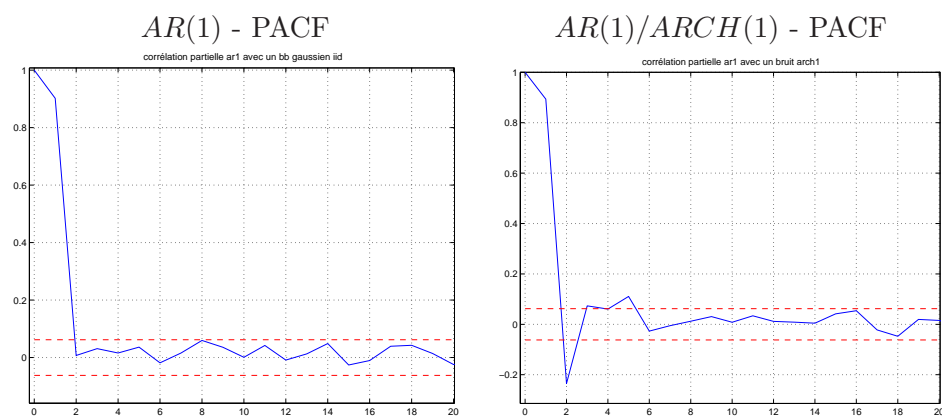


FIG. 8.5 – Fonctions de corrélation partielle

- identifier les paramètres de la partie *ARMA* (i.e. spécifier une fonction de "moyenne conditionnelle"). Ceci peut se faire par les méthodes habituelles des processus *ARMA*, mais peut conduire à des difficultés supplémentaires liées aux propriétés particulières de ϵ_t . En particulier, on dira qu'il n'y a pas de partie *ARMA* si la série satisfait les critères habituels de 'bruit blanc',
- tester la présence d'un effet "*GARCH*" sur le résidu obtenu (voir ci-dessous),
- si un effet "*GARCH*" est détecté, calculer les paramètres du processus *GARCH* (voir ci-dessous),
- tester la validité du modèle (voir ci-dessous) et affiner si nécessaire.

On peut également envisager une identification 'globale' du processus *ARMA/GARCH*.

8.4.1 Tester la présence d'un effet "GARCH" sur le résidu

Il existe plusieurs méthodes, la plus simple est décrite ici.

Si la série de variables décorréelées ϵ_t est affectée d'un effet *ARCH*, celui-ci introduit via la variance σ_t^2 une dépendance qui devient visible sur la fonction de corrélation empirique de la série des ϵ_t^2 .

La figure 8.6 illustre cette propriété.

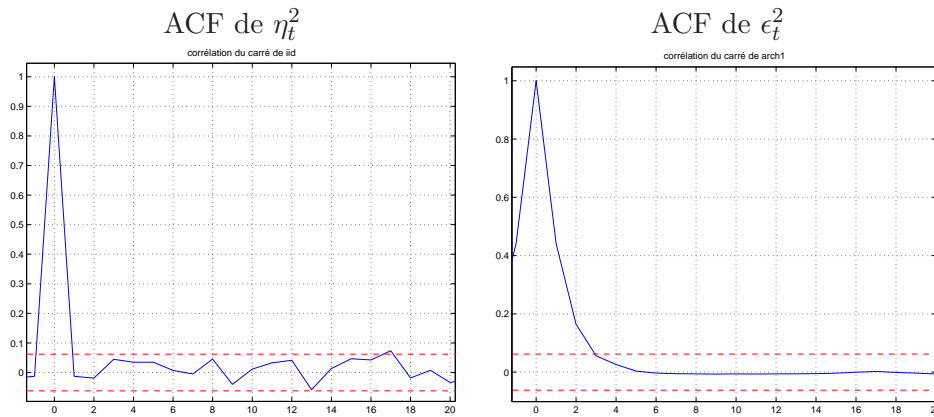


FIG. 8.6 – L'effet "GARCH" sur la corrélation du carré d'une variable aléatoire

Calculons le coefficient de corrélation entre ϵ_{t-1}^2 et ϵ_t^2 , soit $\rho_{\epsilon^2}(1)$.

$$\rho_{\epsilon^2}(1) = \frac{\text{Cov}(\epsilon_t^2, \epsilon_{t-1}^2)}{\text{Var}(\epsilon_t^2)} = \frac{N}{D}$$

avec

$$N = \text{Cov}(\epsilon_t^2, \epsilon_{t-1}^2) = \mathbb{E}(\epsilon_t^2 \epsilon_{t-1}^2) - \mathbb{E}(\epsilon_t^2)^2$$

et

$$D = \mathbb{E}(\epsilon_t^4) - \mathbb{E}(\epsilon_t^2)^2$$

Or on a :

$$\begin{aligned} \epsilon_t &= \sigma_t \eta_t, \\ \epsilon_t^2 &= \sigma_t^2 \eta_t^2 = (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \eta_t^2, \\ \epsilon_t^2 \epsilon_{t-1}^2 &= \alpha_0 \epsilon_{t-1}^2 \eta_t^2 + \alpha_1 \epsilon_{t-1}^4 \eta_t^2, \end{aligned}$$

soit :

$$\begin{aligned} N &= \alpha_0 m_{2_\epsilon} + \alpha_1 m_{4_\epsilon} - m_{2_\epsilon}^2, \\ D &= m_{4_\epsilon} - m_{2_\epsilon}^2, \end{aligned}$$

en utilisant les valeurs trouvées précédemment, soit $m_{2_\epsilon} = \frac{\alpha_0}{1-\alpha_1}$ et $m_{4_\epsilon} = 3m_{2_\epsilon}^2 \frac{1-\alpha_1^2}{1-3\alpha_1^2}$, on trouve finalement :

$$\rho_{\epsilon^2}(1) = \alpha_1$$

qui est différent de zéro si α_1 n'est pas nul, c'est-à-dire lorsqu'il y a un "effet *GARCH*". Cette propriété est reprise au paragraphe suivant pour le calcul des paramètres d'un processus *GARCH*.

8.4.2 Calcul des paramètres d'un processus "GARCH"

On peut envisager deux méthodes, illustrées ci-dessous dans le cas d'un *ARCH*(1).

8.4.2.1 Identifier le carré d'un processus GARCH comme un processus ARMA de moyenne non nulle

Dans le cas d'un *ARCH*(1), on a $\epsilon_t^2 = (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \eta_t^2$, et l'on peut voir que la moyenne non conditionnelle de ϵ_t^2 est égale à $\frac{\alpha_0}{1-\alpha_1}$.

Soit u_t la série donnée par $u_t = \epsilon_t^2 - \frac{\alpha_0}{1-\alpha_1}$, on peut vérifier que u_t satisfait la relation $u_t = \alpha_1 u_{t-1} + v_t$ avec $v_t = \epsilon_t^2 - \sigma_t^2$. On peut vérifier également que les v_t constituent une suite de variables aléatoires non corrélées (mais non *iid*). On peut donc estimer successivement :

- la valeur de α_1 par les techniques classiques d'estimation des processus *ARMA* (paramétriques ou 'robustes'),
- la valeur de α_0 à partir de α_1 et de la moyenne empirique de ϵ_t^2 .

Compte tenu des caractéristiques de v_t (en particulier son caractère non gaussien et sa dépendance à ϵ_t), cette technique donne des résultats approchés (les propriétés statistiques de ces estimateurs n'ont pas été étudiées dans le détail, voir [17] page 121). On constate en particulier dans le cas d'un *ARCH*(1) que, si α_1 se rapproche de 1, son estimation par cette méthode est de qualité médiocre (voir figure de "ACF de ϵ_t^2 " du paragraphe 4.1, où pour $\alpha_1 = 0.9$, on trouve $\widehat{\alpha}_1 = 0.45$; noter que, dans le cas d'un *AR*(1), l'estimation de ϕ par la fonction de corrélation empirique est la même que par le maximum de vraisemblance si l'on suppose le bruit gaussien). Cette technique s'étend au cas d'un processus *GARCH*; on démontre que :

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + v_t + \sum_{j=1}^q (-\beta_j) v_{t-j}$$

qui permet, après le même changement de variable que précédemment ($u_t = \epsilon_t^2 - \mathbb{E}(\epsilon_t^2)$), d'estimer les coefficients α et β à partir d'une identification *ARMA* de u_t .

8.4.2.2 Cas particulier du processus IGARCH

Un processus *GARCH*(1,1) peut être réécrit sous la forme :

$$\epsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 + v_t - \beta_1 v_{t-1}$$

Si $\alpha_1 + \beta_1 = 1$, alors le modèle a une racine unité, et le processus est appelé *IGARCH* (Integrated *GARCH*) : il présente une forte persistance de l'effet des carrés des 'chocs' passés (ϵ_t^2). On a des propriétés voisines si $\alpha_1 + \beta_1$ est proche de 1. On rencontre fréquemment ces processus dans le domaine de la finance (taux de change, indices boursiers).

La figure 8.7 montre une réalisation d'un processus *IGARCH* avec $\alpha_1 = 0.08$ et $\beta_1 = 0.9$, ainsi que les variations de sa variance conditionnelle :

On voit sur ces graphiques que la variance du processus varie lentement. Ces graphiques donnent l'apparence d'un processus non stationnaire, alors qu'il est en fait stationnaire.

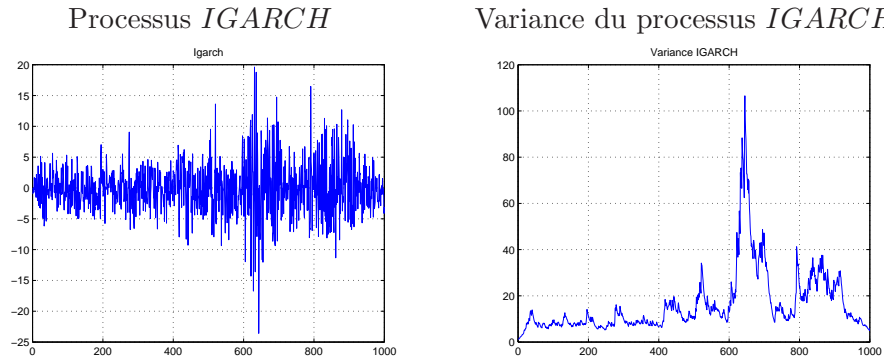


FIG. 8.7 – Processus IGARCH et variance conditionnelle

8.4.2.3 Identifier le processus GARCH par maximum de vraisemblance

Soit $f(\epsilon_2, \dots, \epsilon_n | \alpha_0, \alpha, \epsilon_1)$ la densité de probabilité de $\epsilon_2, \dots, \epsilon_n$ conditionnée par ϵ_1 et pour un couple α_0, α_1 donné. On a dans le cas d'un processus $ARCH(1)$ avec η_t gaussien :

$$f(\epsilon_2, \dots, \epsilon_n | \alpha_0, \alpha_1, \epsilon_1) = \prod_{t=2}^n \frac{1}{\sigma_t \sqrt{2\pi}} \exp - \frac{\epsilon_t^2}{2\sigma_t^2}$$

avec $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$

soit :

$$S = -2 \log(f) = \sum_{t=2}^n \left[\log(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) + \frac{\epsilon_t^2}{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right] + Cte$$

alors on peut déterminer les valeurs de α_0 et α_1 les plus probables par la formule :

$$(\widehat{\alpha}_0, \widehat{\alpha}_1) = \operatorname{argmin}_{\alpha_0, \alpha_1} S(\alpha_0, \alpha_1) = \operatorname{argmin}_{\alpha_0, \alpha_1} \sum_{t=2}^n \left[\log(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) + \frac{\epsilon_t^2}{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right]$$

Dans le cas où η_t est gaussien, cette méthode donne de meilleurs résultats que la précédente (cf paragraphe 4.2.1).

Ce calcul se généralise facilement à un processus $ARCH(p)$.

8.4.3 Tester la validité du modèle (Model Checking)

Si le modèle du processus $ARCH$ a été correctement identifié, alors les résidus "normalisés"

$$\tilde{\epsilon}_t = \frac{\epsilon_t}{\widehat{\sigma}_t}$$

constituent une séquence de variables aléatoires *iid*.

Le modèle $ARCH$ ayant été identifié dans une phase précédente, les valeurs successives de $\widehat{\sigma}_t$ sont données par $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \epsilon_{t-1}^2$

8.4.4 Estimation globale d'un processus ARMA/GARCH

Si l'on a une idée a priori du modèle à estimer, on peut envisager dans certains cas de faire une estimation globale par maximum de vraisemblance. On illustre ceci dans le cas d'un processus $AR(1)/ARCH(1)$.

Dans ce cas, on a trois paramètres à estimer : ϕ pour la partie AR(1), α_0 et α_1 pour la partie ARCH(1). Les valeurs les plus vraisemblables sont données, dans le cas gaussien, par :

$$(\widehat{\phi}, \widehat{\alpha}_0, \widehat{\alpha}_1) = \operatorname{argmin}_{\phi, \alpha_0, \alpha_1} S(\phi, \alpha_0, \alpha_1)$$

avec :

$$S(\phi, \alpha_0, \alpha_1) = \sum_{t=3}^n \left[\log(\alpha_0 + \alpha_1(X_{t-1} - \phi X_{t-2})^2) + \frac{(X_t - \phi X_{t-1})^2}{\alpha_0 + \alpha_1(X_{t-1} - \phi X_{t-2})^2} \right]$$

On pourra tester la validité du modèle de façon analogue à ce qui est décrit au paragraphe 8.4.3 :

- soit $\widehat{\epsilon}_t$ la séquence ARCH(1) estimée : $\widehat{\epsilon}_t = X_t - \widehat{\phi}X_{t-1}$
- soit $\tilde{\epsilon}_t = \frac{\widehat{\epsilon}_t}{\widehat{\sigma}_t}$ avec $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{\epsilon}_{t-1}^2$
- on testera que la série $\tilde{\epsilon}_t$ constitue bien une séquence de variables aléatoires *iid*.

8.5 Les processus M-GARCH

Ce cours n'étudie en détail que les processus univariés. Dans les domaines de l'économie et de la finance, les dépendances entre plusieurs séries peuvent être utilisées pour prendre des décisions d'arbitrage. Il est nécessaire pour cela de considérer des séries multivariées.

Une série temporelle multivariée est une collection de k séries univariées, appelées *composantes*². Dans les conditions de stationnarité d'ordre 2 (stationnarité faible), les covariances et corrélations sont alors des matrices (inter-covariance / corrélation ou cross-covariance / corrélation) :

$$\mathbf{\Gamma}_l \equiv [\Gamma_{ij}(l)] = E[(\mathbf{r}_t - \mu)(\mathbf{r}_{t-l} - \mu)']$$

où μ est le vecteur moyenne de \mathbf{r}_t .

$$\rho_l \equiv [\rho_{ij}(l)] = \mathbf{D}^{-1} \mathbf{\Gamma}_l \mathbf{D}^{-1}$$

où \mathbf{D} est la matrice (k, k) diagonale dont les éléments sont les écarts-type des séries individuelles r_{it} .

On obtient :

$$\rho_{ij}(l) = \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\operatorname{Cov}(r_{it}, r_{j,t-l})}{\operatorname{std}(r_{it})\operatorname{std}(r_{jt})}$$

On définit également les valeurs empiriques de ces matrices : pour l'ensemble de données \mathbf{r}_t , $t = 1, \dots, T$, la valeur empirique de la matrice de covariance est :

$$\widehat{\mathbf{\Gamma}}_l = \frac{1}{T} \sum_{t=l+1}^T (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_{t-l} - \bar{\mathbf{r}})' \quad 0 \leq l \leq T-1$$

où $\bar{\mathbf{r}}$ est la moyenne échantillonnée :

$$\bar{\mathbf{r}} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t$$

La valeur empirique de la matrice de corrélation vaut :

$$\widehat{\rho}_l = \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{\Gamma}}_l \widehat{\mathbf{D}}^{-1} \quad 0 \leq l \leq T-1$$

²une telle série sera notée $\mathbf{r}_t = (r_{1t} r_{2t} \dots r_{kt})'$: la valeur à l'instant t est un vecteur de taille k .

où $\hat{\mathbf{D}}$ est la matrice diagonale (k, k) des écarts-type empiriques des composantes de la série \mathbf{r}_t .

Il est possible de définir et d'identifier des modèles *VARMA* (*ARMA* vectoriels) ; par exemple, un modèle *VAR*(1) comprend les deux équations scalaires suivantes :

$$\begin{aligned} r_{1t} &= \phi_{11}r_{1,t-1} + \phi_{12}r_{2,t-1} + \epsilon_1 , \\ r_{2t} &= \phi_{21}r_{1,t-1} + \phi_{22}r_{2,t-1} + \epsilon_2 , \end{aligned}$$

qui peuvent s'écrire de façon plus concise sous forme matricielle :

$$\mathbf{r}_t = \Phi \mathbf{r}_{t-1} + \epsilon_t$$

où \mathbf{r}_t , ϵ_t sont des vecteurs de dimension 2 et Φ est une matrice $(2, 2)$.

L'intégration dans les formules des modèles *VARMA* de racines égales à l'unité (comme dans les *ARIMA* mono-variés) conduit à la notion de co-intégration.

A partir de ces outils, il est possible d'établir des relations de dépendance temporelle entre les composantes ; par exemple, les variations d'une composante à l'instant t ont des conséquences sur les variations d'une autre composante à l'instant $t + \delta t$.

Dans le cas des processus *M-GARCH* (*GARCH* multivariés), dont la volatilité varie au cours du temps, on constate par exemple que les corrélations entre les indices boursiers de plusieurs places augmentent dans les périodes de forte volatilité (voir [17]).

Conclusion : prédiction et synthèse

Comme indiqué dans l'introduction, un des objectifs, parmi d'autres, de l'étude des séries chronologiques est la prédiction, qui consiste à donner la meilleure valeur estimée du processus pour des instants futurs, compte tenu des valeurs observées précédemment. Ainsi si l'on observe une série x_i aux instants $i = 1, \dots, n$, la prédiction consiste à donner les valeurs de \hat{x}_i pour des instants i tels que $i > n$.

Dans le cas du modèle additif, on a posé

$$x_i = g_i + s_i + w_i$$

où $g_i + s_i$ représente la partie déterministe (tendance plus variations saisonnières) et w_i est la partie aléatoire. Pour la prédiction, on posera donc

$$\hat{x}_i = \hat{g}_i + \hat{s}_i + \hat{w}_i$$

Sans trop détailler, nous allons indiquer les principales propriétés que l'on peut attendre d'une telle prédiction.

La prédiction est fondée sur la modélisation que l'on a su faire de la série pendant la période observée, c'est-à-dire les instants 1 à n . Pour prédire les valeurs ultérieures de la série, on fait ***l'hypothèse forte que le modèle élaboré à partir des observations x_1, \dots, x_n n'est pas modifié pour $i > n$*** , ce qui constitue une première limitation.

Pour la partie déterministe, on a vu au chapitre 3 plusieurs façons de la modéliser, qu'on peut résumer ici en :

- méthodes paramétriques qui supposent que la partie déterministe s'exprime en fonction d'un (petit) nombre de paramètres (par exemple c'est une fonction polynomiale du temps). Dans ce cas, la prédiction se fait aisément en prolongeant cette fonction pour $i > n$, en supposant comme dit précédemment que ce prolongement est justifié,
- méthodes non paramétriques qui sont des estimations fonctionnelles ***locales*** qui par construction ne permettent pas la prédiction, ce qui constitue une deuxième limitation.

Pour la partie aléatoire, après lui avoir trouvé un modèle, on peut mettre en oeuvre les techniques de prédiction développées dans les divers chapitres de ce cours (ARMA, ARIMA, ARFIMA, GARCH). Notons bien que la séquence des valeurs prédites ne ressemble pas à une trajectoire du processus (que l'on peut synthétiser à partir du même modèle), ce qui peut être illustré sur l'exemple simple d'un AR(1). Pour un tel processus, on a vu que le prédicteur à l'instant $n + h$ est $\hat{x}_{n+h} = \phi^h x_n$ qui est une fonction exponentielle qui décroît 'rapidement' vers 0.

Pour résumer cette conclusion, considérons une série dont le modèle 'vrai' est la somme d'une fonction linéaire et d'un processus AR(1), soit

$$x_i = a + bi + w_i$$

avec

$$w_i = \phi w_{i-1} + Z_i$$

alors le prédicteur pour $i > n$ est donné par

$$\hat{x}_i = \hat{a} + \hat{b}i + \hat{\phi}^{i-n}(x_n - \hat{a} - \hat{b}n)$$

alors qu'une trajectoire synthétisée pour $i > n$ est donnée par

$$tr_i = \hat{a} + \hat{b}i + u_i$$

avec

$$u_i = \hat{\phi}u_{i-1} + Z_i$$

et

$$u_n = x_n - \hat{a} - \hat{b}n$$

et Z_i est une trajectoire quelconque du bruit blanc Z pour $i > n$.

Correction des exercices

Vous trouverez ici une correction de la plupart des exercices de ce cours.

Solution 1.5:

1. Soit Z_t un bruit blanc. On veut observer $Y_t = x_t + Z_t$ avec $x_t = at + b$. On remarque que $x_{t+1} = x_t + a$. Donc on peut prendre :

$$X_t = \begin{pmatrix} a \\ x_t \end{pmatrix} ,$$

initialisation :

$$X_0 = \begin{pmatrix} a \\ b \end{pmatrix} ,$$

équation d'état :

$$X_{t+1} = FX_t ,$$

avec

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} ,$$

et la matrice de covariance du bruit d'état est nulle. Observation :

$$Y_t = GX_t + Z_t ,$$

avec

$$G = \begin{pmatrix} 0 & 1 \end{pmatrix} ,$$

et la matrice de covariance du bruit de mesure vaut $R = \sigma^2$

2. Soit le vecteur d'état :

$$X_t := Y_t ,$$

initialisation :

$$X_0 = Y_0 ,$$

équation d'état :

$$X_{t+1} = FX_t + Z_{t+1} ,$$

avec

$$F = \phi ,$$

et la matrice de covariance du bruit d'état vaut $Q = \sigma^2$. Equation d'observation :

$$Y_t = GX_t ,$$

avec

$$G = 1 ,$$

et la matrice de covariance du bruit de mesure est nulle.

3. On veut observer $U_t = x_t + Y_t$ avec Y un $AR(1)$ et $x_t = at + b$. Il suffit de combiner les principes des deux premiers exemples. Soit le vecteur d'état :

$$X_t = \begin{pmatrix} a \\ x_t \\ Y_t \end{pmatrix},$$

initialisation :

$$X_0 = \begin{pmatrix} a \\ b \\ Y_0 \end{pmatrix},$$

équation d'état :

$$X_{t+1} = FX_t + \begin{pmatrix} 0 \\ 0 \\ Z_{t+1} \end{pmatrix},$$

avec

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \phi \end{pmatrix},$$

observation :

$$W_t = GX_t,$$

avec

$$G = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

□

Solution 1.6: On modélise d'abord la partie AR , que l'on note U . Soit :

$$X_t := \begin{pmatrix} U_{t-1} \\ U_t \end{pmatrix},$$

équation d'état :

$$X_{t+1} = FX_t + \begin{pmatrix} 0 \\ Z_{t+1} \end{pmatrix},$$

avec

$$F = \begin{pmatrix} 0 & 1 \\ 0 & \phi \end{pmatrix},$$

ainsi, U vérifie :

$$U_t - \phi U_{t-1} = Z_t.$$

Et on obtient Y en remarquant que :

$$Y_t = U_t + \theta U_{t-1}.$$

D'où l'équation d'observation :

$$Y_t = GX_t,$$

avec

$$G = \begin{pmatrix} \theta & 1 \end{pmatrix}.$$

□

Solution 1.7:

1. P_X est à valeurs réelles :

$$\begin{aligned}
 \overline{P_X(\omega)} &= \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \overline{\gamma_X(t) e^{-jt\omega}} , \\
 &= \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \overline{\gamma_X(t)} e^{jt\omega} , \\
 &= \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma_X(-t) e^{-jt\omega} , \\
 &= \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma_X(t) e^{-jt\omega} , \\
 &= P_X(\omega) ,
 \end{aligned}$$

car $\gamma_X(k) \in \mathbb{R}$.

2. La convergence de $\sum_{n \in \mathbb{Z}} |\gamma_X(n)|$ permet d'intervertir somme et intégrale dans ce qui suit :

$$\begin{aligned}
 \int_{-\pi}^{\pi} P_X(\omega) e^{jk\omega} d\omega &= \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma_X(t) \int_{-\pi}^{\pi} e^{-jt\omega} e^{jk\omega} d\omega , \\
 &= \gamma_X(k) ,
 \end{aligned}$$

Cela implique que la variance d'un processus aléatoire stationnaire X , $\gamma_X(0)$ est égale à la moyenne de sa densité spectrale de puissance.

3. Pour un bruit blanc,

$$P_X(\omega) = \frac{1}{2\pi} \gamma_X(0), \quad \forall \omega \in [-\pi, \pi] .$$

□

Solution 1.1: (a) On trouve $\mathbb{E}(X_t) = a$, et

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases} b^2 + c^2 & \text{si } k = 0 \\ bc & \text{si } |k| = 1 , \\ 0 & \text{sinon.} \end{cases}$$

Donc le processus (X_t) est stationnaire.

(b) On trouve $\mathbb{E}(X_t) = a$, et

$$\text{Cov}(X_t, X_{t+k}) = b^2 .$$

Donc le processus (X_t) est stationnaire.

(c) On trouve $\mathbb{E}(X_t) = 0$, et

$$\text{Cov}(X_t, X_{t+k}) = \cos(ct) \cos(c(t+k)) + \sin(ct) \sin(c(t+k)) = \cos(ck) .$$

Donc le processus (X_t) est stationnaire.

(d) On trouve $\mathbb{E}(X_t) = 0$, et

$$\text{Cov}(X_t, X_{t+k}) = \cos(ct) \cos(c(t+k)) .$$

Donc le processus (X_t) n'est pas stationnaire (sauf si c est un multiple de π).

(e) On trouve $\mathbb{E}(X_t) = 0$, et

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases} 1 & \text{si } k = 0 \\ \cos(ct) \sin(c(t+1)) & \text{si } k = 1 , \\ \cos(c(t-1)) \sin(ct) & \text{si } k = -1 , \\ 0 & \text{sinon.} \end{cases}$$

Donc le processus (X_t) n'est pas stationnaire (sauf si c est un multiple de π).

(f) On trouve $\mathbb{E}(X_t) = 0$, et

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{sinon.} \end{cases}$$

Donc le processus (X_t) est stationnaire.

□

Solution 1.8: On calcule les moyennes et les covariances de (X_t) :

$$\mathbb{E}(X_t) = a\mathbb{E}(X_{t-1}) + m, \quad t \geq 1,$$

D'où :

$$\mathbb{E}(X_t) - \frac{m}{1-a} = a \left(\mathbb{E}(X_{t-1}) - \frac{m}{1-a} \right), \quad t \geq 1,$$

et donc :

$$\mathbb{E}(X_t) - \frac{m}{1-a} = a^t \left(\mathbb{E}(X_0) - \frac{m}{1-a} \right), \quad t \geq 1.$$

Donc la moyenne de X_t est indépendante de t si et seulement si

$$\mathbb{E}(X_0) = \frac{m}{1-a}.$$

Supposons donc que c'est le cas. On calcule ensuite, pour $t \geq 0$ et $k \geq 0$,

$$X_t = \sum_{i=0}^{t-1} a^i U_{t-i} + a^t X_0 = \sum_{j=1}^t a^{t-j} U_j + a^t X_0.$$

En utilisant le fait que U_t soit indépendant de X_0 pour tout $t \geq 1$, et le fait que les U_i soient indépendants entre eux,

$$\begin{aligned} \text{Cov}(X_t, X_{t+k}) &= \sum_{j=1}^t a^{2(t-j)=k} \sigma^2 + a^{2t+k} \text{Var}(X_0), \\ &= \sum_{i=0}^{t-1} a^{2i+k} \sigma^2 + a^{2t+k} \text{Var}(X_0), \\ &= a^k \left(\sigma^2 \frac{1-a^{2t}}{1-a^2} + a^{2t} \text{Var}(X_0) \right), \\ &= a^k \left(\frac{\sigma^2}{1-a^2} + a^{2t} \left(\text{Var}(X_0) - \frac{\sigma^2}{1-a^2} \right) \right), \end{aligned}$$

et si l'on veut que ces covariances ne dependent que de k , il faut et il suffit que :

$$\text{Var}(X_0) = \frac{\sigma^2}{1-a^2}.$$

On obtient ainsi les deux conditions voulues, et en plus que :

$$\gamma_X(k) = a^k \frac{\sigma^2}{1-a^2}.$$

□

Solution 2.2: On utilise la propriété qu'une fonction K absolument sommable est la fonction d'autocovariance d'un processus stationnaire si et seulement si sa transformée de Fourier est positive :

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} K(n) \geq 0 \quad \forall \lambda \in [-\pi, \pi] .$$

Or, $f(\lambda) = \frac{1}{2\pi}[1 + 2r \cos \lambda]$ qui est clairement toujours positive si et seulement si $|r| < 1/2$. \square

Solution 2.3: Comme la série $\sum_{i \in \mathbb{Z}} |\psi_i|$ converge, on peut inverser l'espérance et la somme dans ce qui suit :

$$\mathbb{E}(X_t) = \sum_{i \in \mathbb{Z}} \psi_i \mathbb{E}(Y_{t-i}) = 0,$$

et :

$$\begin{aligned} \mathbb{E}(X_{t+h} \overline{X_t}) &= \sum_{i, l \in \mathbb{Z}} \psi_i \overline{\psi_l} \mathbb{E}(Y_{t+h-i} \overline{Y_{t-l}}) , \\ &= \sum_{i, l \in \mathbb{Z}} \psi_i \overline{\psi_l} \gamma_Y(h + l - i) , \end{aligned}$$

qui ne dépend pas de t . Donc X est stationnaire. De plus, la densité spectrale de X vaut :

$$\begin{aligned} P_X(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-jh\omega} , \\ &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \sum_{i, l \in \mathbb{Z}} \psi_i \overline{\psi_l} \gamma_Y(h + l - i) e^{-jh\omega} , \\ &= \sum_{i, l \in \mathbb{Z}} \psi_i e^{-ji\omega} \overline{\psi_l} e^{jl\omega} \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_Y(h + l - i) e^{-j(h+l-i)\omega} , \\ &= P_Y(\omega) \sum_{i, l \in \mathbb{Z}} \psi_i e^{-ji\omega} \overline{\psi_l} e^{jl\omega} , \\ &= |\psi(e^{-j\omega})|^2 P_Y(\omega) . \end{aligned}$$

\square

Solution 2.5: On calcule :

$$\begin{aligned} \langle \mathbf{e}_k, \mathbf{e}_l \rangle &= \frac{1}{n} \sum_{i=1}^n e^{ji \frac{2k\pi}{n}} e^{-ji \frac{2l\pi}{n}} , \\ &= \frac{1}{n} \sum_{i=1}^n \left(e^{j \frac{2(k-l)\pi}{n}} \right)^i , \\ &= \mathbb{I}_{\{k=l\}} . \end{aligned}$$

\square

Solution 2.6:

$$\begin{aligned}
I_n(\omega_k) &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n X_t e^{-jt\omega_k} X_s^* e^{js\omega_k}, \\
&= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n X_t X_s^* e^{-j(t-s)\omega_k}, \\
&= \frac{1}{n} \sum_{s=1}^n \sum_{h=s-n}^{s-1} X_{s+h} X_s^* e^{-jh\omega_k}, \\
&= \frac{1}{n} \sum_{|h|<n} \sum_{s=\max\{1, h+1\}}^{\min\{n, h+n\}} X_{s+h} X_s^* e^{-jh\omega_k}, \\
&= \frac{1}{n} \sum_{|h|<n} \gamma_{nc}(h).
\end{aligned}$$

□

Solution 2.8: Calculons d'abord la transformée de Fourier discrète $\hat{X}_n(\lambda)$ du vecteur (X_1, \dots, X_n) .

$$\begin{aligned}
\hat{X}_n(\lambda) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n e^{it\omega} e^{-it\lambda}, \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^n e^{it(\omega-\lambda)},
\end{aligned}$$

où on reconnaît la somme partielle d'une série géométrique. Si $\lambda = \omega$,

$$\hat{X}_n(\lambda) = \sqrt{n}.$$

Si $\lambda \neq \omega$,

$$\begin{aligned}
\hat{X}_n(\lambda) &= \frac{1}{\sqrt{n}} \frac{e^{in(\omega-\lambda)} - 1}{e^{i(\omega-\lambda)} - 1}, \\
&= \frac{e^{in(\omega-\lambda)/2}}{\sqrt{n} e^{i(\omega-\lambda)/2}} \frac{\sin(n(\omega-\lambda)/2)}{\sin((\omega-\lambda)/2)}.
\end{aligned}$$

On obtient ainsi le périodogramme : si $\omega_k = \frac{2k\pi}{n}$,

$$I_n(\omega_k) = \begin{cases} n & \text{si } \omega_k = \omega \\ \frac{\sin(n(\omega-\omega_k)/2)^2}{n \sin((\omega-\omega_k)/2)^2} & \text{sinon.} \end{cases}.$$

Remarquons que si la fréquence ω est une des fréquences ω_k , alors le périodogramme est nul partout sauf pour $\omega_k = \omega$. Dans le cas contraire, le périodogramme ne donne plus zéro. □

Solution 3.1:

1. Calculons la moyenne empirique :

$$\begin{aligned}
\bar{x}_n &= \frac{1}{n} \sum_{t=1}^n a + bt, \\
&= a + b \frac{n+1}{2}.
\end{aligned}$$

Puis l'autocovariance empirique $\hat{\gamma}_n(k)$:

$$\begin{aligned}
 \hat{\gamma}_n(k) &= \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x}_n)(x_{t+k} - \bar{x}_n) , \\
 &= \frac{b^2}{n} \sum_{t=1}^{n-k} \left(t - \frac{n+1}{2}\right) \left(t+k - \frac{n+1}{2}\right) , \\
 &= \frac{b^2}{n} \left[\sum_{t=1}^{n-k} (t^2 + kt) - (n+1) \sum_{t=1}^{n-k} t - \frac{n+1}{2} \sum_{t=1}^{n-k} k + \left(\frac{n+1}{2}\right)^2 \sum_{t=1}^{n-k} 1 \right] , \\
 &= \frac{b^2}{n} \left[\frac{(n-k)(n-k+1)(2(n-k)+1)}{6} + k \frac{(n-k)(n-k+1)}{2} \right. \\
 &\quad \left. - (n+1) \frac{(n-k)(n-k+1)}{2} - k \frac{n+1}{2} (n-k) + \left(\frac{n+1}{2}\right)^2 (n-k) \right] ,
 \end{aligned}$$

où on a utilisé le fait que $\sum_{t=1}^n t^2 = \frac{n(n+1)(2n+1)}{6}$.

$$\begin{aligned}
 \hat{\gamma}_n(k) &= \frac{b^2(n-k)}{n} \left[(n-k+1) \left(\frac{(2(n-k)+1)}{6} + \frac{k}{2} - \frac{(n+1)}{2} \right) \right. \\
 &\quad \left. - k \frac{n+1}{2} + \left(\frac{n+1}{2}\right)^2 \right] , \\
 &= \frac{b^2(n-k)}{n} \left[\frac{n^2}{12} + nP_1(k) + P_2(k) \right] ,
 \end{aligned}$$

où P_1 est un polynôme de degré 1 et P_2 un polynôme de degré 2. Ainsi,

$$\hat{\rho}_n(k) = \frac{\hat{\gamma}_n(k)}{\hat{\gamma}_n(0)} = \frac{(n-k)}{n} \frac{\left[\frac{n^2}{12} + nP_1(k) + P_2(k) \right]}{\left[\frac{n^2}{12} + nP_1(0) + P_2(0) \right]} ,$$

qui tend bien vers 1 pour tout k fixé. On voit même que si $k = \alpha n$ pour α assez petit, $\hat{\rho}_n(k) = \hat{\rho}_n(\alpha n)$ tend vers une limite strictement positive lorsque n tend vers l'infini.

2. Par hypothèse, on a :

$$\hat{\gamma}_{X,n}(k) = (\hat{\gamma}_{x,n}(k) + \hat{\gamma}_{Y,n}(k))(1 + o_P(1)) .$$

Comme $\hat{\gamma}_{x,n}(k)$ est équivalent à $\frac{bn^2}{12}$, et que $\hat{\gamma}_{Y,n}(k)$ converge en loi par hypothèse, on a :

$$\hat{\rho}_{X,n}(k) = \frac{(bn^2(1 + o(1)) + O_P(1))(1 + o_P(1))}{(bn^2(1 + o(1)) + O_P(1))(1 + o_P(1))} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1 .$$

□

Solution 3.2:

1. On calcule :

$$\begin{aligned}\frac{\partial S_{lin}(a_0, a_1)}{\partial a_0} &= -2 \sum_{i=1}^n (x_i - a_0 - a_1 i), \\ &= 2 \left(a_1 \sum_{i=1}^n i + n a_0 - \sum_{i=1}^n x_i \right), \\ \frac{\partial S_{lin}(a_0, a_1)}{\partial a_1} &= -2 \sum_{i=1}^n i (x_i - a_0 - a_1 i), \\ &= 2 \left(a_1 \sum_{i=1}^n i^2 + a_0 \sum_{i=1}^n i - \sum_{i=1}^n i x_i \right)\end{aligned}$$

On trouve :

$$\begin{aligned}\hat{a}_1 &= \frac{Cov(t, x)}{var(t)}, \\ \hat{a}_0 &= \bar{x} - \hat{a}_1 \bar{t},\end{aligned}$$

où :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, & \bar{t} &= \frac{1}{n} \sum_{i=1}^n i, \\ Cov(t, x) &= \frac{1}{n} \sum_{i=1}^n i x_i - \bar{x} \bar{t}, & var(t) &= \frac{1}{n} \sum_{i=1}^n i^2 - \bar{t}^2.\end{aligned}$$

2. Le modèle est alors :

$$x_i = a_0 + a_1 i + \sigma Z_i \quad \forall i = 1, \dots, n,$$

où (Z_1, \dots, Z_n) est un vecteur gaussien centré réduit. La loi de l'observation $x = (x_1, \dots, x_n)$ est alors celle d'un vecteur gaussien, de matrice de covariance $\sigma^2 I_n$ et de moyenne $(a_0 + a_1 i)_{i=1, \dots, n}$. On est donc dans un modèle statistique $(\mathbb{P}_\theta)_{\theta \in \Theta}$ d'ensemble de paramètres $\Theta = \{(a_0, a_1, \sigma^2) \in \mathbb{R}^3\}$ où la loi $\mathbb{P}_{(a_0, a_1, \sigma^2)}$ est celle d'un vecteur gaussien, de matrice de covariance $\sigma^2 I_n$ et de moyenne $(a_0 + a_1 i)_{i=1, \dots, n}$. Ce modèle est dominé par la mesure de Lebesgue λ sur \mathbb{R}^n . En notant $f_{(a_0, a_1, \sigma^2)}$ la densité de la loi $\mathbb{P}_{(a_0, a_1, \sigma^2)}$ par rapport à λ , on trouve la vraisemblance :

$$f_{(a_0, a_1, \sigma^2)}(x) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a_0 - a_1 i)^2}.$$

Donc :

$$f_{(a_0, a_1, \sigma^2)}(x) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{\sigma^2} S_{lin}(a_0, a_1)},$$

et le paramètre (a_0, a_1, σ^2) qui maximise la vraisemblance (le "plus probable") pour l'observation x donnée est bien tel que (a_0, a_1) minimise S_{lin} . Plus précisément, on trouve aussi que :

$$\hat{\sigma}^2 = \frac{1}{n} S_{lin}(\hat{a}_0, \hat{a}_1).$$

3. Le modèle devient :

$$x_i = a_0 + a_1 i + \sqrt{V_i} Z_i \quad \forall i = 1, \dots, n,$$

où (Z_1, \dots, Z_n) est un vecteur gaussien centré réduit. La loi de l'observation $x = (x_1, \dots, x_n)$ est alors celle d'un vecteur gaussien, de matrice de covariance Γ_V et de moyenne $(a_0 + a_1 i)_{i=1, \dots, n}$, avec :

$$\Gamma = \begin{pmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_n \end{pmatrix}$$

En notant $V = (V_1, \dots, V_n)$, on est donc dans un modèle statistique $(\mathbb{P}_\theta)_{\theta \in \Theta}$ d'ensemble de paramètres $\Theta = \{(a_0, a_1) \in \mathbb{R}^2\}$ où la loi $\mathbb{P}_{(a_0, a_1)}$ est celle d'un vecteur gaussien, de matrice de covariance Γ et de moyenne $(a_0 + a_1 i)_{i=1, \dots, n}$. Ce modèle est dominé par la mesure de Lebesgue λ sur \mathbb{R}^n . En notant $f_{(a_0, a_1)}$ la densité de la loi $\mathbb{P}_{(a_0, a_1)}$ par rapport à λ , on trouve la vraisemblance :

$$f_{(a_0, a_1)}(x) = \frac{1}{\prod_{i=1}^n \sqrt{2\pi V_i}} e^{-\sum_{i=1}^n \frac{(x_i - a_0 - a_1 i)^2}{V_i}}.$$

Donc :

$$f_{(a_0, a_1)}(x) = \frac{1}{\prod_{i=1}^n \sqrt{2\pi V_i}} e^{-\frac{1}{\sigma^2} S_{lin, V}(a_0, a_1)},$$

avec :

$$S_{lin, V}(a_0, a_1) = \sum_{i=1}^n \frac{(x_i - a_0 - a_1 i)^2}{V_i}.$$

Le paramètre $((\hat{a}_0)_v, (\hat{a}_1)_V)$ qui maximise la vraisemblance est celui qui minimise $S_{lin, V}$. On calcule :

$$\begin{aligned} \frac{\partial S_{lin}(a_0, a_1)}{\partial a_0} &= -2 \sum_{i=1}^n \frac{(x_i - a_0 - a_1 i)}{V_i}, \\ &= 2 \left(a_1 \sum_{i=1}^n \frac{i}{V_i} + a_0 \sum_{i=1}^n \frac{1}{V_i} - \sum_{i=1}^n \frac{x_i}{V_i} \right), \\ \frac{\partial S_{lin}(a_0, a_1)}{\partial a_1} &= -2 \sum_{i=1}^n i \frac{(x_i - a_0 - a_1 i)}{V_i}, \\ &= 2 \left(a_1 \sum_{i=1}^n \frac{i^2}{V_i} + a_0 \sum_{i=1}^n \frac{i}{V_i} - \sum_{i=1}^n \frac{i x_i}{V_i} \right) \end{aligned}$$

On trouve :

$$\begin{aligned} (\hat{a}_1)_V &= \frac{Cov_V(t, x)}{var_V(t)}, \\ (\hat{a}_0)_V &= \bar{x}_V - (\hat{a}_1)_V \bar{t}_V, \end{aligned}$$

où :

$$\begin{aligned} \bar{x}_V &= \frac{1}{n} \sum_{i=1}^n \frac{x_i}{V_i}, & \bar{t}_V &= \frac{1}{n} \sum_{i=1}^n \frac{i}{V_i}, \\ Cov_V(t, x) &= \frac{1}{n} \sum_{i=1}^n \frac{i x_i}{V_i} - \bar{x}_V \bar{t}_V, & var_V(t) &= \frac{1}{n} \sum_{i=1}^n \frac{i^2}{V_i} - (\bar{t}_V)^2. \end{aligned}$$

4. La méthode des moindres carrés est facile à mettre en oeuvre, et correspond à la méthode du maximum de vraisemblance dans le cas gaussien, tant que le modèle s'écrit comme le fait que la moyenne du vecteur appartient à un certain sous-espace vectoriel (ou affine) de \mathbb{R}^n . Cela suppose que le modèle s'écrit de façon linéaire (ou affine) en les paramètres. Si le modèle ne s'écrit pas de façon linéaire, on peut toujours essayer de minimiser les moindres carrés, mais l'annulation des dérivées partielles n'est plus un problème linéaire (quoique parfois, on puisse s'y ramener). Il n'est même pas évident que le minimum soit atteint en un point unique.

□

Solution 3.3:

1. En reprenant les notations précédentes, on peut écrire :

$$f_A(x) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}(x-IA^T)^T(x-IA^T)}$$

et :

$$\hat{A} = \operatorname{argmax}_A f_A(x) = \operatorname{argmin}_A S \text{ avec } S = (x - IA^T)^T(x - IA^T)$$

Par annulation de la dérivée de S par rapport à A , c'est-à-dire à chacune des composantes de A , on obtient $I^T(x - I\hat{A}^T) = 0$, soit le résultat :

$$\hat{A} = (I^T I)^{-1} I^T x$$

2. Dans ce cas, la formule précédente devient :

$$f_A(x) = \frac{1}{\sqrt{\operatorname{Det}(C_w)}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(x-IA^T)^T C_w^{-1}(x-IA^T)}$$

et :

$$\hat{A} = \operatorname{argmin}_A S \text{ avec } S = (x - IA^T)^T C_w^{-1}(x - IA^T)$$

La même méthode que précédemment conduit au résultat :

$$\hat{A} = (I^T C_w^{-1} I)^{-1} I^T C_w^{-1} x$$

□

Solution 3.5:

1. Comme Y_t est combinaison linéaire des Z_t , alors par linéarité de l'espérance, on a immédiatement que, pour tout $t \in \mathbb{Z}$, $\mathbb{E}[Y_t] = 0$.

On peut écrire Y_t sous la forme suivante :

$$\forall t \in \mathbb{Z}, \quad Y_t = \sum_{j=-\infty}^{+\infty} \theta_j Z_{t+j},$$

en posant $\theta_j = 0$ pour $j < -m_1$ ou $j > m_2$. Alors, pour tout $t \in \mathbb{Z}$ et tout $k \in \mathbb{Z}$,

$$Y_t Y_{t+k} = \left(\sum_{j=-\infty}^{+\infty} \theta_j Z_{t+j} \right) \left(\sum_{h=-\infty}^{+\infty} \theta_h Z_{t+k+h} \right) = \sum_{(j,h) \in \mathbb{Z}^2} \theta_j \theta_h Z_{t+j} Z_{t+k+h}.$$

D'où,

$$\operatorname{cov}[Y_t, Y_{t+k}] = \mathbb{E}[Y_t Y_{t+k}] = \sigma^2 \sum_{\substack{(j,h) \in \mathbb{Z}^2 \\ j=h+k}} \theta_j \theta_h = \sigma^2 \sum_{j \in \mathbb{Z}} \theta_j \theta_{j-k}.$$

Cette covariance ne dépend pas de t mais uniquement de k . Donc, on peut poser $\operatorname{cov}[Y_t, Y_{t+k}] = \gamma(k)$.

2. M est un opérateur linéaire, donc $MX_t = Mm_t + Ms_t + MZ_t = m_t + Y_t$. Le bruit dans la représentation additive de MX_t est Y_t . Il est bien stationnaire d'après la question précédente, mais ce n'est pas un bruit blanc, dès que la moyenne mobile n'est pas triviale, i.e dès que deux valeurs des θ_i sont non nulles. La variance de ce bruit est $\gamma(0)$, la variance de Y_t pour tout $t \in \mathbb{Z}$:

$$\gamma(0) = \operatorname{var}[Y_t] = \sigma^2 \sum_{j \in \mathbb{Z}} \theta_j^2,$$

d'après l'équation (3.2). Ainsi, une moyenne mobile transforme la variance d'un bruit blanc d'un facteur τ égal à :

$$\tau = \frac{\text{var}[Y_t]}{\text{var}[Z_t]} = \sum_{j \in \mathbb{Z}} \theta_j^2.$$

La variance sera donc réduite si $\tau < 1$.

3. Pour finir, on a :

$$\sum_{k \in \mathbb{Z}} \gamma(k) z^k = \sigma^2 \sum_{k \in \mathbb{Z}} \theta_j \theta_{j-k} z^j z^{k-j} = \sigma^2 M(z) M\left(\frac{1}{z}\right).$$

□

Solution 3.7: Il suffit de regarder l'effet de $\nabla = (I - B)$, l'opérateur de différentiation, sur $m_t = t^k$. On a :

$$(\nabla m)_t = m_t - m_{t-1} = t^k - (t-1)^k = - \sum_{j=0}^{k-1} \binom{k}{j} t^j (-1)^{k-j},$$

et on voit que ∇ transforme un polynôme de degré k en un polynôme de degré $(k-1)$. Ainsi, il suffit d'appliquer ∇^k à une série à tendance polynomiale d'ordre k pour réduire cette tendance à une constante et il suffit d'appliquer ∇^{k+1} à une série à tendance polynomiale d'ordre k pour réduire cette tendance à 0. Sur un bruit blanc Z de moyenne nulle et variance σ^2 , on remarque que l'espérance de $\nabla^k Z$ est nulle, et que :

$$\begin{aligned} \text{Var}((\nabla^k Z)_t) &= \mathbb{E} \left[\left(\sum_{j=0}^k \binom{k}{j} (-B)^j Z_t \right)^2 \right], \\ &= \mathbb{E} \left[\left(\sum_{j=0}^k (-1)^j \binom{k}{j} Z_{t-j} \right)^2 \right], \\ &= \sum_{j=0}^k \binom{k}{j}^2 \mathbb{E} [Z_{t-j}^2], \\ &= \sum_{j=0}^k \binom{k}{j}^2 \sigma^2 = \binom{2k}{k} \sigma^2. \end{aligned}$$

L'application de ∇^k multiplie donc la variance du bruit par $\binom{2k}{k}$. Cette méthode ne peut donc pas s'appliquer pour k trop grand. □

Solution 3.8: On cherche le minimum, en (a, b, γ) , de $\sum_{t=1}^n (X_t - ae^{bt} - \gamma)^2$. La recherche d'un point critique donne le système suivant d'équations :

$$\begin{cases} 0 &= a \sum_{t=1}^n e^{2bt} + \gamma \sum_{t=1}^n e^{bt} - \sum_{t=1}^n e^{bt} X_t \\ 0 &= a \sum_{t=1}^n t e^{2bt} + \gamma \sum_{t=1}^n t e^{bt} - \sum_{t=1}^n t e^{bt} X_t \\ 0 &= a \sum_{t=1}^n e^{bt} + \gamma \sum_{t=1}^n e^{bt} - \sum_{t=1}^n e^{bt} X_t \end{cases}$$

Ce système, qui est linéaire en (a, γ) n'est pas linéaire en b . On ne connaît pas de solution simple. Par contre, si on connaît trois points équidistants en abscisse et supposés appartenir

à la courbe, il est possible de déterminer algébriquement ses paramètres. En effet, si on note t_1, t_2, t_3 trois abscisses telles que $t_2 - t_1 = t_3 - t_2 := \delta \neq 0$, et y_1, y_2, y_3 trois ordonnées telles que les (t_i, y_i) soient sur la courbe d'équation $y = ae^{bt} + \gamma$, on a le système d'équations :

$$\begin{cases} y_1 &= ae^{bt_1}e^{-b\delta} + \gamma \\ y_2 &= ae^{bt_2} + \gamma \\ y_3 &= ae^{bt_2}e^{b\delta} + \gamma \end{cases}$$

De la seconde équation, on tire $ae^{bt_2} = y_2 - \gamma$, qu'on reporte dans la première et la dernière :

$$\begin{cases} (y_2 - \gamma)e^{-b\delta} + \gamma = y_1 \\ ae^{bt_2} &= y_2 - \gamma \\ (y_2 - \gamma)e^{b\delta} + \gamma &= y_3 \end{cases}$$

En reportant la valeur de $e^{b\delta}$ donnée par la dernière équation dans la première, on a :

$$\frac{(y_2 - \gamma)^2}{(y_3 - \gamma)} = y_1 - \gamma .$$

Ce qui donne :

$$y_2^2 - 2\gamma y_2 + \gamma^2 = y_1 y_3 - \gamma(y_1 + y_3) + \gamma^2 ,$$

et donc :

$$\gamma = \frac{y_2^2 - y_1 y_3}{2y_2 - y_1 - y_3} ,$$

puis

$$b = \frac{1}{\delta} \log \frac{y_3 - \gamma}{y_2 - \gamma} ,$$

et :

$$a = \frac{y_2 - \gamma}{e^{bt_2}} .$$

□

Solution 3.9:

1. Il suffit de différentier à une période p : on définit $\Delta_p = I - B^p$. Alors,

$$(\Delta_p X)_t = (\Delta_p m)_t + (\Delta_p s)_t + (\Delta_p W)_t ,$$

Le terme $(\Delta_p m)_t$ est une nouvelle tendance, qui respecte encore la condition de moyenne nulle sur une période :

$$\sum_{k=1}^p (\Delta_p m)_{t+k} = \sum_{k=1}^p m_{t+k} - \sum_{k=1}^p m_{t+k-p} = 0 .$$

Le terme $(\Delta_p W)_t$ un terme de bruit : stationnaire, aléatoire, de moyenne nulle et de variance :

$$\text{Var}(\Delta_p W_t) = 2\gamma_W(0) + 2\gamma_W(p) .$$

Quant au terme $(\Delta_p s)_t$, lui, il est nul :

$$(\Delta_p s)_t = s_t - s_{t-p} = 0 .$$

Ainsi, la décomposition $(\Delta_p X)_t = (\Delta_p m)_t + (\Delta_p W)_t$ a bien éliminé la composante saisonnière.

2. On a déjà décrit partiellement l'effet sur la tendance : on conserve la propriété de moyenne nulle sur une période. Pour ce qui est de la forme de la tendance obtenue, on peut dire par exemple que si la tendance est polynomiale de degré k , la tendance modifiée est également polynomiale, mais de degré $k - 1$. En effet :

$$(\Delta_p t^k)_t = t^k - (t - p)^k = - \sum_{j=0}^{k-1} \binom{k}{j} t^j (-p)^{k-j} .$$

3. Si on sait comment éliminer la tendance originale, m par un opérateur linéaire dont l'effet ne dépend pas de l'origine du temps, alors le même opérateur élimine $\Delta_p m$. Par ailleurs, si m est polynomiale de degré k , d'après la réponse précédente, on voit que ∇^k élimine $\Delta_p m$.
4. Un avantage important de cette méthode est qu'elle ne fait aucune hypothèse paramétrique sur la saisonnalité, autre que la connaissance de la période. Par contre, un inconvénient est l'éventuelle introduction de corrélation dans le bruit. Par exemple, si le bruit original est blanc, le nouveau bruit devient corrélé et en plus la variance du nouveau bruit est plus grande que celle de l'ancien bruit.

□

Solution 3.10: Le signal y vérifie :

$$(I - \alpha e^{i\omega_0} B)y_t = \alpha(I - e^{i\omega_0} B)x_t .$$

On voit ainsi que l'opération effectuée par le filtre sur le signal x (ou plutôt sur αx) ressemble à celle effectuée dans un $ARMA(1,1)$ sur le bruit blanc. La fonction de transfert du filtre est donc :

$$F(\lambda) = \alpha \frac{1 - e^{i\omega_0} e^{-i\lambda}}{1 - \alpha e^{i\omega_0} e^{-i\lambda}} = \alpha \frac{1 - e^{i(\omega_0 - \lambda)}}{1 - \alpha e^{i(\omega_0 - \lambda)}} .$$

Notamment, $F(\omega_0) = 0$, et donc la pulsation ω est absente du signal y . Plus α est proche de 1, plus F est proche de la fonction valant 0 en ω_0 et 1 ailleurs. Cet effet est linéaire et s'applique également au bruit. D'ailleurs, on a aussi :

$$(I - \alpha e^{i\omega_0} B)y_t = \alpha(I - e^{i\omega_0} B)w_t .$$

Plus précisément, si le bruit w a une transformée de Fourier \hat{w} , on a :

$$\hat{y}(\lambda) = F(\lambda)\hat{w}(\lambda) .$$

Notamment, si w est considéré comme la réalisation d'un processus stationnaire avec une densité spectrale de puissance f_W , on a :

$$f_y(\lambda) = |F(\lambda)|^2 f_W(\lambda) .$$

□

Solution 3.11:

1. On trouve :

$$\mathbb{E}(X_t) = m_t + s_t ,$$

et

$$\text{Cov}(X_t, X_{t+k}) = \gamma(k) .$$

2. On trouve :

$$\mathbb{E}(X_t) = m_t s_t ,$$

et

$$\text{Cov}(X_t, X_{t+k}) = \gamma(k) .$$

3. On trouve :

$$\mathbb{E}(X_t) = m_t s_t ,$$

et

$$\text{Cov}(X_t, X_{t+k}) = m_t m_{t+k} s_t s_{t+k} \gamma(k) .$$

□

Solution 3.12: Une moyenne mobile étant un opérateur linéaire, il suffit d'écrire ce qu'il faut sur une base des polynômes $(1, t, t^2)$ et sur une base des fonctions de période 3 et de moyenne nulle :

$$\left\{ \begin{array}{l} \forall t \in \mathbb{Z}, a_2 + a_1 + a_0 + a_1 + a_2 = 1 \\ \forall t \in \mathbb{Z}, a_2(t-2) + a_1(t-1) + a_0 t + a_1(t+1) + a_2(t+2) = t \\ \forall t \in \mathbb{Z}, a_2(t-2)^2 + a_1(t-1)^2 + a_0 t^2 + a_1(t+1)^2 + a_2(t+2)^2 = t^2 \\ \forall t \in \mathbb{Z}, a_2 s_1(t-2) + a_1 s_1(t-1) + a_0 s_1(t) + a_1 s_1(t+1) + a_2 s_1(t+2) = 0 \\ \forall t \in \mathbb{Z}, a_2 s_2(t-2) + a_1 s_2(t-1) + a_0 s_2(t) + a_1 s_2(t+1) + a_2 s_2(t+2) = 0 \end{array} \right.$$

ce qui est équivalent à :

$$\left\{ \begin{array}{l} 2a_2 + 2a_1 + a_0 = 1 \\ 4a_2 + a_1 = 0 \\ \forall t \in \mathbb{Z}, a_2 + a_1 - a_0 = 0 \end{array} \right.$$

D'où :

$$a_0 = \frac{3}{9}, \quad a_1 = \frac{4}{9}, \quad a_2 = -\frac{1}{9} .$$

□

Solution 3.13: On cherche donc les coefficients $\theta = (\theta_{-m}, \dots, \theta_0, \dots, \theta_m)$, solution du problème de minimisation suivant :

$$\left\{ \begin{array}{l} \max_{\theta} \sum_{i=-m}^m \theta_i^2 \\ \text{s.c.} \quad \sum_{i=-m}^m \theta_i = 1 \end{array} \right.$$

En appliquant la méthode des multiplicateurs de Lagrange, on montre que la solution optimale est :

$$\forall i \in \{-m, \dots, m\}, \quad \theta_i = \frac{1}{2m+1} .$$

□

Solution 4.1:

1. La trace de F vaut 2 et le déterminant vaut 1, il y a une seule valeur propre : 1. L'équation n'est donc pas stable, ce qui colle intuitivement avec le fait qu'une tendance linéaire "dérive".

2. Une unique valeur propre : ϕ . L'équation est donc stable si et seulement si $|\phi| < 1$, ce qui correspond au cas où l' $AR(1)$ est causal.
3. On a deux valeurs propres qui se déduisent des deux exemples précédents : 1 et ϕ . L'équation n'est donc pas stable, ce qui colle toujours avec le fait qu'une tendance linéaire "dérive".

□

Solution 4.2: Soit λ une valeur propre non nulle de F et \vec{u} un vecteur propre associé à λ . On pose :

$$\tilde{X}_t = X_t + \lambda^t \vec{u} ,$$

et :

$$\tilde{Y}_t = G\tilde{X}_t + W_t .$$

On a alors :

$$\tilde{X}_{t+1} = X_{t+1} + \lambda^{t+1} \vec{u} = FX_t + V_t + F(\lambda^t \vec{u}) = F\tilde{X}_t + V_t .$$

Donc, \tilde{X}_t est une autre solution de l'équation d'état (4.7).

□

Solution 4.3: On peut modifier le modèle en ajoutant un bruit sur la pente de la tendance linéaire. Soit Z_t et W_t deux bruits blancs. On peut prendre :

$$X_t = \begin{pmatrix} a \\ x_t \end{pmatrix} ,$$

initialisation :

$$X_0 = \begin{pmatrix} a \\ b \end{pmatrix} ,$$

équation d'état :

$$X_{t+1} = FX_t + \begin{pmatrix} 0 \\ W_t \end{pmatrix} ,$$

avec

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} ; ,$$

observation :

$$Y_t = GX_t + Z_t ,$$

avec

$$G = \begin{pmatrix} 0 & 1 \end{pmatrix} .$$

□

Solution 4.4: L'équation d'état est la même, et l'équation d'observation devient :

$$Y_t = G(X_t) + Z_t ,$$

avec :

$$G(x, y) = y^2 .$$

La matrice jacobienne de G au point (x, y) est :

$$J_y = \begin{pmatrix} 0 & 2y \end{pmatrix} ,$$

ce qui donne les équations suivantes pour le filtre de Kalman étendu :

1. Prédiction de l'état :

$$\begin{aligned} P_{t-1}(X_t) &= F P_{t-1}(X_{t-1}) , \\ C_{t|t-1} &= F C_{t-1|t-1} F^T \end{aligned}$$

2. Prédiction de la mesure :

$$P_{t-1}(Y_t) = G(P_{t-1}(X_t)) = [P_{t-1}(X_t)]_2^2 ,$$

3. Calcul de l'innovation et de sa covariance :

$$\begin{aligned} \nu_t &= Y_t - P_{t-1}(Y_t) , \\ S_t &= J_{P_{t-1}(X_t)} C_{t|t-1} J_{P_{t-1}(X_t)}^T + R = 4P_{t-1}(Y_t) + \sigma^2 , \end{aligned}$$

car $R = \sigma^2$, la variance du bruit de mesure.

4. Calcul du gain :

$$K_t = C_{t|t-1} J_{P_{t-1}(X_t)}^T S_t^{-1} ,$$

5. Estimation de l'état courant :

$$\begin{aligned} P_t(X_t) &= P_{t-1}(X_t) + K_t \nu_t , \\ C_{t|t} &= C_{t|t-1} - K_t S_t K_t^T . \end{aligned}$$

□

Solution 5.1: Soit X une solution stationnaire. On observe que pour tout $k \geq 0$,

$$X_t = \sum_{i=0}^k \phi^i Z_{t-i} + \phi^{k+1} X_{t-k-1} .$$

La série $\sum_{i \geq 0} \phi^i Z_{t-i}$ converge (cf. Exercice 5.16), et $\phi^{k+1} X_{t-k-1}$ converge vers 0 presque sûrement et en moyenne quadratique. On a donc :

$$X_t = \sum_{i=0}^{\infty} \phi^i Z_{t-i} ,$$

et on vérifie facilement que cette somme est bien une solution. C'est donc la seule solution stationnaire, et elle est bien causale. □

Solution 5.2:

1. On peut écrire :

$$X_t = -\phi^{-1} U_{t+1} + \phi^{-1} X_{t+1} .$$

Donc, pour tout $k \geq 1$:

$$X_t = -\sum_{i=1}^k \phi^{-i} U_{t+i} + \phi^{-k} X_{t+k} .$$

La série $-\sum_{i \geq 1} \phi^{-i} U_{t+i}$ converge (cf. Exercice 5.16), et $\phi^{-k} X_{t+k}$ converge vers 0 presque sûrement et en moyenne quadratique. On a donc :

$$X_t = -\sum_{i=1}^{\infty} \phi^{-i} U_{t+i} ,$$

et on vérifie facilement que cette somme est bien une solution. C'est donc la seule solution stationnaire. Cette représentation n'est pas causale.

2. On calcule :

$$\gamma_X(0) = \sigma^2 \sum_{i=1}^{\infty} \phi^{-2i} = \sigma^2 \frac{\phi^{-2}}{1 - \phi^{-2}} ,$$

puis, pour tout $k \geq 0$,

$$\mathbb{E}(X_t X_{t+k}) = \phi \mathbb{E}(X_{t-1} X_{t+k}) + \mathbb{E}(X_{t+k} U_t) = \phi \mathbb{E}(X_{t-1} X_{t+k}) ,$$

et donc :

$$\forall k \geq 0, \gamma_X(k) = \phi \gamma_X(k+1) .$$

Ce qui donne :

$$\forall k \geq 0, \gamma_X(k) = \sigma^2 \frac{\phi^{-2}}{1 - \phi^{-2}} \phi^{-k} = \sigma^2 \frac{\phi^{-k}}{\phi^2 - 1} .$$

3. Si $V_t = X_t - \frac{1}{\phi} X_{t-1}$, soit γ_V la fonction d'autocovariance de V . On a :

$$\gamma_V(0) = \gamma_X(0) \left(1 + \frac{1}{\phi^2}\right) - \frac{2}{\phi} \gamma_X(1) = \frac{\sigma^2}{\phi^2} ,$$

puis, pour $k \geq 1$,

$$\begin{aligned} \gamma_V(k) &= \gamma_X(k) \left(1 + \frac{1}{\phi^2}\right) - \frac{1}{\phi} \gamma_X(k+1) - \frac{1}{\phi} \gamma_X(k-1) , \\ &= \gamma_X(k) \left(1 + \frac{1}{\phi^2} - \frac{1}{\phi} \phi^{-1} - \frac{1}{\phi} \phi\right) , \\ &= 0 . \end{aligned}$$

V est donc un bruit blanc, et on a une représentation en $AR(1)$ causale :

$$X_t = \phi^{-1} X_{t-1} + V_t ,$$

avec V bruit blanc de variance σ^2/ϕ^2 .

□

Solution 5.4: La première question découle de l'équation (8.1). Pour l'application au $MA(2)$, on obtient :

$$\begin{cases} \gamma(0) &= \sigma^2(2 + 1.5^2) , \\ \gamma(1) &= 0 , \\ \gamma(2) &= -\sigma^2 . \end{cases}$$

Pour un $AR(1)$ causal, on peut partir de l'expression en $MA(\infty)$ $X_t = \sum_{i \geq 0} \phi^i Z_{t-i}$ et obtenir :

$$\gamma(k) = \sigma^2 \frac{\phi^k}{1 - \phi^2} ,$$

et pour un $AR(1)$ avec $|\phi| > 1$, voir l'Exercice 5.2.

□

Solution 5.6: Soit X un $AR(1)$, d'après le cours, la densité spectrale est :

$$P_X(\omega) = \frac{\sigma^2}{2\pi |1 - \phi e^{-j\omega}|^2} = \frac{\sigma^2}{2\pi (1 + \phi^2 - 2\phi \cos \omega)} .$$

Pour un $MA(1)$,

$$P_X(\omega) = \frac{\sigma^2 |1 + \theta e^{-j\omega}|^2}{2\pi} = \frac{\sigma^2 (1 + \theta^2 + 2\theta \cos \omega)}{2\pi}.$$

□

Solution 5.8: Avec les notations usuelles,

$$\begin{aligned} \frac{\theta(z)}{\phi(z)} &= \frac{1 + \theta z}{1 - \phi z}, \\ &= (1 + \theta z) \sum_{i=0}^{\infty} \phi^i z^i, \\ &= 1 + \sum_{i=1}^{\infty} \phi^{i-1} (\phi + \theta) z^i. \end{aligned}$$

Donc la représentation en $MA(\infty)$ d'un $ARMA(1, 1)$ causal est :

$$X_t = Z_t + \sum_{i=1}^{\infty} \phi^{i-1} (\phi + \theta) Z_{t-i}.$$

On en déduit la fonction d'autocovariance de X (cf. équation (8.1)) :

$$\begin{cases} \gamma_X(0) &= \sigma^2 \left(1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right) \\ \gamma_X(k) &= \sigma^2 \phi^{k-1} \frac{(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2} \quad \forall k \geq 1 \end{cases}$$

D'où la fonction d'autocorrélation :

$$\rho_X(k) = \phi^{k-1} \frac{(1 + \phi\theta)(\phi + \theta)}{1 + \theta^2 + 2\phi\theta} \quad \forall k \geq 1.$$

□

Solution 5.9: Dans le cas d'un $ARMA(1, 1)$, les équations aux différences s'écrivent, en posant $\frac{\theta(z)}{\phi(z)} = \sum_{i \geq 0} \psi_i z^i$:

$$\begin{cases} \gamma_X(0) - \phi \gamma_X(1) &= \sigma^2 (1 + \theta \psi_1), \\ \gamma_X(1) - \phi \gamma_X(0) &= \sigma^2 \theta, \\ \gamma_X(k) - \phi \gamma_X(k-1) &= 0, \quad \forall k \geq 2 \end{cases}$$

Comme $\psi_1 = \phi + \theta$, on vérifie facilement ces équations.

□

Solution 5.10: Soit $(Z_t)_{t \in \mathbb{Z}}$ un bruit blanc de variance σ^2 .

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_n(0)) &= \mathbb{E}\left(\frac{1}{n} \sum_{t=1}^n Z_t^2 - \left(\frac{1}{n} \sum_{t=1}^n Z_t\right)^2\right), \\ &= \mathbb{E}(Z_1^2) - \text{Var}\left(\frac{1}{n} \sum_{t=1}^n Z_t\right)^2, \\ &= \sigma^2 - \frac{1}{n} \text{Var}(Z_1), \\ &= \left(1 - \frac{1}{n}\right) \sigma^2. \end{aligned}$$

On trouve également :

$$\begin{aligned}
\mathbb{E}(\hat{\gamma}_n(1)) &= \mathbb{E}\left(\frac{1}{n} \sum_{t=1}^{n-1} (Z_t - \bar{Z}_n)(Z_{t+1} - \bar{Z}_n)\right), \\
&= \frac{n-1}{n} \mathbb{E}(Z_1 Z_2) - \frac{1}{n} \mathbb{E}(\bar{Z}_n \sum_{t=1}^{n-1} Z_{t+1}) - \frac{1}{n} \mathbb{E}(\bar{Z}_n \sum_{t=1}^{n-1} Z_t) + \frac{n-1}{n} \mathbb{E}(\bar{Z}_n^2), \\
&= -\mathbb{E}((\bar{Z}_n)^2) + \frac{1}{n} \mathbb{E}(Z_1 \bar{Z}_n) - \mathbb{E}((\bar{Z}_n)^2) + \frac{1}{n} \mathbb{E}(Z_n \bar{Z}_n) + \frac{n-1}{n} \mathbb{E}(\bar{Z}_n^2), \\
&= -\frac{n+1}{n^2} \sigma^2 + \frac{2}{n^2} \sigma^2, \\
&= \frac{1-n}{n^2} \sigma^2.
\end{aligned}$$

Ensuite, on voit qu'il faut que Z ait un moment d'ordre 4 pour pouvoir calculer la variance de $\hat{\gamma}_n(0)$:

$$\begin{aligned}
\mathbb{E}(\hat{\gamma}_n(0)^2) &= \frac{1}{n^2} \sum_{1 \leq t, s \leq n} \mathbb{E}((Z_t - \bar{Z}_n)^2 (Z_s - \bar{Z}_n)^2), \\
&= \frac{1}{n^2} \sum_{1 \leq t, s \leq n} \mathbb{E}(Z_t^2 Z_s^2 + \bar{Z}_n^4 + 4Z_t Z_s \bar{Z}_n^2 + \bar{Z}_n^2 Z_t^2 + \bar{Z}_n^2 Z_s^2 \\
&\quad - 2Z_t^2 Z_s \bar{Z}_n - 2Z_s^2 Z_t \bar{Z}_n - 2Z_s \bar{Z}_n^3 - 2Z_t \bar{Z}_n^3), \\
&= \frac{1}{n^2} \left[\sum_{1 \leq t \leq n} \mathbb{E}(Z_t^4) + n(n-1)\sigma^4 + n^2 \mathbb{E}(\bar{Z}_n^4) + 4 \sum_{1 \leq t \leq n} \mathbb{E}(Z_t^2 \bar{Z}_n^2) \right. \\
&\quad \left. + 2n \sum_{1 \leq t \leq n} \mathbb{E}(\bar{Z}_n^2 Z_t^2) - 4 \sum_{1 \leq t \leq n} \mathbb{E}(Z_t^3 \bar{Z}_n) - 4n \sum_{1 \leq t \leq n} \mathbb{E}(Z_s \bar{Z}_n^3) \right], \\
&= \frac{1}{n^2} \left[n(n-1)\sigma^4 + \mathbb{E}(Z_1^4)(n + \frac{4}{n} + 2 - 4 + \frac{4}{n^2}) + n^2 \mathbb{E}(\bar{Z}_n^4) \right], \\
&= \frac{1}{n^2} \left[n(n-1)\sigma^4 + \mathbb{E}(Z_1^4)(n - 2 + \frac{4}{n} + \frac{4}{n^2}) + n^{-2}(n\mathbb{E}(Z_1^4) + 3n(n-1)\sigma^4) \right],
\end{aligned}$$

donc :

$$\text{Var}(\hat{\gamma}_n(0)) = \mathbb{E}(Z_1^4) \frac{n-2+\frac{4}{n}+\frac{4}{n^2}}{n^2} + \sigma^4 \frac{(n+3)(n-1)}{n^3}.$$

Notamment, cette variance tend vers 0. Comme la moyenne de $\hat{\gamma}_n(0)$ converge vers σ^2 , on a donc convergence dans L^2 , et donc en probabilité, de $\hat{\gamma}_n(0)$ vers σ^2 . \square

Solution 5.11: Le fait que pour $h > q$, $\rho_X(h)$ soit nul se voit facilement, par exemple sur les équations aux différences. Soit Z le bruit blanc fort tel que $X_t = \theta(B)Z_t$. On remarque d'abord que :

$$\hat{\gamma}_X(0) = \frac{1}{n} \sum_{t=1}^n X_t^2 - \left(\frac{1}{n} \sum_{t=1}^n X_t \right)^2.$$

Le Théorème 5.12 implique que :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}(X_1) = 0,$$

De même, le Théorème 5.12 appliqué à X_t^2 implique que :

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}(X_1^2) ,$$

Ainsi,

$$\hat{\gamma}_X(0) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \text{Var}(X_1) = \gamma_X(0) .$$

Soit $h > q$ fixé. On a :

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h} X_t - \bar{X}_n \left(\frac{1}{n} \sum_{t=1}^{n-h} X_t + \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h} \right) .$$

Le Théorème 5.12 appliqué à X_t , qui est q -dépendant, implique que :

$$\frac{1}{\sqrt{\text{Var}(\sum_{t=1}^{n-h} X_t)}} \sum_{t=1}^{n-h} X_t \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) .$$

Or, dès que $n \geq h + q + 1$,

$$\text{Var}(\sum_{t=1}^{n-h} X_t) = (n-h)\text{Var}(X_1) + (n-h-q) \sum_{i=1}^q \text{Cov}(X_t, X_{t+i}) + O(1) .$$

comme \bar{X}_n converge vers 0 en probabilité, on en déduit que $\bar{X}_n \frac{1}{\sqrt{n}} \sum_{t=1}^{n-h} X_t$ converge vers 0 en probabilité. De même,

$$\frac{1}{\sqrt{\text{Var}(\sum_{t=1}^{n-h} X_{t+h})}} \sum_{t=1}^{n-h} X_{t+h} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) ,$$

et donc $\bar{X}_n \frac{1}{\sqrt{n}} \sum_{t=1}^{n-h} X_{t+h}$ converge vers 0 en probabilité. Enfin, le Théorème 5.12 appliqué à $X_t X_{t+h}$, qui est $q + h$ -dépendant donne que :

$$\frac{1}{\sqrt{\text{Var}(\sum_{t=1}^{n-h} X_t X_{t+h})}} \sum_{t=1}^{n-h} X_t X_{t+h} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) .$$

Calculons $\text{Var}(\sum_{t=1}^{n-h} X_t X_{t+h})$, en utilisant que X est q -dépendant et que $h > q$:

$$\begin{aligned} \text{Var}(\sum_{t=1}^{n-h} X_t X_{t+h}) &= (n-h)\mathbb{E}((X_1 X_{h+1})^2) + 2 \sum_{1 \leq i < j \leq (n-h)} \mathbb{E}(X_i X_{i+h} X_j X_{j+h}) , \\ &= (n-h)\mathbb{E}(X_1^2)\mathbb{E}(X_{h+1}^2) + 2 \sum_{i=1}^{n-h-q-1} \sum_{j=i+q+1}^{n-h} \mathbb{E}(X_i X_{i+h} X_j X_{j+h}) \\ &\quad + 2 \sum_{i=1}^{n-h-q-1} \sum_{j=i+q+1}^{i+q+1} \mathbb{E}(X_i X_{i+h} X_j X_{j+h}) + 2 \sum_{i=n-h-q}^{n-h} \sum_{j=i+1}^{n-h} \mathbb{E}(X_i X_{i+h} X_j X_{j+h}) , \\ &= (n-h)\gamma_X(0)^2 + 0 + 2 \sum_{i=1}^{n-h-q-1} \sum_{j=i+q+1}^{i+q+1} \mathbb{E}(X_i X_{i+h} X_j X_{j+h}) + O_P(1) . \end{aligned}$$

Remarquons maintenant qu'en notant $\theta_0 = 1$,

$$\begin{aligned}
\mathbb{E}(X_i X_{i+h} X_j X_{j+h}) &= \sum_{0 \leq k_1, k_2, k_3, k_4 \leq q} \theta_{k_1} \theta_{k_2} \theta_{k_3} \theta_{k_4} \mathbb{E}(Z_{i-k_1} Z_{i+h-k_2} Z_{j-k_3} Z_{j+h-k_4}) , \\
&= \sum_{0 \leq k_1, k_2, k_3, k_4 \leq q} \theta_{k_1} \theta_{k_2} \theta_{k_3} \theta_{k_4} \sigma^4 \mathbb{I}_{i-k_1=j-k_3} \mathbb{I}_{i-k_2=j-k_4} , \\
&= \sigma^4 \sum_{k_1=0}^q \sum_{k_2=0}^q \theta_{k_1} \theta_{k_2} \theta_{k_1+j-i} \theta_{k_2+j-i} , \\
&= (\sigma^2 \theta_{k_1} \theta_{k_1+j-i})^2 , \\
&= \gamma_X(j-i)^2 .
\end{aligned}$$

Donc, pour $h > q$:

$$\text{Var}\left(\sum_{t=1}^{n-h} X_t X_{t+h}\right) = (n-h)\gamma_X(0)^2 + 2(n-h-q-2) \sum_{k=1}^q \gamma_X(k)^2 + O_P(1) ,$$

et ainsi, pour $h > q$,

$$\sqrt{n} \hat{\gamma}_X(h) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \gamma_X(0)^2 + 2 \sum_{k=1}^q \gamma_X(k)^2\right) .$$

Comme $\hat{\gamma}_X(0)$ converge en probabilité vers $\gamma_X(0)$, on en déduit :

$$\sqrt{n} \hat{\rho}_X(h) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, 1 + 2 \sum_{k=1}^q \rho_X(k)^2\right) .$$

□

Solution 5.13: Il suffit d'écrire que $\psi(z)\phi(z) = \theta(z)$, développer le produit $\phi(z)\psi(z)$ et écrire que le coefficient devant z^j est nul pour $j > q$. On obtient :

$$\forall i = 1, \dots, p, \quad \sum_{j=q+i}^p \hat{\psi}_{q+i-j} \hat{\phi}_j = \hat{\psi}_{q+i} ,$$

ce qui est exactement l'identité matricielle demandée.

□

Solution 5.14: On rappelle que pour un $AR(1)$,

$$\gamma(k) = \sigma^2 \frac{\phi^k}{1 - \phi^2} ,$$

On en déduit par récurrence les formules de Durbin-Levinson :

$$\left\{ \begin{array}{ll} \phi_{1,1} &= \phi \\ v_1 &= \sigma^2 \\ \phi_{m,m} &= 0 \quad \forall m \geq 2 \\ \phi_{m,1} &= \phi \\ \phi_{m,j} &= 0 \quad \forall m \geq 2 \quad \forall 2 \leq j \leq m \\ v_m &= \sigma^2 \end{array} \right.$$

□

Solution 5.16: Proposition 3.1.1 p.83 + Théorème 3.1.1 p.85 de [5].

□

Solution 5.17: Comme dans l'exercice 2.3, on remarque que :

$$\gamma_X(h) = \sum_{i,l \in \mathbb{Z}} \psi_i \bar{\psi}_l \gamma_Z(h+l-i),$$

où γ_Z est la fonction d'autocovariance de Z . Donc :

$$\gamma_X(h) = \sigma^2 \sum_{l \in \mathbb{Z}} \psi_{h+l} \bar{\psi}_l, \quad (8.1)$$

d'où, pour $r^{-1} < |z| < r$:

$$\begin{aligned} G_X(z) &= \sum_{h \in \mathbb{Z}} \gamma_X(h) z^h, \\ &= \sigma^2 \sum_{h \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \psi_{h+l} \bar{\psi}_l z^h, \\ &= \sigma^2 \sum_{h \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \psi_{h+l} z^{h+l} \bar{\psi}_l z^{-l}, \\ &= \sigma^2 \psi(z) \bar{\psi}(z^{-1}). \end{aligned}$$

Pour un $ARMA(p, q)$ comme dans l'énoncé, on peut écrire $\phi(z)$ comme un produit de polynômes de degré 1 ayant des racines hors du cercle unité. Soit r le minimum des modules de ces racines. Dans ce cas, $\theta(z)/\phi(z)$ peut se développer en série entière à l'intérieur du disque $D(0, r)$. On note ψ_i les coefficients de ce développement :

$$\frac{\theta(z)}{\phi(z)} = \sum_{i=1}^{\infty} \psi_i z^i \quad \forall |z| < r.$$

Alors on sait que si ϕ n'a pas de racine à l'intérieur du disque unité, on a la convergence absolue presque sûre et en moyenne quadratique (cf. Exercice 5.16) :

$$X_t = \sum_{i=1}^{\infty} \psi_i Z_{t-i},$$

et on a donc, pour $r^{-1} < |z| < r$:

$$G_X(z) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}.$$

En fait, il suffit que ϕ et θ n'aient pas de racine de module 1 (cf. paragraphe 3.5 page 102 dans [5]).

□

Solution 5.18: L'autocovariance de Y est égale à celle du processus recentré $X_t = Y_t - \mu$. En utilisant l'équation (8.1),

$$\begin{cases} \gamma_Y(0) &= \sigma^2(1 + \theta_1^2 + \theta_{12}^2) \\ \gamma_Y(1) &= \sigma^2 \theta_1 \\ \gamma_Y(k) &= 0 & \forall k = 2, \dots, 11 \\ \gamma_Y(11) &= \sigma^2 \theta_1 \theta_{12} \\ \gamma_Y(12) &= \sigma^2 \theta_{12} \\ \gamma_Y(k) &= 0 & \forall k \geq 13 \end{cases}$$

D'où la fonction d'autocorrélation :

$$\left\{ \begin{array}{ll} \rho_Y(0) &= 1 \\ \rho_Y(1) &= \frac{\theta_1}{1+\theta_1^2+\theta_{12}^2} \\ \rho_Y(k) &= 0 \quad \forall k = 2, \dots, 11 \\ \rho_Y(11) &= \frac{\theta_1\theta_{12}}{1+\theta_1^2+\theta_{12}^2} \\ \rho_Y(12) &= \frac{\theta_{12}}{1+\theta_1^2+\theta_{12}^2} \\ \rho_Y(k) &= 0 \quad \forall k \geq 13 \end{array} \right.$$

□

Solution 5.19: X est un $AR(1)$, et on a donc :

$$P_X(\omega) = \frac{\sigma^2}{2\pi|1 - \alpha e^{-j\omega}|^2}.$$

Notons U_t le processus $X_t + Z_t$. La densité spectrale de U_t est la somme des densités spectrales de X et Z . En effet, ces deux processus sont décorrélés par hypothèse, donc la fonction de covariance de U est $\gamma_X + \gamma_Z$. On a donc :

$$P_U(\omega) = \frac{\sigma^2}{2\pi|1 - \alpha e^{-j\omega}|^2} + \frac{\sigma^2}{2\pi}.$$

Enfin, on obtient la densité spectrale de Y par la formule du cours :

$$P_Y(\omega) = P_U(\omega) \frac{1}{|1 - \alpha e^{-j\omega}|^2}.$$

Donc :

$$P_Y(\omega) = \frac{\sigma^2}{2\pi} \left(\frac{1}{|1 - \alpha e^{-j\omega}|^2} + \frac{1}{|1 - \alpha e^{-j\omega}|^4} \right).$$

□

Solution 5.20: On obtient la densité spectrale de X par la formule du cours :

$$P_X(\omega) = \frac{\sigma^2}{2\pi} \left(\frac{1}{|1 - 0.99e^{-3j\omega}|^2} \right) = \frac{\sigma^2}{2\pi} \left(\frac{1}{1 + 0.99^2 - 2 * 0.99 \cos(3\omega)} \right).$$

Un graphe de la dsp de X avec $\sigma^2 = 1$ est présenté à la Figure 8.8. On remarque des “pics” aux pulsations $-2\pi/3$, 0 et $2\pi/3$, qui suggèrent que les trajectoires de X présenteront des oscillations avec une périodicité approchée valant 3. Un exemple de réalisation est donné à la Figure 8.9

□

Solution 5.21: Soient γ_X et f_X (resp. γ_Y et f_Y) la fonction d'autocovariance et la densité spectrale de X (resp. de Y). On note deux représentations causales de X et Y comme suit :

$$\phi_X(B)X_t = \theta_X(B)Z_t^X \quad \text{et} \quad \phi_Y(B)Y_t = \theta_Y(B)Z_t^Y,$$

où Z^X (resp. Z^Y) est un bruit blanc de variance σ_X^2 (resp. σ_Y^2), et on suppose que θ_X et θ_Y n'ont pas de racine de module < 1 (une telle représentation peut toujours être trouvée, cela

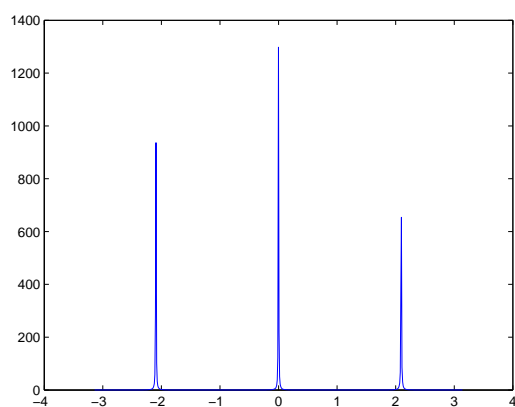
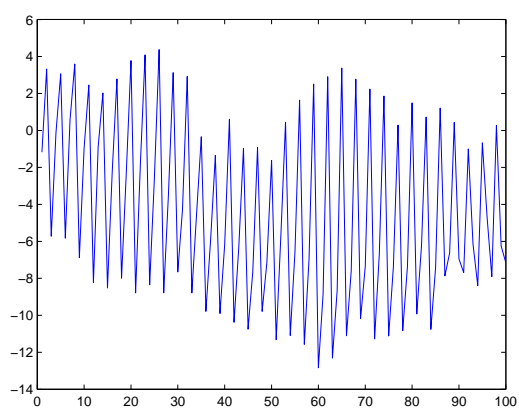


FIG. 8.8 – DSP de l'Exercice 5.20

FIG. 8.9 – Un $AR(3)$ presque 3-périodique

découle notamment du Théorème de Fejer-Riesz). Clairement, T est stationnaire et on a, grâce à la décorrélation entre X et Y , $\gamma_T = \gamma_X + \gamma_Y$. Par conséquent, $f_T = f_X + f_Y$. Donc :

$$\forall \theta \in [0, 2\pi], f_T(\theta) = \frac{\sigma_X^2 |\theta_X(e^{-j\theta})|^2}{2\pi |\phi_X(e^{-j\theta})|^2} + \frac{\sigma_Y^2 |\theta_Y(e^{-j\theta})|^2}{2\pi |\phi_Y(e^{-j\theta})|^2}.$$

La densité spectrale de T , f_T , est donc égale à $F(e^{-i\theta})$ où F est une fraction rationnelle. Or f_T est positive, réelle et intégrable sur $[0, 2\pi]$. Donc il existe, grâce au Théorème 5.14, une fraction irréductible $\frac{Q}{P}$ à coefficients réels, avec P n'ayant que des racines de module > 1 et Q des racines de module ≥ 1 telle que :

$$\forall \theta \in [0, 2\pi] f_T(\theta) = \left| \frac{Q}{P}(e^{-i\theta}) \right|^2.$$

D'autre part,

$$f_T(\theta) = \frac{1}{2\pi} \frac{\sigma_X^2 |\theta_X(e^{-i\theta})\phi_Y(e^{-i\theta})|^2 + \sigma_Y^2 |\theta_Y(e^{-i\theta})\phi_X(e^{-i\theta})|^2}{|\phi_X(e^{-i\theta})\phi_Y(e^{-i\theta})|^2}.$$

Comme deux fractions rationnelles en z coïncident dès qu'elles coïncident sur $|z| = 1$,

$$\frac{|Q(z)|^2}{|P(z)|^2} = \frac{1}{2\pi} \frac{\sigma_X^2 |\theta_X(z)\phi_Y(z)|^2 + \sigma_Y^2 |\theta_Y(z)\phi_X(z)|^2}{|\phi_X(z)\phi_Y(z)|^2}.$$

Si on note $Z_t^T = \frac{P(B)}{Q(B)}T_t$ (si Q a des racines de module 1, cela fait tout de même sens, cf. [2], paragraphe 7.3.6 page 66) on voit que la densité spectrale de Z^T est constante égale à 1. Donc la fonction d'autocovariance de Z^T vaut 2π en 0 et 0 ailleurs. Donc Z^T est un bruit blanc de variance 2π . Donc T est un ARMA. Pour connaître l'ordre de T , on remarque que le degré de P est inférieur ou égal à $p_1 + p_2$ et que le degré de Q est inférieur ou égal à $\max\{p_1 + q_2, q_1 + p_2\}$. \square

Solution 5.22: Soit $m > p$. En utilisant la formule de la transcomatrice, on observe que :

$$(\Gamma_m^{-1})_{mm} = \frac{\det(\Gamma_{m-1})}{\det(\Gamma_m)}. \quad (8.2)$$

Appelons L_1, \dots, L_m , les m lignes de la matrice Γ_m . Notons Γ'_m la matrice obtenue à partir de Γ_m en changeant la ligne L_m en L'_m , où :

$$L'_m = L_m - \sum_{i=1}^p L_{m-i} \phi_i.$$

Bien sûr, Γ_m et Γ'_m ont le même déterminant. D'autre part, on se rappelle les équations de Yule-Walker :

$$\Gamma_m \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{pmatrix}, \quad (8.3)$$

et aussi :

$$\sigma^2 = \gamma(0) - \sum_{i=1}^p \phi_i \gamma(i). \quad (8.4)$$

Les lignes $m - p$ à $m - 1$ du système donné par (8.3) et l'égalité (8.4) impliquent que :

$$L'_m = (0, \dots, 0, \sigma^2) .$$

Ainsi, si on développe le déterminant de Γ'_m selon sa dernière ligne, L'_m , on obtient que :

$$\det(\Gamma'_m) = \sigma^2 \det(\Gamma_{m-1}) ,$$

ce qui donne le résultat via (8.2). □

Solution 5.23: Clairement,

$$\mathbb{E}(Y|X) = X^2 .$$

Pour Y^* ,

$$a + b\mathbb{E}(X) = \mathbb{E}(X^2) ,$$

et :

$$a\mathbb{E}(X) + b\mathbb{E}(X^2) = \mathbb{E}(X^3) .$$

Le déterminant de ce système est égal à la variance de X , qui est non nulle par hypothèse, donc il y a une unique solution :

$$b = \frac{\mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2)}{\text{Var}(X)} ,$$

et

$$a = \frac{\mathbb{E}(X^2)^2 - \mathbb{E}(X)\mathbb{E}(X^3)}{\text{Var}(X)} .$$

Lorsque X est une normale centrée réduite, cela donne $a = 1$ et $b = 0$. On trouve donc $Y^* = 1$. L'erreur quadratique de prédiction linéaire vaut :

$$\mathbb{E}((Y - 1)^2) = \mathbb{E}(X^4) + 2\mathbb{E}(X^2)\mathbb{E}(Z - 1) + \mathbb{E}((Z - 1)^2) = \mathbb{E}(X^4) = 3 ,$$

et l'erreur de prédiction par l'espérance conditionnelle est :

$$\mathbb{E}([Y - \mathbb{E}(Y|X)]^2) = \mathbb{E}(Z^2) = 1 .$$

□

Solution 5.24: On peut inverser la représentation de X :

$$W_t = (I - \alpha B)^{-1} X_t = \sum_{j=0}^{\infty} \alpha^j X_{t-j} ,$$

D'où :

$$X_{t+1} = W_{t+1} - \sum_{j=1}^{\infty} \alpha^j X_{t+1-j} ,$$

Donc :

$$\mathbb{E}(X_{t+1} | (X_{t-k})_{k \geq 0}) = - \sum_{j=1}^{\infty} \alpha^j X_{t+1-j} .$$

L'erreur quadratique moyenne de prédiction est :

$$\mathbb{E}[(\hat{X}_{t+1} - X_{t+1})^2] = \mathbb{E}[W_{t+1}^2] = \sigma^2 .$$

□

Solution 5.25: Puisque X est un $AR(p)$,

$$X_{t+1} = \phi_1 X_t + \dots + \phi_p X_{t+1-p} + Z_{t+1} .$$

Or Z_{t+1} est indépendant de X_t, \dots, X_{t+1-p} et de moyenne nulle, donc :

$$\hat{X}_{t+1} = \mathbb{E}(X_{t+1} | X_t, \dots, X_{t+1-p}) = \phi_1 X_t + \dots + \phi_p X_{t+1-p} .$$

Remarque : cette espérance conditionnelle étant linéaire en X_t, \dots, X_{t+1-p} , c'est aussi le meilleur prédicteur linéaire de X_{t+1} sachant X_t, \dots, X_{t+1-p} .

On peut réécrire ce qu'on vient de faire de manière formelle, et pour X_{t+h} . Les polynômes $\phi(B)$ et B^h sont premiers entre eux, donc il existe (par Bezout) deux polynômes P et Q tels que :

$$1 = \phi(B)P(B) + B^h Q(B) ,$$

et on peut sans perte de généralité supposer que le degré de P est inférieur ou égal à $h-1$ (sinon, on peut faire passer "ce qui dépasse" dans $B^h Q(B)$). Du coup, le degré de Q est inférieur ou égal à $p-1$. Maintenant, on écrit :

$$\begin{aligned} X_{t+h} = \frac{1}{\phi(B)} Z_{t+h} &= \frac{\phi(B)P(B) + B^h Q(B)}{\phi(B)} Z_{t+h} , \\ &= P(B)Z_{t+h} + B^h Q(B) \frac{1}{\phi(B)} Z_{t+h} , \\ &= P(B)Z_{t+h} + B^h Q(B) X_{t+h} . \end{aligned}$$

Comme le degré de P est inférieur ou égal à $h-1$, $P(B)Z_{t+h}$ est indépendant de X_t, \dots, X_{t+1-p} . Comme le degré de Q est inférieur ou égal à $p-1$, $B^h Q(B)X_{t+h}$ est une fonction (linéaire) de X_t, \dots, X_{t+1-p} , et on a :

$$\hat{X}_{t+h} = \mathbb{E}(X_{t+h} | X_t, \dots, X_{t+1-p}) = Q(B)X_t .$$

L'erreur quadratique de prédiction est égale à :

$$\mathbb{E}[(X_{t+h} - \hat{X}_{t+h})^2] = \sigma^2 \sum_{i=0}^{h-1} p_i^2 ,$$

où on a noté $P(B) = \sum_{i=0}^{h-1} p_i B^i$. On peut calculer récursivement les p_i grâce à l'identité de Bezout ci-dessus, en développant le produit $\phi(B)P(B)$ et en se rappelant que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$:

$$\begin{cases} p_0 &= 1 \\ p_1 &= \phi_1 \\ p_j &= \sum_{i=0}^{j-1} p_i \phi_{j-i} \quad \forall 1 \leq j \leq h-1 \end{cases}$$

□

Solution 6.2: C'est évident : on a déjà vu que $(I-B)^d$ élimine les polynômes de degré inférieur ou égal à $d-1$. □

Solution 8.1:

loi conditionnelle pour un AR(1) : $\mathcal{N}(\phi X_{t-1}, \sigma_Z^2)$

loi non conditionnelle pour un AR(1) : $\mathcal{N}(0, \frac{\sigma_Z^2}{1-\phi^2})$ indépendante de t

loi conditionnelle pour un ARCH(1) : $\mathcal{N}(0, \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2)$

loi non conditionnelle pour un ARCH(1) : $\mathcal{L}(0, m_{2\epsilon}, m_{4\epsilon}, \dots)$.

loi conditionnelle pour un AR(1) / ARCH(1) : $\mathcal{N}(\phi X_{t-1}, \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2)$

loi non conditionnelle pour un AR(1) / ARCH(1) : à préciser (moyenne = 0). □

Solution 8.2:

Les moments d'ordre impair de η sont nuls.

D'après le cours, on a :

$$m_{2k_\epsilon} = \left(\sum_{i=0}^k C_k^i \alpha_0^{k-i} \alpha_1^i m_{2i_\epsilon} \right) m_{2k_\eta}$$

On en déduit :

$$m_{2k_\eta} = \frac{m_{2k_\epsilon}}{\sum_{i=0}^k C_k^i \alpha_0^{k-i} \alpha_1^i m_{2i_\epsilon}}$$

Pour que ϵ soit gaussien de variance 1, il faut :

$$m_{2k_\epsilon} = \prod_{j=1}^k (2j-1) \quad k = 1, 2, \dots$$

et $m_{0_\epsilon} = 1$.

Donc la loi de η est donnée par ses moments d'ordres pairs :

$$m_{2k_\eta} = \frac{\prod_{j=1}^k (2j-1)}{\sum_{i=0}^k C_k^i \alpha_0^{k-i} \alpha_1^i \prod_{j=1}^i (2j-1)}$$

avec par convention $\prod_{j=1}^i (2j-1) = 1$ pour $i = 0$. □

Bibliographie

- [1] D.W. Allan. Statistics of atomic frequency standards. *IEEE Proceedings*, 54 :221–235, 1966.
- [2] R. Azencott and D. Dacunha-Castelle. *Séries d’observations irrégulières. Modélisation et prévision*. Masson, 1984.
- [3] L. Bauwens, S. Laurent, and J. Rombouts. Multivariate garch models : A survey. *Journal of Applied Econometrics*, 21 :79–109, 2006.
- [4] R. Biscay, M. Lavielle, A. Gonzales, I. Clark, and P. Valdés. Maximum a posteriori estimation of change points in the eeg. *International Journal of Bio-Medical Computing*, 38 :189–196, 1995.
- [5] P. Brockwell and R. Davis. *Time Series : Theory and Methods*. Springer Series in Statistics. Springer, second edition, 1991.
- [6] D. Dacunha-Castelle and M. Duflo. *Probabilités et statistiques. Tome 1*. Collection Mathématiques Appliquées pour la Maîtrise. Paris : Masson, 1983.
- [7] J. de Gooijer, B. Abraham, A. Gould, and L. Robinson. Methods for determining the order of an autoregressive-moving average process : A survey. *International Statistical Review*, 53 :301–329, 1985.
- [8] P. Doukhan. *Séries temporelles, cours de troisième année*. ENSAE, 2004.
- [9] L. Elie, N. El Karoui, T. Jeantheau, and A. Pferzel. Les modèles arch sur les cours de change. Travail de coopération entre le CMAP (Centre de Mathématiques Appliquées de l’Ecole Polytechnique) et la Salle des Marchés de la BNP, disponible sur le web : [http://www.actuaries.org/AFIR/colloquia/Rome/Elie Karoui Jeantheau Pferzel.pdf](http://www.actuaries.org/AFIR/colloquia/Rome/Elie%20Karoui%20Jeantheau%20Pferzel.pdf).
- [10] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4) :987–1007, Jul. 1982.
- [11] W.A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity : Half a century of research. *Signal Processing*, 86 :639–697, 2006.
- [12] X. Gourdon. *Les Maths en tête*, volume Analyse. Paris : Ellipses, 1994.
- [13] X. Gourdon. *Les Maths en tête*, volume Algèbre. Paris : Ellipses, 1994.
- [14] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning. Data mining, inference, and prediction. 2nd ed.* Springer Series in Statistics. New York, NY : Springer. xxii, 745 p., 2009.
- [15] M. Misiti, Y. Misiti, G. Oppenheim, and J.M. Poggi. *Les ondelettes et leurs applications*. Hermes Science, IC2. Lavoisier, 2003.
- [16] B. Picinbono. Statistical properties of randomly modulated laser beams. *Physical Review A*, 4(6) :2398–2407, 1971.
- [17] R. Tsay. *Analysis of financial time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2005.

Index

- acf , voir autocorrélation (fonction d')
- AIC, 74
- AICC, 74
- AR, 9
- ARFIMA, 93
 - autocorrélation, 94
 - densité spectrale, 94
 - identification, 94
 - par maximum de vraisemblance, 95
 - variance d'Allan, 97
- ARIMA, 79
 - corrélation empirique, 79
 - identification, 81
 - modèle d'état, 82
 - prédiction, 80
- ARMA, 8, 51
 - modele d'etat, 56
 - autocorrélation partielle, 54
 - autocovariance, 54, 58
 - causal, 52
 - densité spectrale, 58
 - estimation spectrale, 75
 - identification, 60
 - estimation de l'ordre, 95
 - par maximum de vraisemblance, 71
 - paramétrique, 71
 - robuste, 65
 - inversible, 53
 - pacf , voir autocorrélation partielle
 - prédiction, 76
- ARMA/GARCH, 101
- autocorrélation (fonction d'), 6
 - empirique, 27
- autocovariance (fonction d'), 6
- autocovariance (fonction d'), 15
 - empirique, 26
- bruit blanc
 - faible, 8
 - fort, 8
- corrélation (fonction de), 6
 - empirique, 27
- covariance (fonction de), 6, 15
 - empirique, 26
- décomposition de Wold, 18
- densité spectrale, 16
 - d'un ARMA, 17
- distribution spectrale, 17
- Durbin-Levinson (algorithme de), 67
- élimination de composantes spectrales, 38
- élimination de la saisonnalité
 - par différentiation, 37
- élimination de la tendance
 - par différentiation, 32
 - par moindres carrés, 29
- équation aux différences, 59
- filtrage, 11
- filtre de Kalman, 45
 - étendu, 49
- GARCH, 101
 - effet, 107
 - identification, 105
 - moments, 102
 - validation, 109
- IGARCH, 108
- innovations (algorithme des), 65
- irrégulière (observation), 49
- Kalman , voir filtre
- lissage, 11
- lissage exponentiel, 31
- M-GARCH, 110
- MA, 9
 - manquante (observation), 48
- mémoire intermédiaire (processus à), 90
- mémoire longue (processus à), 87, 90
- modèle
 - additif, 7, 25

- hybride, 26
- multiplicatif, 25
- modèle d'état, 9, 43
 - associé à un ARMA , voir ARMA
- moindres carrés, 29
- moyenne mobile, 31, 36

- pacf , voir ARMA, autocorrélation partielle
- périodogramme, 19
- prédiction, 11
- prédiction linéaire, 41

- résidus (analyse), 72
- Représentation spectrale, 15
- représentation spectrale, 10

- saisonnalité, 3
- saisonniers (facteurs), voir saisonnalité
- SARIMA, 82
 - identification, 83
- série
 - stationnaire, 3
- simulation, 12
- stationnaire, voir série stationnaire
- synthèse, 12

- tendance, 3, 25

- validation, vii, 75

- Wold , voir décomposition de Wold

- Yule Walker (equations), 68