

**Master 2 Professionnel  
Cours de Séries Temporelles**

**Année 2008–2009**

G. Oppenheim<sup>1</sup>, A. Philippe<sup>2</sup>, M.-C. Viano<sup>3</sup>

---

<sup>1</sup>Université Marne la vallée

<sup>2</sup>Université Nantes

<sup>3</sup>Université Lille 1



## Table des matières

|   |    |
|---|----|
| Table des figures   | 5  |
| Chapitre 1. Introduction  | 7  |
| Chapitre 2. Qu'est ce qu'une série temporelle? Exemples commentés. Tendances, saisonnalités | 9  |
| 1. Exemples   | 9  |
| 2. Tendances, composantes saisonnières  | 12 |
| Chapitre 3. Lissages exponentiels   | 14 |
| 1. Introduction   | 14 |
| 2. Lissage exponentiel simple   | 14 |
| 3. Lissage exponentiel amélioré, ou double  | 16 |
| 4. Méthode de Holt-Winters  | 20 |
| Chapitre 4. Premiers indices descriptifs  | 24 |
| 1. Indices de tendance centrale   | 24 |
| 2. Indices de dispersion  | 24 |
| 3. Indices de dépendance  | 24 |
| 4. Comment utiliser l'auto-corrélation ?  | 29 |
| Chapitre 5. Vers la modélisation : suites stationnaires de variables aléatoires             | 32 |
| 1. Définitions. Auto-covariance et auto-corrélation   | 32 |
| 2. Exemples   | 33 |
| 3. Auto-corrélation partielle   | 35 |
| 4. Estimation de ces indices caractéristiques   | 39 |
| Chapitre 6. Les processus ARMA  | 46 |
| 1. Les processus autorégressifs et leurs indices caractéristiques.                          | 46 |
| 2. Les processus moyennes mobiles et leurs indices caractéristiques                         | 50 |
| 3. Les ARMA mixtes  | 52 |
| Chapitre 7. Comment prévoir dans un processus ARMA ?  | 55 |
| 1. Principe général   | 55 |
| 2. Prévision à horizon 1  | 55 |
| 3. Prévision à horizon supérieur à 1  | 57 |
| 4. Comportement de l'erreur de prévision quand l'horizon croît                              | 58 |

## TABLE DES MATIÈRES

4

|   |    |
|---|----|
| 5. Intervalles de confiance   | 59 |
| 6. Mise à jour  | 59 |
| Chapitre 8. Identification  | 61 |
| 1. Estimation des paramètres d'un modèle ARMA                           | 61 |
| 2. Choix du modèle  | 62 |
| 3. Identification et prévision  | 64 |
| Chapitre 9. Récapitulons : les prévisions <i>ARIMA</i> et <i>SARIMA</i> | 67 |
| 1. Résumé des étapes déjà vues  | 67 |
| 2. Filtrage inverse : modélisation <i>ARIMA</i> et <i>SARIMA</i>        | 67 |
| Chapitre 10. Initiation à la prévision non paramétrique                 | 71 |
| 1. Idée de base   | 71 |
| 2. Quelques mots sur l'asymptotique                                     | 72 |
| 3. Autre écriture et généralisation à d'autres noyaux                   | 72 |
| 4. Généralisation à d'autres mémoires                                   | 73 |
| 5. Retour à la prévision  | 74 |
| Bibliographie   | 76 |

## Table des figures

|   |    |
|---|----|
| 2.1 Taches solaires   | 11 |
| 2.2 Ventes de champagne   | 11 |
| 2.3 Ventes de voitures  | 11 |
| 2.4 Consommation d'électricité  | 12 |
| 2.5 Trafic aérien   | 12 |
| 2.6 Population des États-Unis   | 12 |
| <br>  |    |
| 3.1 Lissage et prévision par la méthode de lissage exponentiel simple d'un bruit blanc [gauche] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t$ , $\epsilon_t \sim N(0, 1)$ [droite] $(\alpha = .1, .5, .9)$  | 17 |
| 3.2 Lissage et prévision par la méthode de lissage exponentiel double d'un bruit blanc [gauche] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t$ , $\epsilon_t \sim N(0, 1)$ .[droite] $(\alpha = .1, .5, .9)$ | 19 |
| 3.3 Lissage et prévision par la méthode de lissage exponentiel double sur la série $X(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6)$ , $\epsilon_t \sim N(0, 1)$ . $(\alpha = .1, .5, .9)$                                | 19 |
| 3.4 Lissage et prévision par la méthode de Holt-Winters sur la série $x(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6)$ , $\epsilon_t \sim N(0, 1)$ .  | 22 |
| 3.5 Prévision par la méthode de Holt-Winters sur les données de vente de champagne.   | 23 |
| <br>  |    |
| 4.1 Série 1   | 27 |
| 4.2 Série 1 : Nuage de points $\mathcal{N}_k$ pour $k = 1, \dots, 8$ et auto-corrélation  | 27 |
| 4.3 Série 2   | 28 |
| 4.4 Série 2 : Nuage de points $\mathcal{N}_k$ pour $k = 1, \dots, 8$ et auto-corrélation  | 28 |
| 4.5 Série simulée non bruitée (en haut à droite), bruitée (en bas à droite). Auto-corrélation de la série bruitée, de la série bruitée différenciée, puis de la série bruitée différenciée et désaisonnalisée       | 30 |
| 4.6 Champagne : série et auto-corrélation   | 31 |
| 4.7 Champagne : série différenciée $X_t - X_{t-1}$ et auto-corrélation  | 31 |
| 4.8 Champagne : série différenciée et désaisonnalisée : $X_t - X_{t-12}$ et auto-corrélation  | 31 |
| <br>  |    |
| 5.1 Comparaison des covariances théoriques et empiriques pour le modèle ARMA(1,1)<br>$X_n + 0.8X_{n-1} = \epsilon_n - 0.8\epsilon_{n-1}$  | 41 |

## TABLE DES FIGURES

6

|  |    |
|--|----|
| 5.2 bruit blanc gaussien : série, ACF et PACF  | 42 |
| 5.3 modele AR(1) a=0.8 : série, ACF et PACF  | 42 |
| 5.4 modele AR(1) a=-0.8 : série, ACF et PACF   | 43 |
| 5.5 modele MA(1) a=0.8 : série , ACF et PACF   | 43 |
| 5.6 modele MA(1) a=-0.8 : série, ACF et PACF   | 44 |
| 5.7 modele AR(2) $X_n + 0.81X_{n-2} = \epsilon_n$ : série, ACF et PACF                                   | 44 |
| 5.8 modele ARMA(1,1) $X_n + 0.8X_{n-1} = \epsilon_n - 0.8\epsilon_{n-1}$ : série, ACF et PACF            | 45 |
| 6.1 Auto-corrélation et auto-corrélation partielle de quelques <i>ARMA</i>                               | 52 |
| 8.1 Convergence des estimateurs de $a_1$ et $a_2$ dans un modèle <i>AR</i> <sub>2</sub>                  | 63 |
| 8.2 Identification et prévision dans un <i>AR</i> <sub>4</sub> et un <i>MA</i> <sub>3</sub> simulés      | 65 |
| 8.3 La série des lynx. Ajustement d'un <i>AR</i> <sub>8</sub> et prévision                               | 66 |
| 9.1 Série du trafic aérien et série des ventes de champagne : ajustement d'un <i>SARIMA</i> et prévision | 70 |

## CHAPITRE 1

### Introduction

Ce cours est une initiation aux méthodes probabilités de prévision.

- Pourquoi vouloir prévoir ?

Très souvent pour des raisons socio-économiques (prévoir le temps qu'il va faire, l'évolution des ventes d'un certain produit, prévoir la consommation d'électricité pour ajuster au mieux la production).

- Comment prévoir ?

Toujours en s'appuyant sur ce qui s'est passé avant l'instant à partir duquel démarre la prévision. Et parfois en utilisant des renseignements annexes. Dans la prévision des ventes pour l'année  $j+1$  on s'appuiera sur l'évolution des ventes durant les années  $j, j-1, \dots$ , mais on tiendra compte aussi, éventuellement, d'indices exogènes (la conjoncture économique, certains événements récents inattendus etc....).

- Peut-on prévoir parfaitement bien ?

Jamais. Il y a toujours une erreur de prévision, et les méthodes sérieuses fournissent non pas une prévision, mais un intervalle de prévision. D'autre part il arrive souvent qu'une amélioration minime de la qualité de la prévision ait une grosse incidence sur les coûts.

- De quelles méthodes dispose-t-on pour prévoir ?

Elles sont nombreuses.

1- Vous connaissez la plus rudimentaire, qui consiste à faire une régression et à se baser sur elle pour prévoir. Par exemple on ajuste sur la série  $x_1, \dots, x_n$  un modèle de type

$$x_j = a_1 j^2 + a_2 j + a_3 + e_j \quad j = 1, \dots, n.$$

On estime les coefficients par une méthode de régression. On valide le résultat, puis on prévoit  $x_{n+1}$  par

$$\hat{a}_1(n+1)^2 + \hat{a}_2(n+1) + \hat{a}_3,$$

et  $x_{n+2}$  ainsi que les suivants de façon analogue.

2- Les lissages exponentiels sont une batterie de procédures extrêmement simples à mettre en œuvre. Ils sont exposés au chapitre 3.

3- Les méthodes de prévision les plus populaires depuis une quarantaine d'années sont celles liées à la modélisation ARMA. Elles consistent en gros à dégager les tendances évidentes dans le phénomène qu'on observe (croissance, périodicités etc...) et à se concentrer

sur ce qui reste lorsqu'on les a supprimées. On procède à une modélisation linéaire fine du résidu obtenu et on s'appuie sur cette modélisation pour obtenir la prévision. Ces méthodes, plus sophistiquées mais plus performantes que la régression et que les lissages exponentiels, sont coûteuses en temps calcul. Leur utilisation systématique dans tous les domaines, que ce soit économique, industriel, physique etc..., est due au développement exponentiellement rapide de l'outil informatique. Elles seront l'objet principal de ce cours.

4- Les méthodes à noyau, qui sont abordées au dernier chapitre, sont plus récentes et figurent parmi ce qu'on pourrait appeler les techniques non paramétriques de prévision. Plus souples que la méthode ARMA, elles demandent aussi un peu plus de savoir faire de la part de l'utilisateur.

Il y a bien d'autres méthodes dont on ne parlera pas ici. Il faut savoir aussi que les méthodes probabilités ne sont pas les seules. Par exemple la prévision de la pollution atmosphérique peut très bien se faire par des méthodes liées à la mécanique des fluides, qui n'ont aucunement recours aux probabilités.

Certaines données sont "faciles" et, sur elles, toutes les méthodes donnent de bonnes prévisions. Tout l'art du statisticien, face à des données difficiles, est de confronter plusieurs méthodes et de choisir celle qui convient le mieux.

## CHAPITRE 2

### Qu'est ce qu'une série temporelle ? Exemples commentés. Tendances, saisonnalités

Une série temporelle (ou encore une série chronologique) est une suite finie  $(x_1, \dots, x_n)$  de données indexées par le temps. L'indice temps peut être selon les cas la minute, l'heure, le jour, l'année etc.... Le nombre  $n$  est appelé la longueur de la série. Il est la plupart du temps bien utile de représenter la série temporelle sur un graphe construit de la manière suivante : en abscisse le temps, en ordonnée la valeur de l'observation à chaque instant. Pour des questions de lisibilité, les points ainsi obtenus sont reliés par des segments de droite. Le graphe apparaît donc comme une ligne brisée.

#### 1. Exemples

EXEMPLE 2.1. Les taches solaires. Il s'agit du nombre annuel de taches observées à la surface du soleil pour les années 1700–1980. Il y a une donnée par an, soit donc 281 points reliés par des segments. C'est une des séries temporelles les plus célèbres. On y distingue la superposition de deux phénomènes périodiques (une période d'à peu près 11 ans et une autre d'à peu près 60 ans, cette dernière étant moins crédible que la première vu la longueur de la série). ||

EXEMPLE 2.2. Les ventes de champagne. Il s'agit du nombre mensuel de bouteilles vendues entre mars 1970 et septembre 1978. Soit 103 points. On distingue sur le graphe une augmentation moyenne, une dispersion croissante (les “pics” sont de plus en plus marqués) et deux saisonnalités superposées (1 an et 6 mois).

Cette série, comme les trois qui suivent, est assez typique de ce que l'on rencontre dans le domaine de l'économétrie.

Elle est facile à modéliser et donne lieu à de “bonnes” prévisions pour toutes les méthodes couramment utilisées. ||

EXEMPLE 2.3. Les ventes de voitures. Les données courent sur 35 ans, et il y en a une par mois. Soit 420 points. On y voit une croissance non linéaire (peut être de type logarithmique). Les deux saisonnalités d'un an et de 6 mois se voient moins que dans la série précédente à cause du plus grand resserrement des abscisses. On voit aussi une rupture très nette aux environs de l'année 1983. Cette rupture fait que la série est extrêmement difficile à modéliser et donne de très mauvaises prévisions quelle que soit la méthode utilisée. ||

EXEMPLE 2.4. La consommation d'électricité sur toute la France. Une donnée toutes les demi-heures, pendant 35 jours en juin-juillet 1991. Soit 1680 données. On y voit bien

deux composantes périodiques correspondant à la journée (48 unités de temps) et à la semaine, ainsi qu'un effet "week-end" massif.

EXEMPLE 2.5. Le nombre de passagers (en milliers) dans les transports aériens. Une donnée par mois de 1949 à 1960. Tendance linéaire (ou plus) très marquée, ainsi qu'une forte augmentation de la variabilité et une composante périodique correspondant à l'année.

||

EXEMPLE 2.6. La population des États-Unis (une donnée tous les 10 ans, de 1790 à 1980, soit 21 points). Le seul phénomène évident est la croissance qui semble non-linéaire.

||

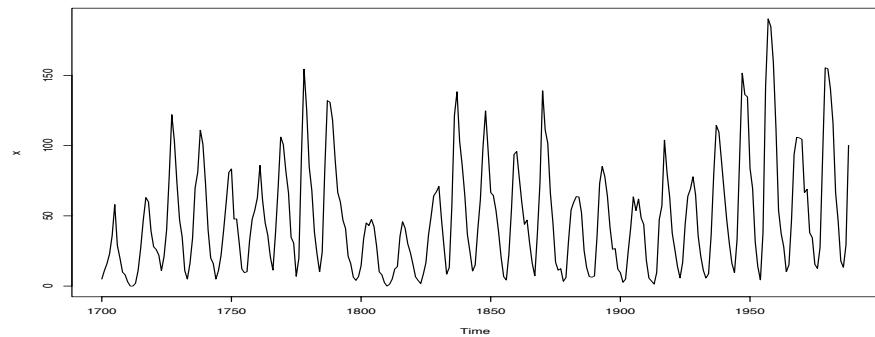


FIG. 2.1. Taches solaires

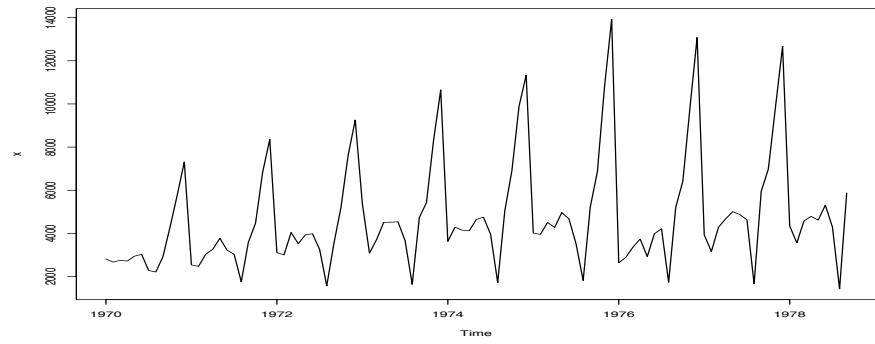


FIG. 2.2. Ventes de champagne

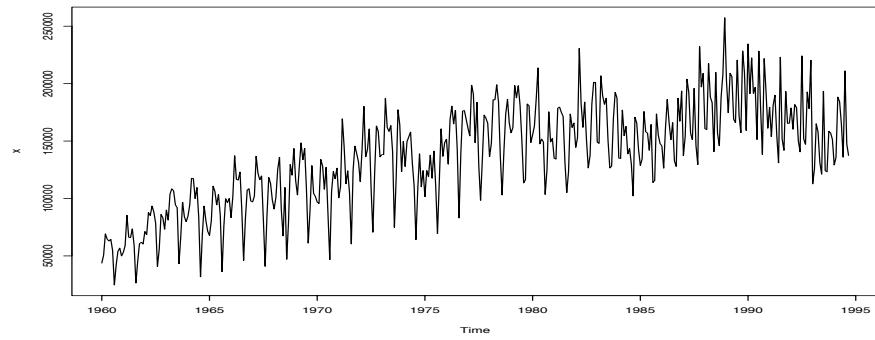


FIG. 2.3. Ventes de voitures

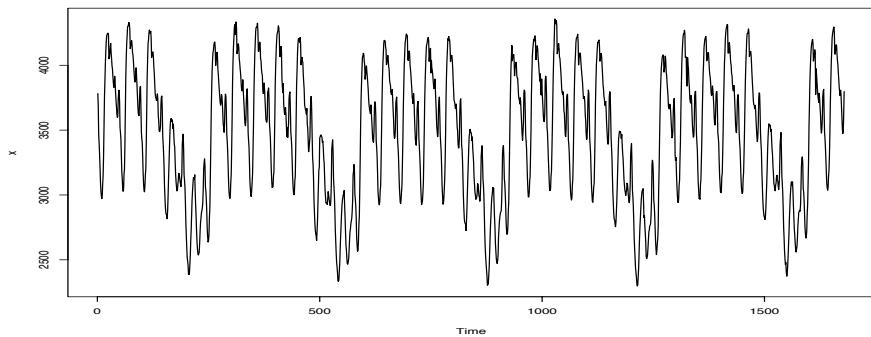


FIG. 2.4. Consommation d'électricité

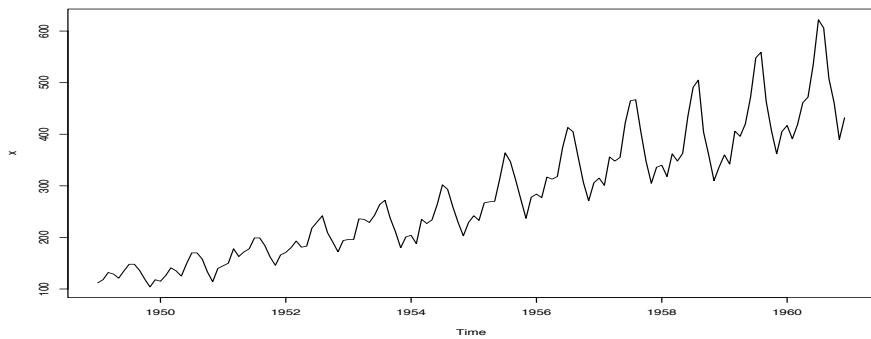


FIG. 2.5. Trafic aérien

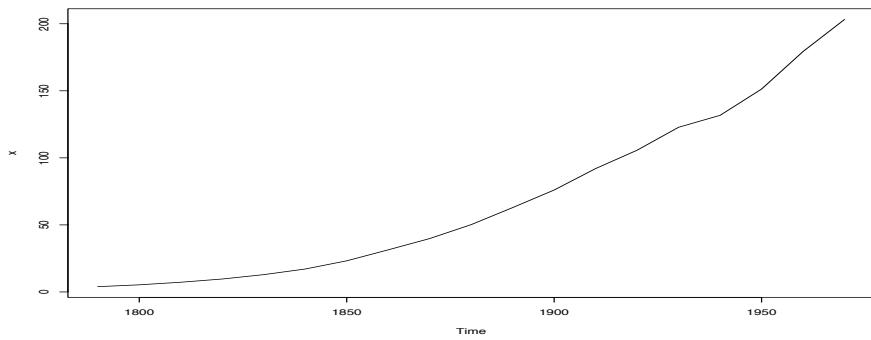


FIG. 2.6. Population des États-Unis

## 2. Tendances, composantes saisonnières

Il est temps de définir ces notions qui ont été évoquées au paragraphe précédent.

## 2.1. Tendances.

- On parle de tendance linéaire lorsque la série peut se décomposer en

$$x_n = an + b + e_n \quad n = 1, 2, \dots$$

Plus généralement, on parle de tendance polynômiale lorsque la série peut se décomposer en

$$x_n = a_1 n^p + a_2 n^{p-1} + \dots + a_{p+1} + e_n \quad n = 1, 2, \dots$$

expression dans laquelle  $e_n$  est un résidu où ne figure plus la tendance et qui, de ce fait, a une allure relativement homogène dans le temps. Ceci sera précisé par la suite.

De même on pourra définir des tendances logarithmiques, exponentielles etc....Ainsi, le graphe de l'exemple 2.6 semble à première vue présenter une tendance polynômiale ou exponentielle. Pour trancher entre les deux, on peut tracer le graphe du logarithme des données et voir si il présente une tendance linéaire.

- La tendance peut être multiplicative dans certaines séries :

$$x_n = t_n e_n \quad n = 1, 2, \dots$$

où  $t_n$  prend l'une des formes (linéaire, polynômiale etc...) évoquées plus haut. C'est alors le logarithme des données (si elles sont positives !) qui présente une tendance additive.

## 2.2. Composantes périodiques, saisonnalités.

- On parle de composante périodique, ou de périodicité, lorsque la série se décompose en

$$x_n = s_n + e_n \quad n = 1, 2, \dots$$

où  $s_n$  est périodique (c'est à dire  $s_{n+T} = s_n$  pour tout  $n$ , où  $T$  est la période, supposée entière) et où  $e_n$  est un résidu non-périodique et sans tendance.

Par exemple  $s_n = \cos \frac{n\pi}{6}$  a une période  $T = 12$ .

Lorsque la période  $T$  est de 6 mois ou un an, comme c'est le cas dans beaucoup de phénomènes socio-économiques, on a plutôt l'habitude de parler de saisonnalité ou de composante saisonnière.

- Comme précédemment, il arrive qu'on ait recours à un modèle multiplicatif.

## 2.3. Superposition de tendances et de périodes.

On peut bien imaginer des modèles de type

$$x_n = s_n^{(1)} + s_n^{(2)} + t_n + e_n \quad n = 1, 2, \dots$$

dans lequel se superposent deux séries périodiques de périodes  $T_1$  et  $T_2$  et une tendance. Moyennant une légère modification (en l'occurrence le passage au logarithme qui gomme l'effet d'augmentation de la dispersion) la série des ventes de champagne (exemple 2.2) est assez bien expliquée de cette façon, avec une tendance linéaire et la superposition de deux périodes de 6 mois et d'un an. Le commentaire est le même pour le trafic aérien. On y reviendra.

## CHAPITRE 3

# Lissages exponentiels

### 1. Introduction

Dans l'industrie il est courant que l'on doive effectuer des prévisions de vente de production ou de stock pour une centaine de produits. Ces prévisions sont effectuées à partir de données par exemple mensuelles et l'horizon de prévision est court, un an maximum. Un certain nombre de techniques "autoprojectives" sont regroupées sous la rubrique du lissage exponentiel. Ces approches construisent des prévisions assez précises. Elles sont peu coûteuses et les calculs peuvent être rendus très rapides par l'utilisation de formules récursives de mise à jour.

Elles ont beaucoup de succès dans les milieux dans lesquels un grand nombre de chroniques unidimensionnelles doivent être analysées séparément dans un but de prévision. Les succès sont importants malgré l'absence de bases théoriques solides comparables à celles des méthodes ARMA, ARIMA et SARIMA.

Nous présentons les bases empiriques des techniques et précisons les conditions d'utilisation optimales.

Les procédures automatiques présentent quelques dangers. On risque des mauvaises prévisions si la technique utilisée n'est pas adaptée à la chronique analysée.

Le lissage exponentiel est brièvement présenté dans Gourieroux-Monfort ([4]). Une présentation détaillée figure dans Mélard ([5]) page 139-170. Un article de Gardner ([3]) fait un point sur la question.

**1.1. Idées générales.** On dispose d'une série chronologique  $x = (x_1, \dots, x_n)$  de longueur  $n$  enregistrée aux dates  $1, \dots, n$ . On se situe à la date  $n$  et on souhaite prévoir la valeur  $x_{n+h}$  non encore observée à l'horizon  $h$ . On note cette prévision :  $\hat{x}_{n,h}$ .

L'entier  $n$  est parfois appelé base de la prévision.

La plus ancienne des techniques est celle du lissage exponentiel simple. Dans toutes ces techniques il s'agit d'ajuster à la chronique, localement, une fonction simple :

- une constante dans le lissage exponentiel simple,
- une droite dans le lissage exponentiel double,
- des fonctions polynomiales ou périodiques dans les lissages plus généraux.

### 2. Lissage exponentiel simple

Dans cette partie la prévision très simple, trop simple, ne dépend pas de l'horizon  $h$  de prévision. Nous conservons l'indice pour être cohérent avec la suite.

**2.1. Définition.** La prévision  $\hat{x}_{n,h}$  est construite :

- en prenant en compte toute l'histoire de la chronique,
- de sorte que, plus on s'éloigne de la base  $n$  de la prévision, moins l'influence des observations correspondantes est importante. Cette décroissance de l'influence est de type exponentiel. De là vient le nom de la technique.

On se donne  $\alpha$ , appelé constante de lissage, avec  $0 < \alpha < 1$ , et on définit (la prévision  $\hat{x}_{n,h}$  ne dépend pas de  $h$ ) :

$$(1) \quad \hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Remarquons que plus  $\alpha$  est voisin de 1, plus l'influence des observations éloignées dans le temps de la base  $n$  est grande.

**2.2. Formules récursives de mise à jour.** La définition (1) vérifie les formules de récurrences suivantes

$$(2) \quad \hat{x}_{n,h} = (1 - \alpha)x_n + \alpha\hat{x}_{n-1,h}$$

$$(3) \quad \hat{x}_{n,h} = \hat{x}_{n-1,h} + (1 - \alpha)(x_n - \hat{x}_{n-1,h})$$

Par conséquent, la connaissance de la prévision à l'horizon sur la base  $n - 1$ , soit  $\hat{x}_{n-1,h}$ , et l'observation  $x_n$  suffisent pour calculer immédiatement la prévision suivante  $\hat{x}_{n,h}$ .

Cependant, l'utilisation des formules récursives nécessite d'initialiser la récurrence en prenant par exemple  $\hat{x}_{1,h} = x_1$ , on remarque que pour  $n$  assez grand la valeur initiale a peu d'influence.

Remarquons aussi que la formule (2) permet d'interpréter  $\hat{x}_{n,h}$  comme le barycentre de  $x_n$  et de  $\hat{x}_{n-1,h}$  affectés respectivement des masses  $1 - \alpha$  et  $\alpha$ . L'influence de la dernière observation est d'autant plus grande que  $\alpha$  est proche de 0.

**2.3. Interprétation géométrique.** Asymptotiquement  $\hat{x}_{n,h}$  est la meilleure approximation au sens des moindres carrés (pour des pondérations exponentielles) de  $x$  par le vecteur constant  $a\delta$  où  $\delta = (1, \dots, 1) \in \mathbb{R}^n$ .

En effet la valeur de  $a$  qui minimise

$$(4) \quad \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - a)^2.$$

est fournie par :

$$\hat{a}(n) = \frac{1 - \alpha}{1 - \alpha^n} \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Si l'on transforme légèrement le modèle en considérant que

$$x = (x_1, \dots, x_n) = (\dots, 0, \dots, 0, x_1, \dots, x_n)$$

c'est à dire que les coordonnées du nouveau vecteur sont nulles pour tous les indices négatifs ou nul. On cherche donc la valeur de  $a$  qui minimise

$$\sum_{j=0}^{\infty} \alpha^j (x_{n-j} - a)^2.$$

la solution est fournie par

$$\hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Il est facile de montrer que les deux estimateurs sont asymptotiquement équivalents, on a

$$\lim_{n \rightarrow \infty} \hat{a}(n) = \lim_{n \rightarrow \infty} \hat{x}_{n,h}.$$

Dans la minimisation de (4) là aussi, l'influence des observations décroît exponentiellement lorsqu'on s'éloigne de la base. On peut dire que l'ajustement par la constante  $a$  est local et se fait au voisinage de  $n$ . La taille réelle de la plage des observations influentes dépend de  $\alpha$ . Plus  $\alpha$  est grand plus la taille est grande.

**2.4. Choix de la constante de lissage.** La méthode n'est pas bonne en présence de tendance, de composante saisonnière ou de fluctuations de hautes fréquences. En pratique le lissage exponentiel simple est assez peu utilisé.

Des valeurs de  $\alpha$  comprises entre 0.7 et 0.99 produisent des prévisions qui tiennent compte du passé lointain. Elles sont rigides et peu sensibles aux variations conjoncturelles. Il n'en n'est plus de même si  $\alpha$  est compris entre 0.01 et 0.3, la prévision est souple c'est à dire influencée par les observations récentes.

On peut aussi déterminer un  $\alpha$  adapté à la prévision d'une série donnée. Pour un horizon  $h$  on cherche la valeur de  $\alpha$  qui minimise  $\sum_{t=1}^{n-h} (x_{t+h} - (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j x_{t-j})^2$ , c'est à dire la valeur qui en moyenne réalise la meilleure prévision à l'horizon  $h$  sur les bases  $1, 2, \dots, n - h$  en utilisant le critère des moindres carrés. On choisit une grille de valeurs de  $\alpha$  pour lesquelles on calcule le critère et on retient la valeur qui le minimise. Cette technique de choix de constante peut être reprise dans les méthodes qui vont suivre. Les calculs sont souvent plus longs car ce n'est, en fait, plus une constante de lissage que nous sommes amenés à déterminer mais souvent deux ou trois.

### 3. Lissage exponentiel amélioré, ou double

**3.1. Définition.** On ajuste au voisinage de  $n$  une droite d'équation  $y_t = a_1 + a_2(t - n)$ . Le prédicteur sera :

$$(5) \quad \hat{x}_{n,h} = \hat{a}_1 + \hat{a}_2 h.$$

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

$$(6) \quad \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (\hat{a}_1 - \hat{a}_2 j))^2 = \inf_{a_1 \in R, a_2 \in R} \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2.$$

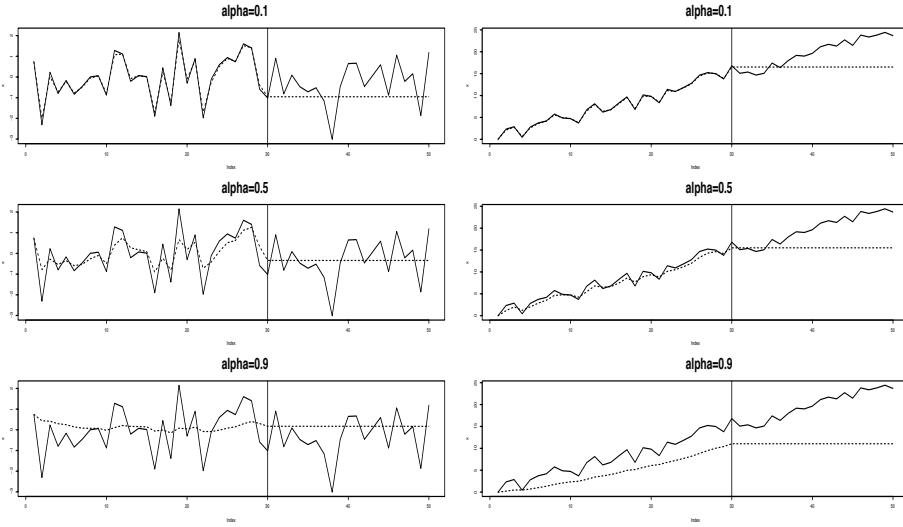


FIG. 3.1. Lissage et prévision par la méthode de lissage exponentiel simple d'un bruit blanc [gauche] et d'une tendance linéaire  $X(t) = 0.5t + \epsilon_t$ ,  $\epsilon_t \sim N(0, 1)$  [droite] ( $\alpha = .1, .5, .9$ )

En remplaçant, comme dans le lissage simple, la sommation finie par une somme de 0 à  $+\infty$ , les équations précédentes deviennent

$$\sum_{j=0}^{\infty} \alpha^j (x_{n-j} - (\hat{a}_1 - \hat{a}_2 j))^2 = \inf_{a_1 \in R, a_2 \in R} \sum_{j=0}^{\infty} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2.$$

**3.2. Calcul des coefficients.** On annule les dérivées partielles par rapport à  $a_1$  et  $a_2$  du critère (6) et on vérifie à l'aide des dérivées secondes que l'on a bien un minimum. En notant :  $C = C(a_1, a_2) = \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2$

$$\begin{aligned} \frac{\delta C}{\delta a_1} &= -2 \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j)) \\ \frac{\delta C}{\delta a_2} &= 2 \sum_{j=0}^{n-1} j \alpha^j (x_{n-j} - (a_1 - a_2 j)), \end{aligned}$$

on a :

$$(7) \quad \begin{aligned} \sum_{j=0}^{n-1} \alpha^j x_{n-j} - a_1 \sum_{j=0}^{n-1} \alpha^j + a_2 \sum_{j=0}^{n-1} j \alpha^j &= 0 \\ \sum_{j=0}^{n-1} j \alpha^j x_{n-j} - a_1 \sum_{j=0}^{n-1} j \alpha^j + a_2 \sum_{j=0}^{n-1} j^2 \alpha^j &= 0 \end{aligned}$$

Puisque  $\sum_{j=0}^{\infty} \alpha^j = \frac{1}{1-\alpha}$ ,  $\sum_{j=0}^{\infty} j\alpha^j = \frac{\alpha}{(1-\alpha)^2}$ ,  $\sum_{j=0}^{\infty} j^2\alpha^j = \frac{\alpha(1+\alpha)}{(1-\alpha)^3}$ , les relations (7) deviennent :

$$(1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j} - a_1 + a_2 \frac{\alpha}{1 - \alpha} = 0$$

$$(1 - \alpha)^2 \sum_{j=0}^{n-1} j\alpha^j x_{n-j} - a_1 \alpha + a_2 \frac{\alpha(1 + \alpha)}{1 - \alpha} = 0.$$

Introduisons deux lissages exponentiels simples successifs, celui des  $x$  produit  $L_1$ , celui de  $L_1$  produit  $L_2$  :

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

$$L_2(n) = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j L_1(n - j).$$

On a donc

$$L_2(n) - (1 - \alpha)L_1(n) = (1 - \alpha)^2 \sum_{j=1}^{n-1} j\alpha^j x_{n-j},$$

et  $L_1$  et  $L_2$  sont liés à  $a_1$  et  $a_2$  par les relations

$$L_1(n) - a_1 + a_2 \frac{\alpha}{1 - \alpha} = 0$$

$$L_2(n) - (1 - \alpha)L_1(n) - a_1 \alpha + a_2 \frac{\alpha(1 + \alpha)}{1 - \alpha} = 0.$$

Les solutions cherchées vérifient donc,

$$(8) \quad \begin{aligned} \hat{a}_1(n) &= 2L_1(n) - L_2(n) \\ \hat{a}_2(n) &= \frac{1 - \alpha}{\alpha} [L_1(n) - L_2(n)]. \end{aligned}$$

Soit aussi :

$$(9) \quad \begin{aligned} L_1(n) &= \hat{a}_1(n) - \frac{\alpha}{1 - \alpha} \hat{a}_2(n) \\ L_2(n) &= \hat{a}_1(n) - \frac{2\alpha}{1 - \alpha} \hat{a}_2(n). \end{aligned}$$

### 3.3. Formules récursives de mise à jour.

Commençons par  $L_1(n)$  et  $L_2(n)$ .

$$(10) \quad L_1(n) = (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j x_{n-j} = (1 - \alpha)x_n + \alpha L_1(n - 1)$$

$$(11) \quad \begin{aligned} L_2(n) &= (1 - \alpha)L_1(n - 1) + \alpha L_2(n - 1) \\ &= (1 - \alpha)^2 x_n + \alpha(1 - \alpha)L_1(n - 1) + \alpha L_2(n - 1). \end{aligned}$$

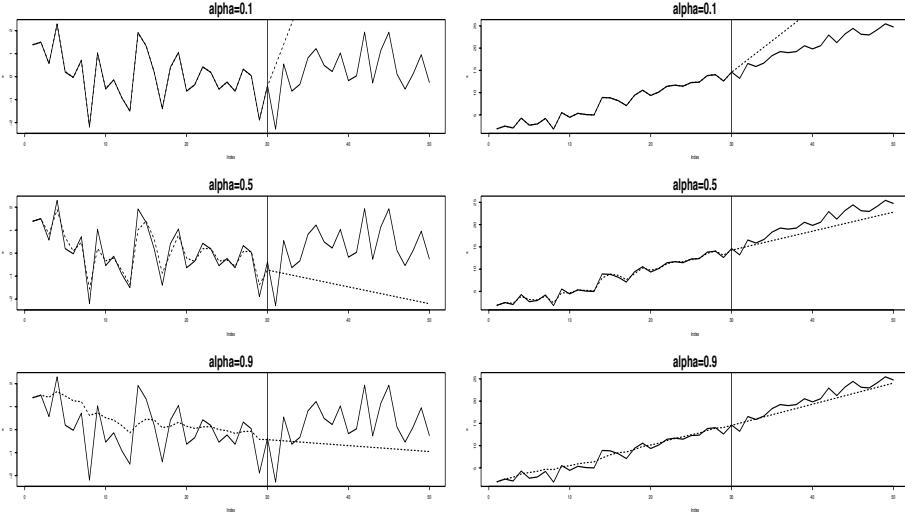


FIG. 3.2. Lissage et prévision par la méthode de lissage exponentiel double d'un bruit blanc [gauche] et d'une tendance linéaire  $X(t) = 0.5t + \epsilon_t$ ,  $\epsilon_t \sim N(0, 1)$ . [droite] ( $\alpha = .1, .5, .9$ )

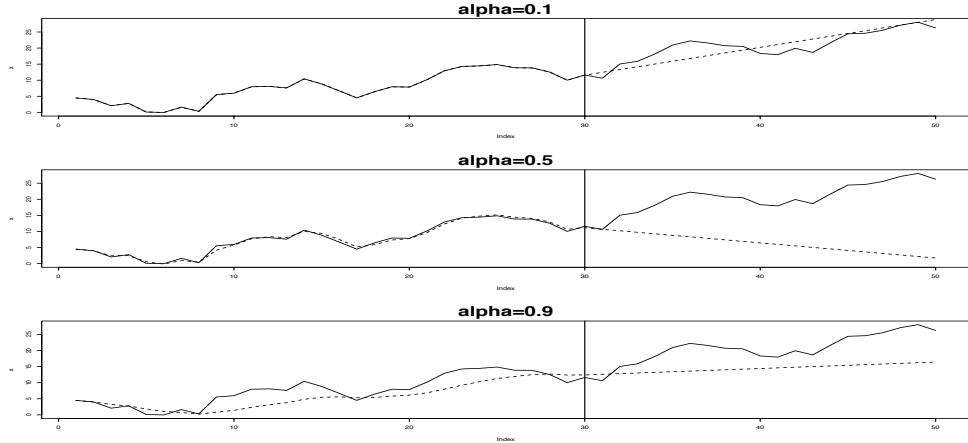


FIG. 3.3. Lissage et prévision par la méthode de lissage exponentiel double sur la série  $X(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6)$ ,  $\epsilon_t \sim N(0, 1)$ . ( $\alpha = .1, .5, .9$ )

récrivons (8) en utilisant (10) et (11).

$$\begin{aligned}\hat{a}_1(n) &= 2[(1-\alpha)x_n + \alpha L_1(n-1)] - [(1-\alpha)^2 x_n + \alpha(1-\alpha)L_1(n-1) + \alpha L_2(n-1)] \\ &= (1-\alpha^2)x_n + \alpha(1+\alpha)L_1(n-1) - \alpha L_2(n-1),\end{aligned}$$

en tenant compte de (9)

$$\begin{aligned}\hat{a}_1(n) &= (1 - \alpha^2)x_n + \alpha(1 + \alpha)[\hat{a}_1(n - 1) - \frac{\alpha}{1 - \alpha}\hat{a}_2(n - 1)] \\ &\quad - \alpha[\hat{a}_1(n - 1) - \frac{2\alpha}{1 - \alpha}\hat{a}_2(n - 1)], \\ &= (1 - \alpha^2)x_n + \alpha^2[\hat{a}_1(n - 1) + \hat{a}_2(n - 1)].\end{aligned}$$

Les formules analogues pour  $\hat{a}_2(n)$  sont :

$$\begin{aligned}\hat{a}_2(n) &= \frac{1 - \alpha}{\alpha}[\alpha(1 - \alpha)x_n + \alpha^2L_1(n - 1) - \alpha L_2(n - 1)] \\ &= (1 - \alpha)^2x_n + \alpha(1 - \alpha)L_1(n - 1) - (1 - \alpha)L_2(n - 1) \\ &= (1 - \alpha)^2x_n - (1 - \alpha)^2\hat{a}_1(n - 1) + (2\alpha - \alpha^2)\hat{a}_2(n - 1).\end{aligned}$$

En utilisant (5) avec  $h=1$  et  $n - 1$  comme instant de base, on obtient :

$$(12) \quad \hat{a}_1(n) = \hat{a}_1(n - 1) + \hat{a}_2(n - 1) + (1 - \alpha^2)[x_n - \hat{x}_{n-1,1}]$$

$$(13) \quad \hat{a}_2(n) = \hat{a}_2(n - 1) + (1 - \alpha)^2[x_n - \hat{x}_{n-1,1}].$$

On peut prendre comme valeurs initiales  $\hat{a}_1(0) = x_1, \hat{a}_2(0) = x_2 - x_1$ .

#### 4. Méthode de Holt-Winters

**4.1. Introduction.** Cette approche a pour but d'améliorer et de généraliser le L.E.S. Nous étudions plusieurs cas particuliers de cette méthode :

- (1) ajustement d'une droite affine (sans saisonnalité)
- (2) ajustement d'une droite affine + une composante saisonnière
- (3) ajustement d'une constante + une composante saisonnière

le dernier cas étant déduit du second. On ajuste au voisinage de l'instant  $n$  une droite

$$x_t = a_1 + (t - n)a_2.$$

La variante proposée par Holt-Winters porte sur les formules de mise à jour dans l'estimation de  $a_1$  et  $a_2$  par les formules (12) et (13). La formule (14) s'interprète comme le barycentre de l'observation  $x_n$  et de la prévision à l'horizon 1 à la date  $n - 1$ . La formule (15) s'interprète comme le barycentre de la pente estimée à l'instant  $n - 1$  et la différence  $[\hat{a}_1(n) - \hat{a}_1(n - 1)]$  entre les ordonnées à l'origine aux dates  $n$  et  $n - 1$ .

Soit  $0 < \beta < 1$  et  $0 < \gamma < 1$  deux constantes fixées et les formules de mise à jour

$$(14) \quad \hat{a}_1(n) = \beta x_n + (1 - \beta)[\hat{a}_1(n - 1) + \hat{a}_2(n - 1)]$$

$$(15) \quad \hat{a}_2(n) = \gamma[\hat{a}_1(n) - \hat{a}_1(n - 1)] + (1 - \gamma)\hat{a}_2(n - 1),$$

La prévision prend la forme :

$$(16) \quad \hat{x}_{n,h} = \hat{a}_1(n) + h\hat{a}_2(n).$$

Cette méthode est plus souple que le lissage exponentiel amélioré, dans la mesure où elle fait intervenir deux constantes  $\beta$  et  $\gamma$  au lieu d'une  $\alpha$ . Les équations (12) et (13), qui sont un cas particulier des formules (14) et (15) peuvent s'écrire :

$$\begin{aligned}\hat{a}_1(n) &= \alpha^2[\hat{a}_1(n-1) + \hat{a}_2(n-1)] + \\ &\quad +(1-\alpha^2)[x_n - \hat{x}_{n-1,1} + \hat{a}_1(n-1) + \hat{a}_2(n-1)]) \\ &= \alpha^2[\hat{a}_1(n-1) + \hat{a}_2(n-1)] + (1-\alpha^2)x_n \\ \hat{a}_2(n) &= \hat{a}_2(n-1) + \frac{(1-\alpha)^2}{1-\alpha^2}[\hat{a}_1(n) - \alpha^2[\hat{a}_1(n-1) + \hat{a}_2(n-1)]] \\ &\quad -(1-\alpha^2)[\hat{a}_1(n-1) + \hat{a}_2(n-1)]) \\ &= \hat{a}_2(n-1) + \frac{(1-\alpha)^2}{1-\alpha^2}[\hat{a}_1(n) - \hat{a}_1(n-1) - \hat{a}_2(n-1)],\end{aligned}$$

qui sont identiques à (14) et (15) si  $\beta = 1 - \alpha^2$  et  $\gamma = \frac{1-\alpha}{1+\alpha}$ . On remarque que si  $\beta$  et  $\gamma$  sont petits, le lissage est fort puisqu'alors  $\alpha$  est grand et que l'on tient compte du passé lointain.

**4.2. Ajout d'une composante saisonnière. Décomposition additive.** On ajuste au voisinage de  $n$ , la chronique

$$x_t = a_1 + (t-n)a_2 + s_t$$

où  $s_t$  est une composante périodique de période  $T$ . Les formules récursives de mise à jour sont : soit  $0 < \beta < 1$ ,

$$\begin{aligned}\hat{a}_1(n) &= \beta(x_n - \hat{s}_{n-T}) + (1-\beta)[\hat{a}_1(n-1) + \hat{a}_2(n-1)] \\ \hat{a}_2(n) &= \gamma[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1-\gamma)\hat{a}_2(n-1), \\ \hat{s}_n &= \delta[x_n - \hat{a}_1(n)] + (1-\delta)\hat{s}_{n-T}.\end{aligned}$$

La prévision prend la forme :

$$\begin{aligned}\hat{x}_{n,h} &= \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-T} \quad 1 \leq h \leq T \\ \hat{x}_{n,h} &= \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-2T} \quad T+1 \leq h \leq 2T\end{aligned}$$

et ainsi de suite pour  $2T < h$ . Dans cette procédure la difficulté de choisir les constantes de lissage est encore plus grande que pour les autres situations.

**Initialisation :** Les valeurs initiales suivantes permettent (mais on pourrait en proposer d'autres) de commencer le calcul récursif des prévisions à partir de l'instant  $T+2$  :

$$\begin{aligned}\hat{a}_1(T+1) &= x_{T+1} \\ \hat{a}_2(T+1) &= \frac{x_{T+1} - x_1}{T} \\ \hat{s}_j &= x_j - (x_1 + (T-1)\hat{a}_2(T+1))) \quad \text{pour } j \in \{1, \dots, T\}\end{aligned}$$

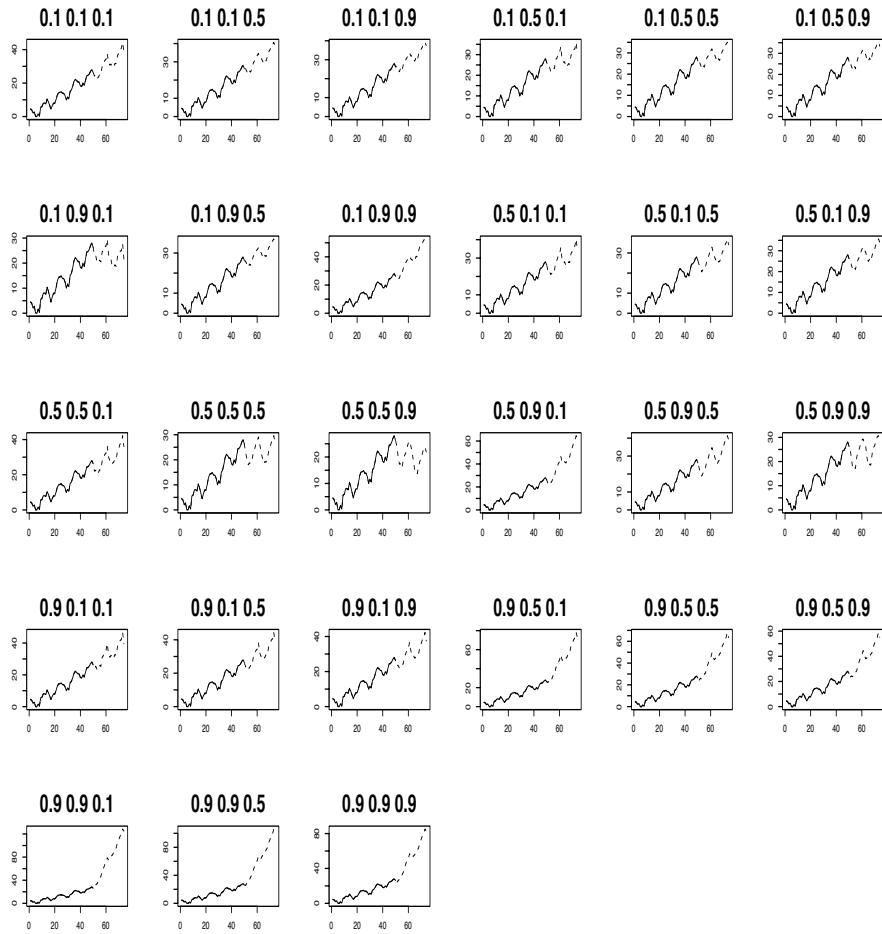


FIG. 3.4. Lissage et prévision par la méthode de Holt-Winters sur la série  $x(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6)$ ,  $\epsilon_t \sim N(0, 1)$ .

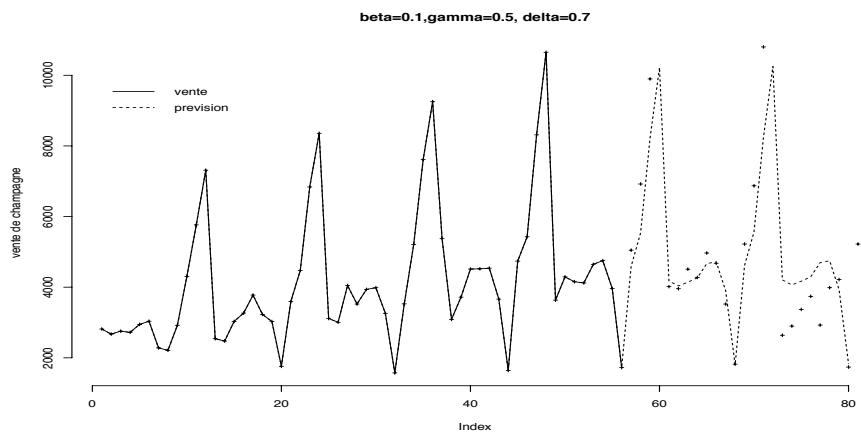


FIG. 3.5. Prévision par la méthode de Holt-Winters sur les données de vente de champagne.

## CHAPITRE 4

### Premiers indices descriptifs

Il est bien utile de disposer de quelques indices numériques qui "résument" une série  $(x_1, \dots, x_n)$ . Ils constituent aussi le premier pas vers l'estimation des modèles paramétriques.

#### 1. Indices de tendance centrale

Ce sont, classiquement, la moyenne et la médiane. La plupart du temps c'est la moyenne qui est utilisée (elle le sera systématiquement dans ce cours)

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j.$$

Il faut cependant garder en mémoire que la moyenne est un indice intéressant, en particulier pour sa robustesse par rapport aux valeurs extrêmes.

#### 2. Indices de dispersion

Les plus courants sont la variance empirique (ou plus exactement l'écart type empirique qui en est la racine carrée) et l'étendue (différence entre la plus grande valeur et la plus petite). Elles indiquent la dispersion des observations autour de leur indice de tendance centrale. Les commentaires précédents sur moyenne-médiane sont encore valables ici. Dans ce cours c'est la variance empirique qui sera utilisée :

$$\hat{\sigma}_n(0) = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2.$$

#### 3. Indices de dépendance

##### 3.1. Auto-covariances empiriques.

Ces indices  $\hat{\sigma}_n(1), \hat{\sigma}_n(2), \dots$  donnent une idée de la dépendance entre les données. Plus exactement

$$\hat{\sigma}_n(1) = \frac{1}{n} \sum_{j=1}^{n-1} (x_j - \bar{x}_n)(x_{j+1} - \bar{x}_n)$$

indique la dépendance entre deux données successives,

$$\hat{\sigma}_n(2) = \frac{1}{n} \sum_{j=1}^{n-2} (x_j - \bar{x}_n)(x_{j+2} - \bar{x}_n)$$

indique la dépendance entre deux données écartées de deux unités de temps et ainsi de suite jusqu'à

$$\hat{\sigma}_n(n-1) = \frac{1}{n} (x_1 - \bar{x}_n)(x_n - \bar{x}_n).$$

En réalité ce dernier indice, qui ne fait intervenir que la première et la dernière donnée, n'a pas grand sens statistique, surtout si la longueur  $n$  de la série est grande. Il en est de même pour  $\hat{\sigma}_n(n-1)$ . Le bon sens, ainsi que des considérations de convergence dépassant la limite de ce cours, recommandent de ne prendre en considération que les auto-covariances empiriques  $\hat{\sigma}_n(1), \dots, \hat{\sigma}_n(h)$  pour un  $h$  "pas trop grand" par rapport à la longueur de la série.

### 3.2. Auto-corrélations empiriques.

Ce sont les quotients des covariances par la variance :

$$\hat{\rho}_n(j) = \frac{\hat{\sigma}_n(j)}{\hat{\sigma}_n(0)}, \quad j = 0, \dots, h.$$

Évidemment  $\hat{\rho}_n(0) = 1$ , si bien qu'en passant des auto-covariances aux auto-corrélations on perd une information : la dispersion de la série autour de sa moyenne. Il faut garder ce fait en mémoire. Ceci dit, ce sont ces auto-corrélations qui caractérisent les dépendances. Pour le voir, il faut remarquer par exemple que

$$\hat{\rho}_n(1) = \frac{\sum_{j=1}^{n-1} (x_j - \bar{x}_n)(x_{j+1} - \bar{x}_n)}{\sum_{j=1}^n (x_j - \bar{x}_n)^2}$$

est presque (à vous de réfléchir à ce qui diffère) le coefficient de corrélation entre la série  $(x_1, x_2, \dots, x_{n-1})$  et la série décalée d'une unité de temps  $(x_2, x_3, \dots, x_n)$ .

### 3.3. Visualisation graphique des auto-corrélations empiriques.

En rassemblant ce que vous savez sur la régression linéaire, vous pouvez déduire de la remarque précédente que le "nuage"  $\mathcal{N}_1$  de points

$$(x_j, x_{j+1}), \quad j = 1, \dots, n-1,$$

(ce nuage est une autre représentation de la série), constitue une bonne illustration de la valeur de  $\hat{\rho}_n(1)$ . Si cette auto-corrélation est proche de  $\pm 1$  le nuage est allongé selon une droite. La pente de cette droite a le signe de  $\hat{\rho}_n(1)$ . Le nuage est d'autant plus "arrondi" que  $\hat{\rho}_n(1)$  est plus proche de zéro.

Un commentaire analogue (en remplaçant  $\hat{\rho}_n(1)$  par  $\hat{\rho}_n(k)$ ) peut être fait pour les nuages  $\mathcal{N}_k$  pour  $k = 2, \dots, h$  :

$$(x_j, x_{j+k}), \quad j = 1, \dots, n-k,$$

Les graphiques qui suivent représentent chacun une série simulée, (de longueur 500), le graphe de son auto-corrélation, les nuages  $\mathcal{N}_1, \dots, \mathcal{N}_8$  et le graphe de son auto-corrélation.

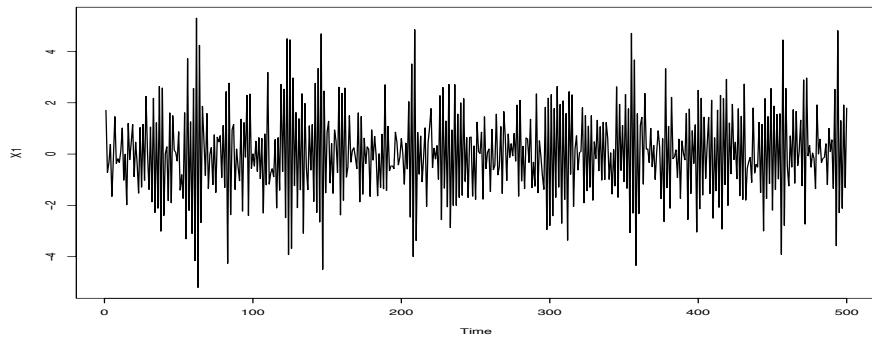
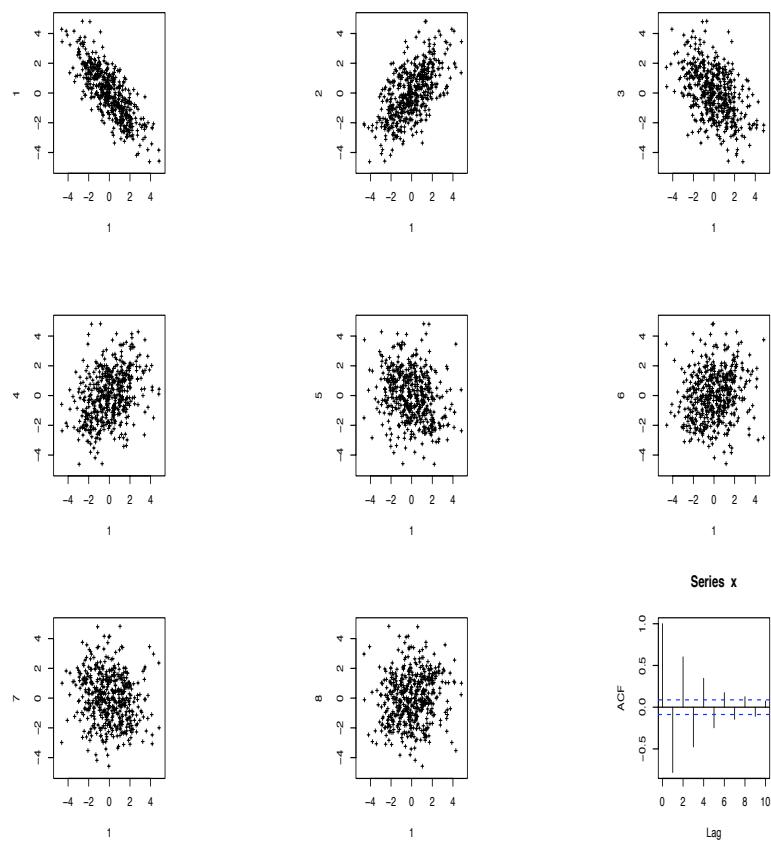


FIG. 4.1. Série 1

FIG. 4.2. Série 1 : Nuage de points  $\mathcal{N}_k$  pour  $k = 1, \dots, 8$  et auto-corrélation

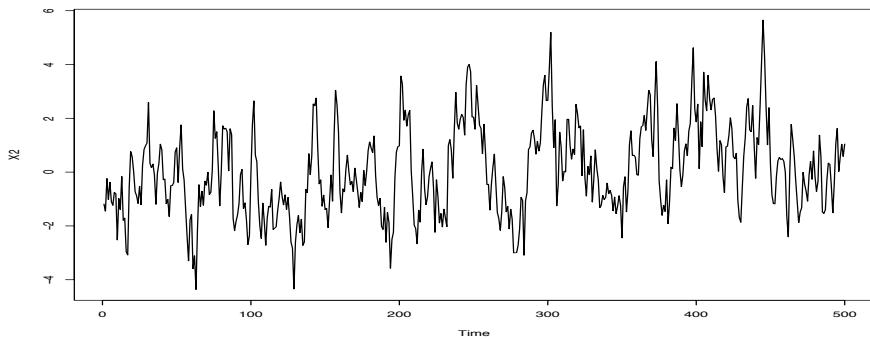
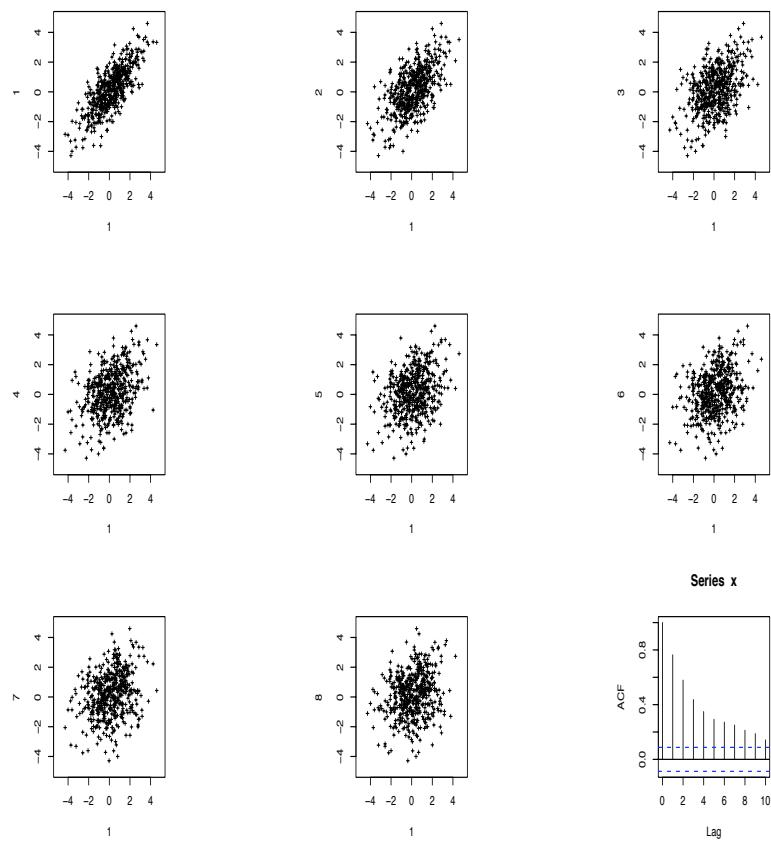


FIG. 4.3. Série 2

FIG. 4.4. Série 2 : Nuage de points  $\mathcal{N}_k$  pour  $k = 1, \dots, 8$  et auto-corrélation

#### 4. Comment utiliser l'auto-corrélation ?

L'auto-corrélation est surtout utilisée sur les séries résiduelles  $e_n$  qui s'ajoutent aux tendances et aux saisonnalités. La démarche habituelle consiste à retirer de la série la tendance que l'on a détectée, puis, si il y a lieu, de la désaisonnalisier (pour ce traitement préalable des données on consultera par exemple [1] ou [2]). Sur la série résiduelle on calcule les auto-covariances empiriques qui serviront (voir chapitre 8) à ajuster un modèle prédictif sur ces résidus.

Cependant il est très utile de comprendre quelle sera l'allure de cette auto-corrélation sur une série brute, comportant tendance ou/et saisonnalité. C'est le but des propositions qui suivent.

**PROPOSITION 4.1.** *On considère une tendance linéaire pure :*

$$x_j = aj + b \quad j = 1, \dots, n.$$

*Pour  $k$  fixé,  $\hat{\rho}_n(k) \rightarrow 1$  quand  $n$  tend vers l'infini.*

La preuve est laissée en exercice. Donner d'abord l'expression de la moyenne. Montrer ensuite que  $\hat{\rho}_n(k)$  ne dépend ni de  $a$  ni de  $b$ . Enfin, utiliser les résultats connus sur les sommes de puissances d'entiers successifs

$$1 + 2 + \dots + N = \frac{N(N+1)}{2} \quad \text{et} \quad 1 + 2^2 + \dots + N^2 = \frac{N(N+1)(2N+1)}{6},$$

et terminer en évitant les calculs inutiles !

Le résultat est le même pour toute tendance polynomiale. On l'admettra.

**PROPOSITION 4.2.** *On considère une série périodique pure :*

$$x_j = a \cos \frac{2j\pi}{T} \quad j = 1, \dots, n.$$

*Pour  $k$  fixé,  $\hat{\rho}_n(k) \rightarrow \cos \frac{2k\pi}{T}$  quand  $n$  tend vers l'infini.*

On admettra ce résultat.

Il indique que, si la série est assez longue, la période ( $T$ , en l'occurrence) se voit aussi bien sur l'auto-corrélation que sur la série elle-même.

En réalité, on n'a jamais à faire avec une tendance ou une composante périodique pures, mais les propositions précédentes donnent une bonne indication sur ce qui se passe pour une série présentant une tendance ou une composante périodique additives.

En résumé, pour une longue série présentant une tendance polynomiale, l'auto-corrélation  $\hat{\rho}_n(k)$  est positive et assez proche de 1 lorsque  $k$  évolue tout en restant petit devant  $n$ . Pour une série présentant une périodicité, cette périodicité se voit bien sur l'auto-corrélation. Ces deux constatations contrastent avec ce que l'on verra plus loin (chapitre 4) pour les séries résiduelles, où  $\hat{\rho}_n(k)$  devient assez vite très petit quand  $k$  grandit.

On conclut que si l'auto-corrélation du résidu présente une des propriétés ci-dessus, c'est sans doute parce que l'on a oublié d'enlever une tendance ou que l'on a mal désaisonné.

Les graphiques 3-5 montrent une série  $t_n = an + b \cos(\frac{\pi n}{3}) + c \cos(\frac{\pi n}{6})$ , superposition d'une tendance linéaire et de deux saisonnalités (périodes 6 et 12), puis une simulation de la série bruitée  $x_n = t_n + e_n$ . On donne la fonction d'auto-corrélation de cette série bruitée, puis l'auto-covariance de la série  $y_n = x_n - x_{n-1}$ , et enfin celle de la série  $z_n = y_n - y_{n-12}$ .

Les graphiques 3-6, 3-7 et 3-8 montrent l'auto-corrélation empirique des ventes de champagne, celle de la série différenciée ( $y_n = x_n - x_{n-1}$ ), puis celle de la série  $y_n$  désaisonnalisée ( $z_n = x_n - x_{n-12}$ ).

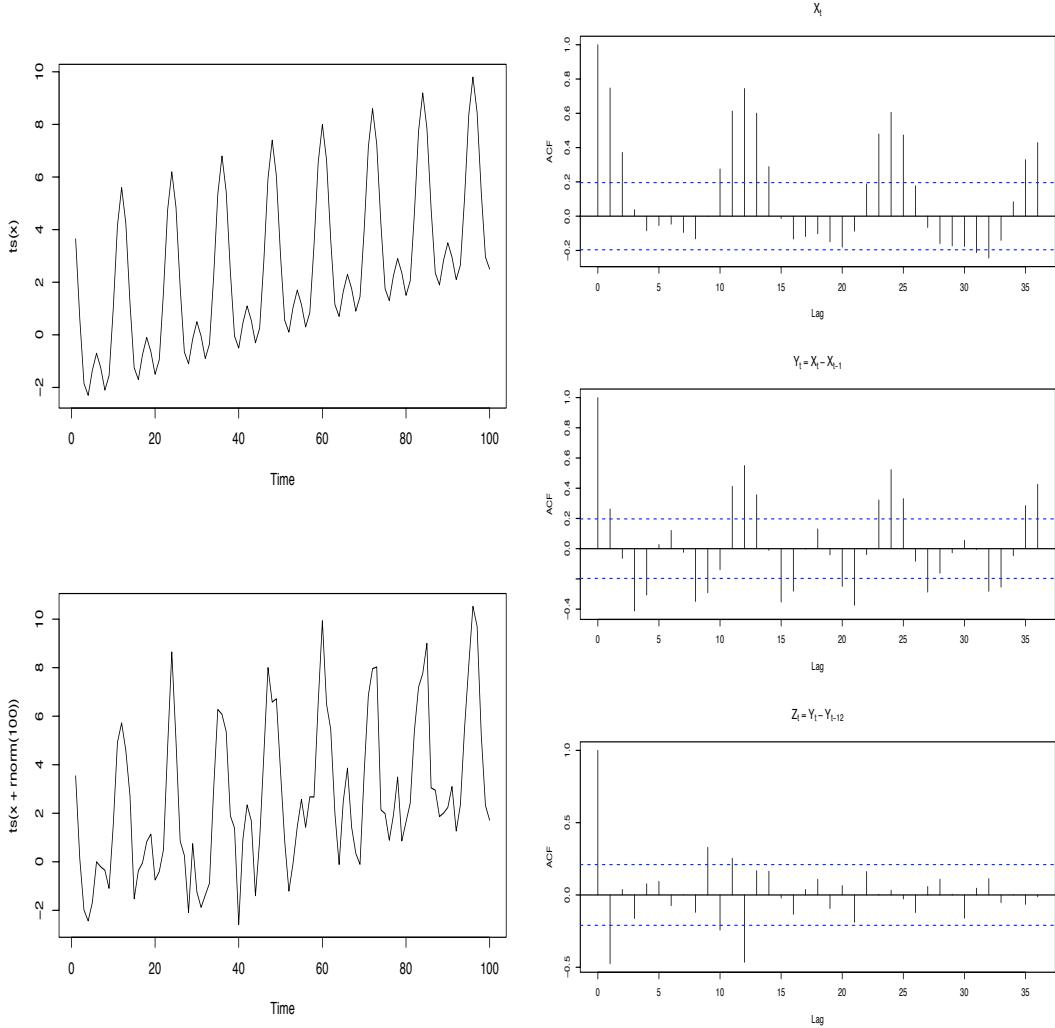


FIG. 4.5. Série simulée non bruitée (en haut à droite), bruitée (en bas à droite). Auto-corrélation de la série bruitée, de la série bruitée différenciée, puis de la série bruitée différenciée et désaisonnalisée

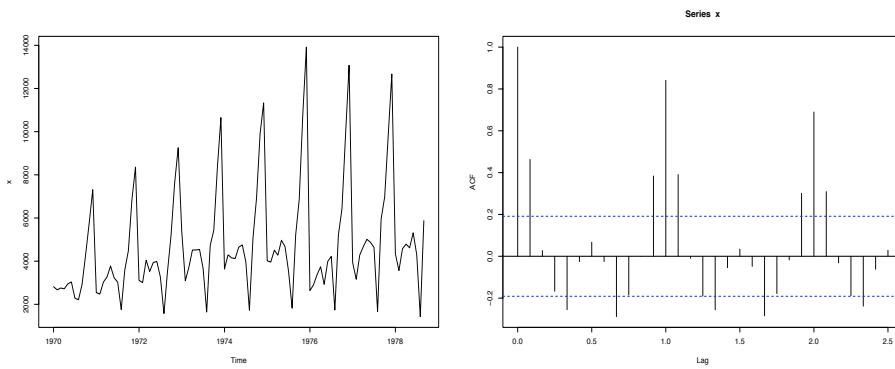
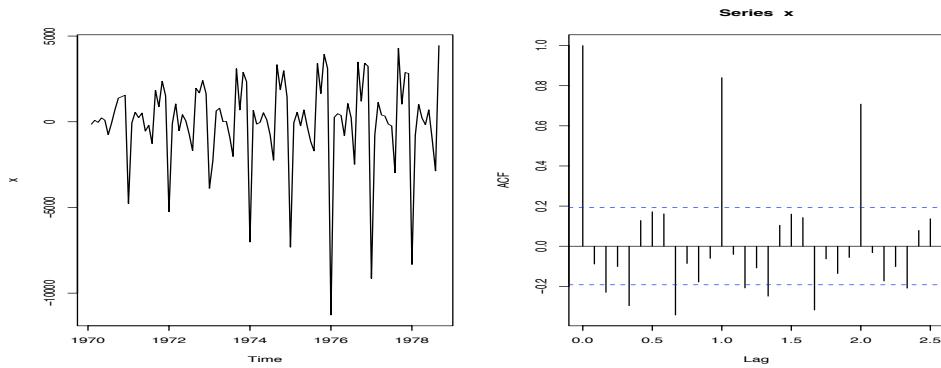
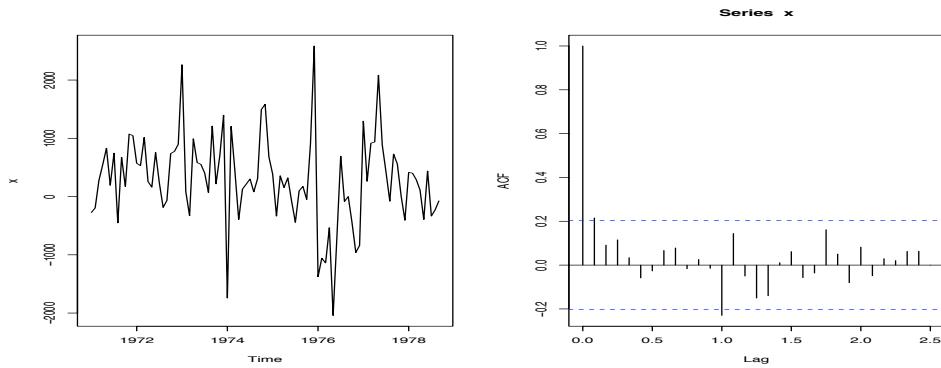


FIG. 4.6. Champagne : série et auto-corrélation

FIG. 4.7. Champagne : série différenciée  $X_t - X_{t-1}$  et auto-corrélationFIG. 4.8. Champagne : série différenciée et désaisonnalisée :  $X_t - X_{t-12}$  et auto-corrélation

## CHAPITRE 5

### Vers la modélisation : suites stationnaires de variables aléatoires

Ce chapitre est un premier pas vers la représentation de la série résiduelle par un modèle théorique qui se prête bien à la prévision. Il va permettre par la suite de travailler sur la série  $(x_1, x_2, \dots, x_n)$  comme si elle était la réalisation des  $n$  premiers termes d'une suite infinie  $(X_1, X_2, \dots) = (X_n)_{n \in \mathbb{N}}$  de variables aléatoires. Pour mieux comprendre ce qui suit, il faut se rendre compte que pour pouvoir prévoir de façon efficace le futur d'une suite finie  $(X_1, \dots, X_n)$  de variables aléatoires, il *faut* que ces variables admettent entre elles un peu de dépendance. Ce n'est par exemple pas la peine d'essayer de prévoir ce que va donner le cinquantième lancer d'une pièce de monnaie lorsqu'on a les résultats des quarante-neuf précédents. L'erreur de prévision sera de toutes façons de 50%. Par contre, si on sait que la donnée  $n + 1$  dépend des données précédentes, et si par chance on a une idée de la nature de cette dépendance, on pourra rechercher une procédure de prévision adaptée. Voilà pourquoi il faut consacrer quelque temps aux suites de variables dépendantes.

Par commodité mathématique, on considérera des suites dont l'indice peut être négatif. Toutes les variables aléatoires considérées dans ce chapitre ont un moment d'ordre deux fini.

#### 1. Définitions. Auto-covariance et auto-corrélation

**DÉFINITION 5.1.** *On dit qu'une suite de variables aléatoires  $(\dots, X_{-1}, X_0, X_1, \dots)$  est stationnaire si l'espérance de  $X_n$  ne dépend pas de  $n$*

$$\mathbb{E}(X_n) =: \mu \quad \forall n$$

*et si les covariances sont stables par translation du temps c'est-à-dire, pour tout  $h$ ,*

$$\text{Cov}(X_n, X_{n+h}) =: \sigma(h) \quad \forall n.$$

On rappelle que la covariance est définie par

$$\text{Cov}(U, V) := \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) = \mathbb{E}((U - \mathbb{E}(U))(V - \mathbb{E}(V))).$$

Une telle suite est appelée *processus stationnaire*.

On remarque que  $\text{Var}(X_n) = \sigma(0)$  pour tout  $n$ . Donc la variance est constante.

Si l'espérance  $\mu$  est nulle on dit que le processus est *centré*. Dans la suite, nos processus seront centrés. C'est bien normal si on pense qu'ils sont censés modéliser des résidus.

### DÉFINITION 5.2.

1- La suite  $\sigma(h)$  est appelée fonction d'auto-covariance du processus stationnaire.

2- La suite  $\rho(h) := \frac{\sigma(h)}{\sigma(0)}$  est sa fonction d'auto-corrélation.

On remarque que  $\sigma(h)$  et  $\rho(h)$  ont les propriétés élémentaires suivantes :

- Ces suites sont symétriques : pour tout  $h \geq 0$ ,  $\sigma(h) = \sigma(-h)$ ; même chose pour  $\rho(h)$ .
- Puisque c'est une variance,  $\sigma(0) \geq 0$ .
- Par définition,  $\rho(0) = 1$ .

L'égalité à zéro de la variance  $\sigma(0)$  est possible mais tout à fait sans intérêt, car alors la probabilité que  $X_n = \mu$  est égale à 1 : la suite est constante. On supposera dans la suite que  $\sigma(0) > 0$ , ce qui justifie l'utilisation de la corrélation.

DÉFINITION 5.3. On dit que le processus  $(X_n)_{n=\pm 1, \dots}$  est gaussien si toutes les lois marginales sont gaussiennes. C'est à dire si, quels que soient  $k$  et  $j_1, \dots, j_k$ , le vecteur  $(X_{j_1}, \dots, X_{j_k})$  est gaussien.

Il est clair qu'un processus gaussien stationnaire est strictement stationnaire, au sens que pour tout ensemble d'indices  $k$  et  $j_1, \dots, j_k$  la loi des vecteurs  $(X_{j_1}, \dots, X_{j_k})$  et  $(X_{j_1+h}, \dots, X_{j_k+h})$  sont les mêmes quel que soit  $h$ .

EXERCICE 1. Soit  $(X_n)_{n=\pm 1, \dots}$  une suite de variables aléatoires gaussiennes indépendantes, centrées et toutes de variance 1. Parmi les processus suivants, lesquels sont gaussiens ? lesquels sont stationnaires ? Donner alors l'expression de leur fonction d'auto-covariance.

- $Y_n = X_n X_{n+1}$
- $Y_n = X_n X_{n+1} \cdots X_{n+k}$
- $Y_n = X_n^2 X_{n+1}^2 \cdots X_{n+k}^2$
- $Y_n = X_n^2 + X_{n+1}^2 + X_{n+2}^2$
- $Y_n = X_1 \cos\left(\frac{n\pi}{2}\right)$
- $Y_n = n + X_n$
- $Y_n = n X_n$
- $Y_n = X_n^2 X_{2n}^2$ .

△△

## 2. Exemples

### 2.1. Bruits blancs.

Ce sont les processus stationnaires les plus élémentaires. Ils ont une fonction d'auto-covariance nulle (en dehors de  $\sigma(0)$ ). Peu intéressants en eux-mêmes pour la prévision (voir la remarque en tête de chapitre), ils servent de base à la construction de tous les modèles ARMA.

DÉFINITION 5.4. Un bruit blanc est une suite de variables aléatoires non corrélées et d'espérance et de variance constante. Autrement dit, pour tout  $n$

$$\mathbb{E}(X_n) = \mu, \quad \text{Var}(X_n) = \sigma^2 \quad \text{et} \quad \text{Cov}(X_n, X_{n+h}) = 0 \quad \text{pour } h \neq 0.$$

Si l'espérance  $\mu$  est nulle, on dit que le bruit blanc est *centré*.

Le cas particulier le plus courant est la suite de variables aléatoires gaussiennes standard (espérance nulle et variance égale à 1) et indépendantes. C'est par exemple ce qui a été choisi dans l'exercice 1.

## 2.2. Moyennes mobiles.

Ce sont des processus obtenus à partir d'un bruit blanc.

DÉFINITION 5.5. Soit  $(\varepsilon_n)_{n=\pm 1, \dots}$  un bruit blanc centré de variance  $\sigma^2$ . On appelle moyenne mobile un processus de la forme

$$X_n = \varepsilon_n + a_1 \varepsilon_{n-1} + \cdots + a_q \varepsilon_{n-q},$$

où  $q$  est un nombre entier fixé non nul, appelé l'ordre de la moyenne mobile.

Un processus moyenne mobile est évidemment stationnaire, et on a

$$\begin{aligned} (17) \quad \sigma(0) &= \sigma^2(1 + a_1^2 + \cdots + a_q^2) \\ \sigma(1) &= \sigma^2(a_1 + a_1 a_2 + \cdots + a_{q-1} a_q) \\ &\vdots \\ \sigma(q) &= \sigma^2 a_1 a_q \\ \sigma(q+1) &= \sigma(q+2) = \cdots = 0 \end{aligned}$$

Nous retrouverons ces processus au chapitre 6.

## 2.3. Processus autorégressif d'ordre 1.

C'est l'exemple le plus simple de toute une famille qui sera introduite au chapitre 6.

DÉFINITION 5.6. Soit  $(\varepsilon_n)_{n=\pm 1, \dots}$  un bruit blanc centré de variance  $\sigma^2$  et  $a \in ]-1, 1[$  un paramètre fixé. Un processus autorégressif d'ordre 1 est un processus stationnaire centré qui vérifie les équations de récurrence

$$(18) \quad X_n = aX_{n-1} + \varepsilon_n \quad \forall n = 0, \pm 1, \dots$$

Tout lecteur attentif aura remarqué que cette définition implicite demande au moins une information supplémentaire, à savoir : existe-t-il un processus stationnaire centré solution de la récurrence (18) ?

La réponse est oui et la solution est unique (l'unicité est à admettre). C'est le processus défini comme la suite de sommes de séries :

$$(19) \quad \forall n, \quad X_n = \varepsilon_n + a\varepsilon_{n-1} + a^2\varepsilon_{n-2} + \cdots$$

EXERCICE 2. Montrer que la série (18) converge à la fois presque sûrement et en moyenne quadratique.  $\triangle\triangle$

En regardant l'expression (19), on voit qu'un tel processus n'est guère qu'un processus moyenne mobile d'ordre infini.

Son espérance et sa fonction d'auto-covariance s'obtiennent en utilisant les propriétés de la convergence dans  $L^2$ . On voit par exemple que

$$(20) \quad \mathbb{E}(X_n) = 0$$

$$(21) \quad \text{Var}(X_n) = \sigma^2(1 + a^2 + a^4 + \dots) = \sigma^2 \frac{1}{1 - a^2}$$

$$(22) \quad \sigma(h) = \sigma(-h) = \sigma^2(a^h + a^{h+2} + \dots) = \sigma^2 \frac{a^h}{1 - a^2} \quad \forall h \geq 0$$

$$(23) \quad \text{Cov}(X_n, \varepsilon_{n+k}) = 0 \quad \forall n \quad \text{et} \quad \forall k > 0.$$

On remarque, comme conséquence de (21) et (22), que l'auto-corrélation tend vers zéro à vitesse exponentielle quand l'écart  $h$  tend vers l'infini :

$$\rho(h) = a^{|h|} \quad \forall h$$

(on évitera de dire qu'il s'agit d'une *décroissance* exponentielle car si  $a$  est négatif, la suite est alternativement positive et négative).

**EXERCICE 3.** Les expressions (20), (21) et (22) peuvent se retrouver facilement à partir de la récurrence (18) et, pour les covariances  $\sigma(h)$ , en utilisant la nullité de la covariance entre le processus au temps  $n$  et le bruit futur (23). Faites ces calculs, ils vous aideront beaucoup par la suite.  $\triangle\triangle$

### 3. Auto-corrélation partielle

Cette notion, importante dans le domaine des séries temporelles, est un peu délicate et demande quelques rappels de la géométrie de  $L^2$ .

**3.1. Rappels.** Les variables aléatoires réelles de carré intégrable définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$  forment un espace vectoriel qui, si on le munit du produit scalaire  $\langle X, Y \rangle = \mathbb{E}(XY)$ , est un espace de Hilbert noté en général  $L^2(\Omega, \mathcal{A}, P)$ , ou plus brièvement  $L^2$  lorsqu'il n'y a pas d'ambiguité.

- *Orthogonalité.* Deux variables  $X$  et  $Y$  de  $L^2$  sont orthogonales lorsque  $\mathbb{E}(XY) = 0$ . Ainsi, en notant  $\mathbb{I}$  la variable constante et égale à 1, on ré-interprète  $\mathbb{E}(X) = 0$  comme étant l'orthogonalité de  $X$  et de  $\mathbb{I}$ .

- *Distance.* De même on définit la distance entre deux variables comme

$$d(X, Y) = \sqrt{\mathbb{E}((X - Y)^2)}.$$

3.1.1. *Espace engendré*. . On considère  $N$  variables aléatoires  $X_1, \dots, X_N$ . Le sous espace de Hilbert engendré par ces variables, qui sera noté  $[X_1, \dots, X_N]$  ou encore  $[X]_1^N$ , est l'ensemble de toutes les combinaisons linéaires  $\lambda_1 X_1 + \dots + \lambda_N X_N$  de ces variables.

3.1.2. *Projection.* . Le projeté d'une variable aléatoire  $Y$  sur l'espace  $[X]_1^N$ , qui sera noté

$$P_{[X]_1^N}(Y),$$

est la variable aléatoire de cet espace qui est la plus proche de  $Y$  au sens de la distance vue plus haut. C'est donc la combinaison linéaire  $P_{[X]_1^N}(Y) = \gamma_1 X_1 + \cdots + \gamma_N X_N$  telle que

$$(24) \quad \mathbb{E}(Y - \gamma_1 X_1 - \cdots - \gamma_N X_N)^2 \leq \mathbb{E}(Y - \lambda_1 X_1 - \cdots - \lambda_N X_N)^2 \quad \forall \lambda_1, \dots, \lambda_N.$$

En quelque sorte, le projeté est *la meilleure explication linéaire* de la variable  $Y$  en termes de  $X_1, \dots, X_N$ . C'est pourquoi on l'utilise pour *prévoir*  $Y$  lorsqu'on ne connaît que  $X_1, \dots, X_N$ .

- *Calcul du projeté.* Il est commode, pour calculer le projeté, d'utiliser la propriété suivante, facile à prouver

**PROPOSITION 5.1.** *Le projeté  $P_{[X_1, \dots, X_N]}(Y) = \gamma_1 X_1 + \cdots + \gamma_N X_N$  est caractérisé par le fait que  $Y - P_{[X_1, \dots, X_N]}(Y)$  est orthogonal aux variables  $X_1, \dots, X_N$ .*

Ceci signifie que les coefficients  $\gamma_1, \dots, \gamma_N$  vérifient les  $N$  équations

$$\mathbb{E}((Y - \gamma_1 X_1 - \cdots - \gamma_N X_N) X_k) = 0 \quad \forall k = 1, \dots, N.$$

Soit aussi

$$(25) \quad \gamma_1 \mathbb{E}(X_1 X_k) + \cdots + \gamma_N \mathbb{E}(X_N X_k) = \mathbb{E}(Y X_k) \quad \forall k = 1, \dots, N.$$

On voit donc que, si ce système d'équations est régulier, c'est à dire si la matrice

$$\Sigma_{[X]_1^N} = \begin{pmatrix} \mathbb{E}(X_1^2) & \dots & \mathbb{E}(X_1 X_N) \\ \vdots & \vdots & \vdots \\ \mathbb{E}(X_1 X_N) & \dots & \mathbb{E}(X_N^2) \end{pmatrix}$$

est inversible, les coefficients qui définissent le projeté de la variable aléatoire  $Y$  sur l'espace  $[X]_1^N$  sont définis par

$$(26) \quad \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_N \end{pmatrix} = \Sigma_{[X]_1^N}^{-1} \begin{pmatrix} \mathbb{E}(Y X_1) \\ \vdots \\ \mathbb{E}(Y X_N) \end{pmatrix}.$$

Avant de continuer, regardons un exemple facile.

**EXERCICE 4.** Reprendre le processus autorégressif de la section 2.3. A partir de la récurrence (18) et des propriétés (20) et (23), donner l'expression du projeté de  $X_n$  sur  $[X]_{n-2}^{n-1}$ , puis sur  $[X]_{n-3}^{n-1}$  et ainsi de suite.

Cet exercice généralisé au chapitre 7 car c'est précisément le projeté de  $X_n$  qui sera pris comme prédicteur de  $X_n$ .  $\triangle\triangle$

- *Propriétés de la projection.*

D'après la définition il est facile de comprendre que, si  $Y$  est elle-même combinaison linéaire des variables  $X_1, \dots, X_N$  sur lesquelles se fait la projection, alors  $Y$  est son propre projeté. En termes de prévision :  $Y$  est dans ce cas sa propre prévision.

D'après (25), on voit que, si  $Y$  est orthogonale à toutes les  $X_j$ , le projeté de  $Y$  est nul. En termes de prévision, on prévoit alors  $Y$  par zéro.

Enfin, l'expression (26) rappelle que l'opération qui consiste à projeter est linéaire : le projeté d'une combinaison linéaire  $a_1 Y_1 + \dots + a_m Y_m$  est la même combinaison linéaire des projetés  $a_1 P_{[X]_1^N}(Y_1) + \dots + a_m P_{[X]_1^N}(Y_m)$ .

### 3.2. Coefficient de corrélation partielle entre deux variables compte non-tenu de l'influence d'un groupe de variables.

On se donne  $N \geq 3$  variables aléatoires  $X_1, \dots, X_N$ , et on va définir un coefficient de corrélation entre  $X_1$  et  $X_N$  qui fasse abstraction de l'influence de  $X_2, \dots, X_{N-1}$ .

Cette notion est une réponse à la question suivante : il arrive que deux phénomènes soient fortement corrélés, mais que cette corrélation traduise non pas un fort lien entre ces deux phénomènes, mais plutôt l'influence commune d'un facteur extérieur. Parmi les exemples célèbres (et, il faut le dire, un peu caricaturaux) il y a celui-ci : on note une forte corrélation entre le nombre de pasteurs baptistes et le nombre d'arrestations pour ivresse sur la voie publique, aux États-Unis, à la fin du dix-neuvième siècle. Bien sûr cette corrélation n'est pas intéressante en elle-même, car l'augmentation simultanée de ces deux indices est simplement due à la forte augmentation de la population urbaine à cette époque (tiré de "Les quatre antilopes de l'apocalypse" de S.J. Gould). Pour évacuer ce genre d'influences non pertinentes, on a défini le coefficient de corrélation partielle. Dans ce qui suit,  $\text{corr}(U, V)$  désigne le coefficient de corrélation entre  $U$  et  $V$ , soit

$$\text{corr}(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var } U \text{ Var } V}}$$

**DÉFINITION 5.7.** *Le coefficient de corrélation partielle entre  $X_1$  et  $X_N$  abstraction faite de l'influence de  $X_2, \dots, X_{N-1}$  est le coefficient de corrélation entre les deux variables auxquelles on a retranché leurs meilleures explications en terme de  $X_2, \dots, X_{N-1}$ , soit*

$$r_{[X]_2^{N-1}}(X_1, X_N) = \text{corr} \left( X_1 - P_{[X]_2^{N-1}}(X_1), X_N - P_{[X]_2^{N-1}}(X_N) \right).$$

- **Propriété.** Il est facile de voir, à partir de la formule (26), que  $r_{[X]_2^{N-1}}(X_1, X_N)$  ne dépend que des espérances des produits  $X_i X_j$ .

Un exemple permet de comprendre la simplicité de la notion.

**EXERCICE 5.** Considérons trois variables indépendantes  $X, Z_1, Z_2$  centrées et de variance 1. On considère les variables

$$X_1 = X + Z_1 \quad \text{et} \quad X_2 = X + Z_2.$$

Calculer  $\rho(X_1, X_2)$ , puis  $r_{[X]}(X_1, X_2)$ , et commenter.  $\triangle\triangle$

### 3.3. Auto-corrélation partielle d'un processus stationnaire.

On considère maintenant un processus stationnaire centré ( $\dots, X_{-1}, X_0, X_1, \dots$ ).

3.3.1. *Définition, premieir calcul et exemple.* On définit la fonction d'auto-corrélation partielle  $r(h)$ , pour  $h \neq 0$ , de la façon suivante

DÉFINITION 5.8.

$$(27) \quad r(h) = r(-h) \quad \forall h \neq 0,$$

$$(28) \quad r(1) = \rho(1)$$

$$(29) \quad r(h) = r_{[X]_2^h}(X_1, X_{h+1}) \quad \forall h \geq 2.$$

La définition de  $r(1)$  est bien conforme à (29) puisque, si on fait  $h = 1$  dans cette formule, on constate qu'il n'y a pas de variable entre  $X_1$  et  $X_2$ . D'autre part, dans (29), le coefficient peut aussi bien se définir comme

$$r(h) = r_{[X]_{k+1}^{k+h-1}}(X_k, X_{k+h}) \quad \forall k \geq 1.$$

En effet, on a vu que la corrélation partielle ne dépend que des espérances des produits des variables en jeu (ici, de leurs covariances, puisque les espérances sont nulles). Or ces covariances ne changent pas si on translate les indices puisque le processus est stationnaire.

Retenant les expressions des projections (25) et (26) on obtient

$$r(h) = \text{corr}(X_n - \sum_{j=1}^{h-1} \lambda_j X_{n+j}, X_{n+h} - \sum_{j=1}^{h-1} \lambda_{h-j} X_{n+j})$$

avec

$$(30) \quad \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{h-1} \end{pmatrix} = \begin{pmatrix} \sigma(0) & \dots & \sigma(h-2) \\ \vdots & \ddots & \vdots \\ \sigma(h-2) & \dots & \sigma(0) \end{pmatrix}^{-1} \begin{pmatrix} \sigma(1) \\ \vdots \\ \sigma(h-1) \end{pmatrix}$$

**EXERCICE 6.** Reprendre le processus autorégressif d'ordre 1 de la section 2.3. Prendre  $h > 1$ . Quelle est la projection de  $X_{h+1}$  sur  $[X]_2^h$ ? (utiliser l'exercice 2). En déduire  $X_{h+1} - P_{[X]_2^h}(X_{h+1})$ , puis, sans plus de calculs, la valeur du coefficient  $r(h)$ .  $\triangle\triangle$

3.3.2. *Algorithme de Durbin-Watson.* Cette méthode itérative de calcul rapide de l'autocorrélation d'un processus stationnaire est basé sur la propriété suivante qui ne sera pas démontrée ici (elle se vérifie sans difficulté).

**PROPOSITION 5.2.** *Si le processus  $(X_n)_{n \in \mathbb{Z}}$  est stationnaire, pour tout  $h \geq 2$   $r(h)$  est le coefficient de  $X_n$  dans l'expression du projeté de  $X_{n+h}$  sur l'espace  $[X]_n^{n+h-1}$ .*

Partant de ce résultat l'algorithme se construit de la façon suivante.

$$P_{[X]_n^{n+h-1}}(X_{n+h}) = \sum_{j=0}^{h-1} \alpha_{h-j}^{(h)} X_j,$$

et, d'après la Proposition 5.2,  $r(h) = \alpha_h^{(h)}$ . Par ailleurs, par définition du projeté,

$$(31) \quad \begin{pmatrix} \alpha_1^{(h)} \\ \vdots \\ \alpha_h^{(h)} \end{pmatrix} = \begin{pmatrix} \sigma(0) & \dots & \sigma(h-1) \\ \vdots & \ddots & \vdots \\ \sigma(h-1) & \dots & \sigma(0) \end{pmatrix}^{-1} \begin{pmatrix} \sigma(1) \\ \vdots \\ \sigma(h) \end{pmatrix}$$

**EXERCICE 7.** Utiliser (31) pour trouver l'expression du coefficient d'autocorrélation partielle du processus

$$X_n = \varepsilon_n + b\varepsilon_{n-1}$$

où  $(\varepsilon_n)_{n \in \mathbb{Z}}$  est un bruit blanc centré.  $\triangle\triangle$

#### 4. Estimation de ces indices caractéristiques

Jusqu'ici, dans ce chapitre, on a travaillé sur des suites  $(\dots, X_{-1}, X_0, X_1, \dots)$  de variables aléatoires. Maintenant on va considérer des réalisations de trajectoires de longueur finie de telles suites. Autrement dit on considère un nombre fini d'indices,  $(1, \dots, n)$ , et des réalisations  $(x_1, x_2, \dots, x_n)$  de  $(X_1, \dots, X_n)$ . Entrent dans cette rubrique les séries obtenues par simulation. Par exemple on dispose de simulateurs de bruits blancs gaussiens. À partir d'eux on peut évidemment simuler des trajectoires de processus moyennes mobiles (voir 2.2). On peut aussi simuler des trajectoires de processus autorégressif d'ordre 1. Pour cela on simule une trajectoire de longueur voulue d'un bruit blanc, on choisit arbitrairement une valeur de  $x_0$ , puis on fait agir la récurrence (18). Voir les figures 5.2 à 5.8 où le graphique du haut représente la série obtenue par simulation.

On considérera aussi bien sûr les résidus de séries temporelles auxquelles on a enlevé les tendances et que l'on a désaisonnalisées.

Dans le premier cas (simulations) on va se demander comment estimer les paramètres  $\mu$ ,  $\sigma(h)$ ,  $\rho(h)$  et  $r(h)$  correspondant à chaque modèle.

Dans le second on va se demander comment des indices bien choisis, calculés sur la série des résidus, peuvent permettre d'ajuster un modèle de processus stationnaire sur cette série.

Les deux problèmes sont en fait les mêmes et les indices empiriques introduits au chapitre 4 sont de bons estimateurs des paramètres théoriques correspondants, du moins pour tous les modèles ARMA considérés au chapitre 6.

##### 4.1. Estimation de l'espérance, des auto-covariances et des auto-corrélations.

- On estime  $\mu$  par la moyenne empirique  $\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$ .
- On estime l'auto-covariance  $\sigma(h)$  par l'auto-covariance empirique

$$\hat{\sigma}_n(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_j - \bar{x}_n)(x_{j+h} - \bar{x}_n)$$

- finalement, on estime l'auto-corrélation  $\rho(h)$  par l'auto-corrélation empirique  $\hat{\rho}_n(h) = \frac{\hat{\sigma}_n(h)}{\hat{\sigma}_n(0)}$ .

On peut se demander ce qui permet de dire que ces indices empiriques sont de bons estimateurs des indices théoriques correspondants. La proposition suivante, qui va au delà des limites de ce cours, est éclairante.

**PROPOSITION 5.3.** *Si  $(\dots, X_{-1}, X_0, X_1, \dots)$  est un des processus ARMA définis au chapitre 6, les indices aléatoires*

$$\begin{aligned}\overline{X}_n &= \frac{1}{n} \sum_{j=1}^n X_j, \\ \hat{\sigma}_n(h) &= \frac{1}{n} \sum_{j=1}^{n-h} (X_j - \overline{X}_n)(X_{j+h} - \overline{X}_n), \\ \hat{\rho}_n(h) &= \frac{\hat{\sigma}_n(h)}{\hat{\sigma}_n(0)}\end{aligned}$$

convergent respectivement, lorsque la taille  $n$  de la trajectoire tend vers l'infini, vers les valeurs théoriques correspondantes (c'est-à-dire respectivement vers  $\mu = 0$ ,  $\sigma(h)$  et  $\rho(h)$ ). Cette convergence a lieu avec la probabilité 1.

Ceci signifie que, pour des longueurs de séries assez grandes, les indices empiriques approchent bien les indices théoriques. Le graphique 5.1 montre l'évolution de l'auto-corrélation empirique vers la théorique quand la longueur de la série croît.

#### 4.2. Estimation des auto-corrélations partielles.

On utilise l'algorithme de Durbin-Watson, qui permet, pour tout  $h \geq 2$ , pour calculer les auto-corrélations partielles estimées  $\hat{\rho}_n(h)$  à partir des auto-corrélations empiriques  $\hat{r}_n(1), \dots, \hat{r}_n(h)$ .

On montre que les bonnes propriétés données par la proposition 5.3 sont aussi valables pour les estimateurs ainsi obtenus.

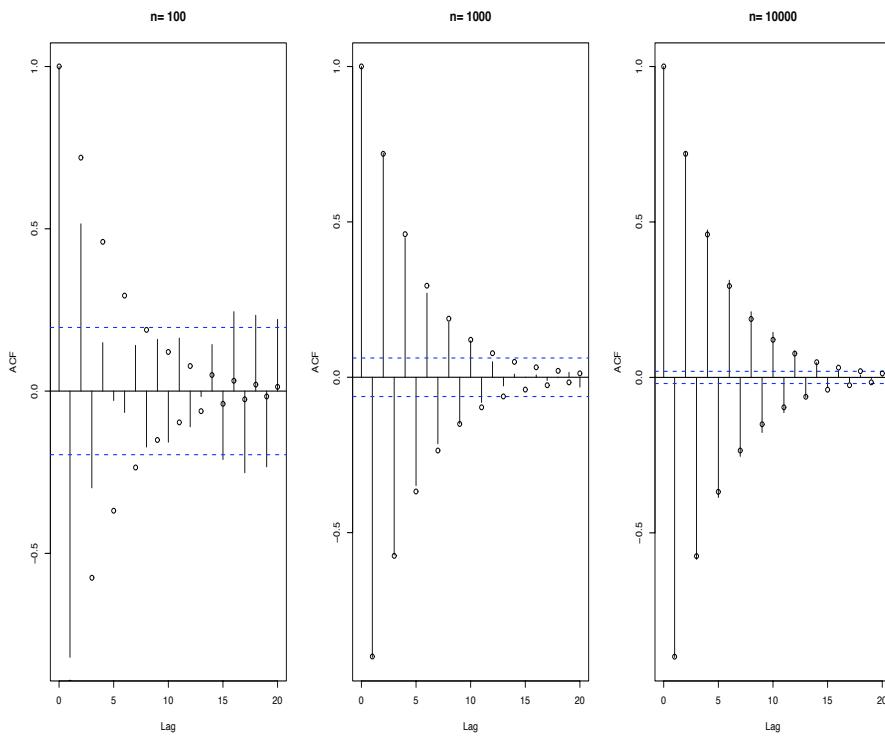


FIG. 5.1. Comparaison des covariances théoriques et empiriques pour le modèle ARMA(1,1)  $X_n + 0.8X_{n-1} = \epsilon_n - 0.8\epsilon_{n-1}$

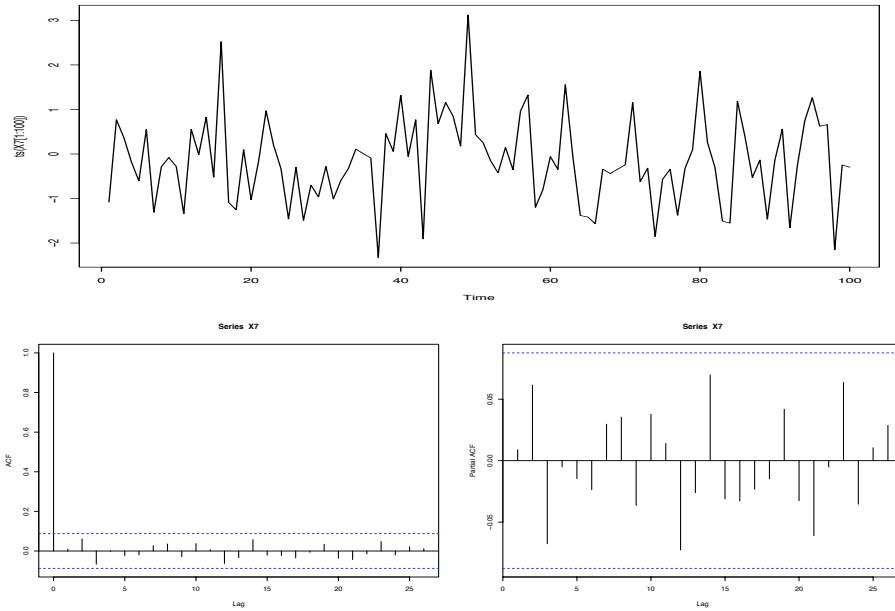
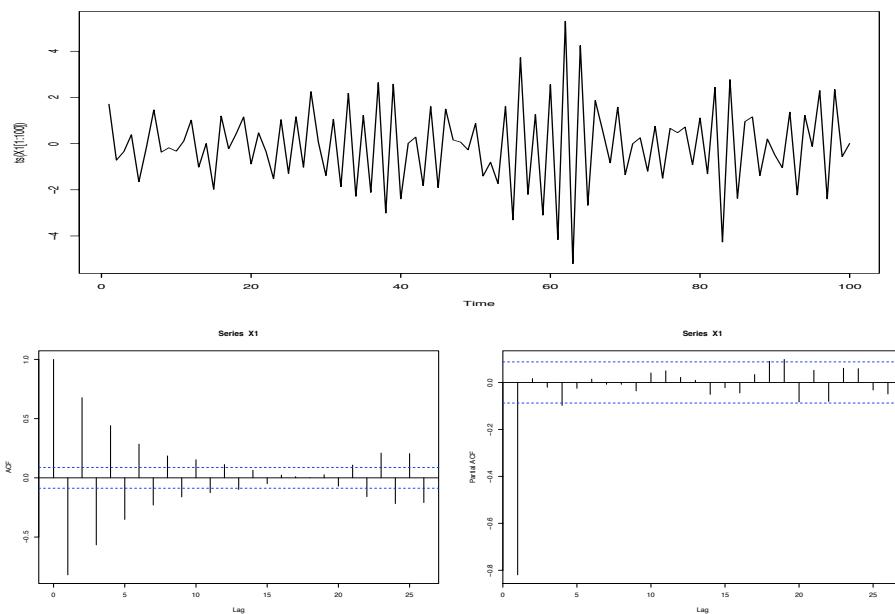
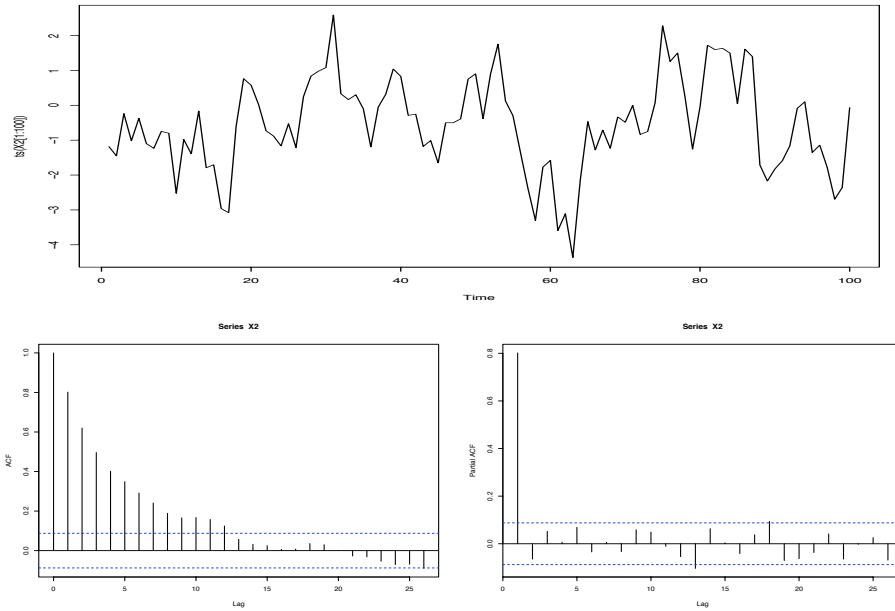
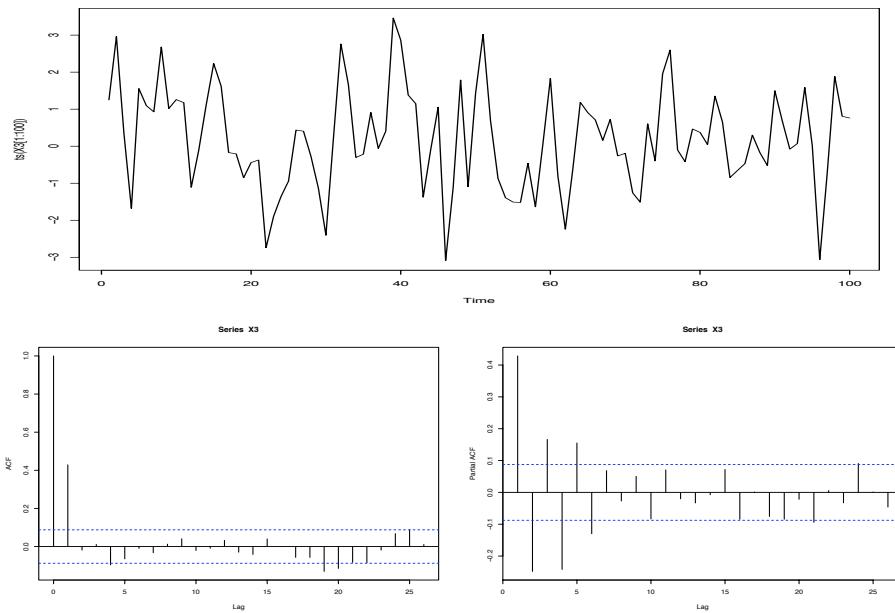
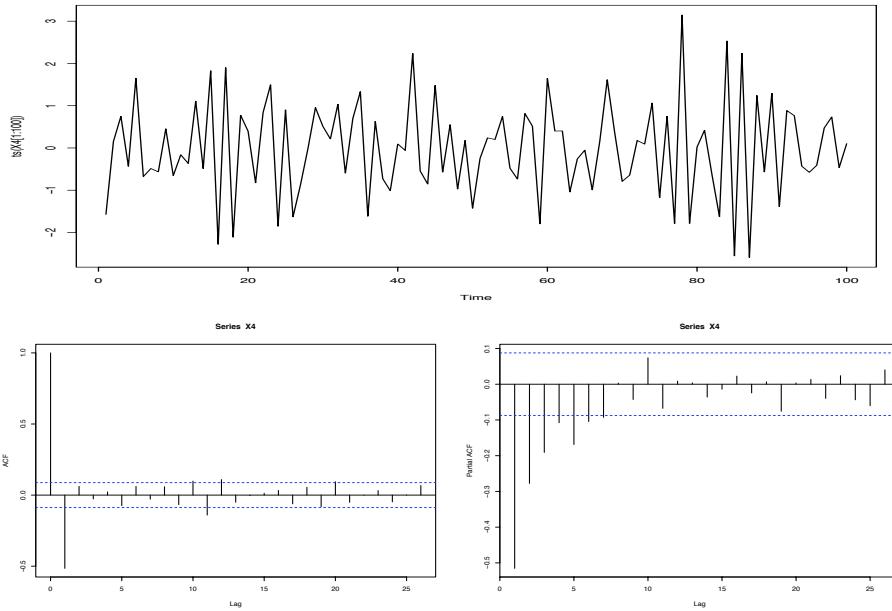
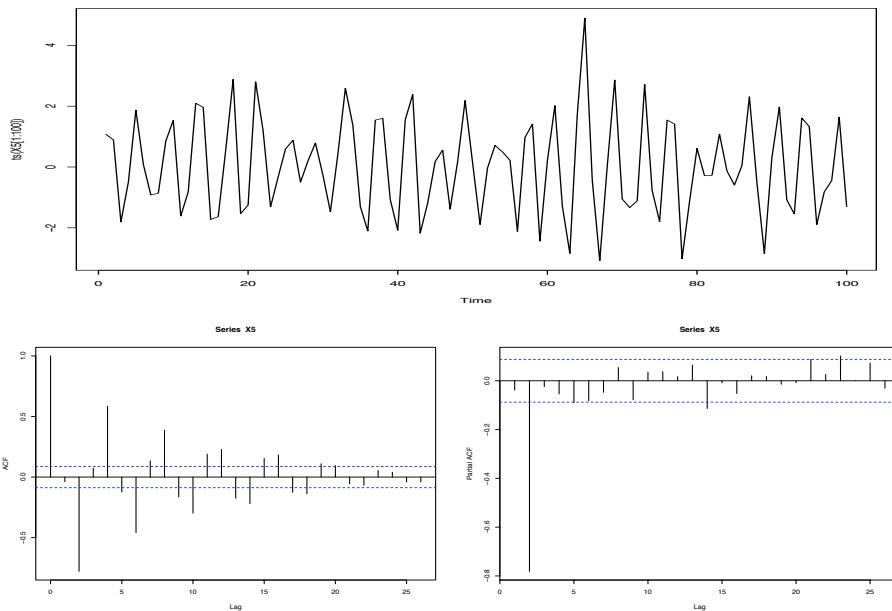


FIG. 5.2. bruit blanc gaussien : série, ACF et PACF

FIG. 5.3. modèle AR(1)  $a=0.8$  : série, ACF et PACF

FIG. 5.4. modèle AR(1)  $\alpha = -0.8$  : série, ACF et PACFFIG. 5.5. modèle MA(1)  $\alpha = 0.8$  : série , ACF et PACF

FIG. 5.6. modèle MA(1)  $a=-0.8$  : série, ACF et PACFFIG. 5.7. modèle AR(2)  $X_n + 0.81X_{n-2} = \epsilon_n$  : série, ACF et PACF

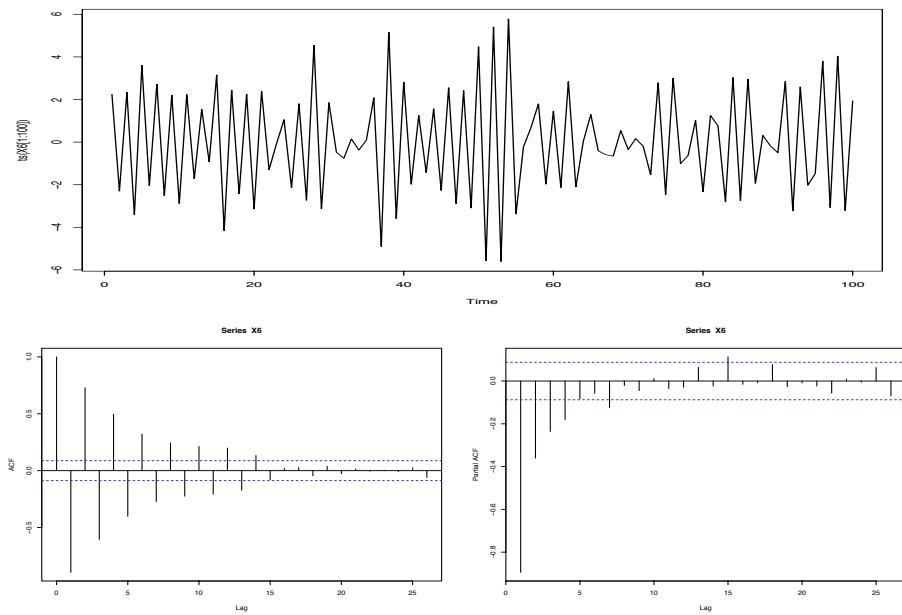


FIG. 5.8. modèle ARMA(1,1)  $X_n + 0.8X_{n-1} = \epsilon_n - 0.8\epsilon_{n-1}$  : série, ACF et PACF

## CHAPITRE 6

### Les processus ARMA

Il s'agit d'une très large famille de processus stationnaires qui présente deux avantages : d'une part ces processus sont d'excellents outils de prévision, et d'autre part on dispose de méthodes parfaitement au point pour estimer leurs paramètres.

#### **1. Les processus autorégressifs et leurs indices caractéristiques.**

Ces processus généralisent les processus autorégressifs d'ordre 1 introduits en 2.3 du chapitre 5. Reprenons cet exemple. En regardant les équations (18) et (19), et en se rappelant que la fonction  $(1 - az)^{-1}$  se développe au voisinage de zéro en

$$\frac{1}{1 - az} = 1 + az + a^2z^2 + \cdots + a^kz^k + \cdots,$$

on constate que l'expression (19) est tout simplement obtenue en remplaçant les  $z^k$  par les  $\varepsilon_{n-k}$  dans le développement. A partir de cette constatation, on obtient une méthode de construction de la famille des processus autorégressifs.

**1.1. Définition, polynôme caractéristique.** On constate d'abord qu'étant donné un polynôme  $A(z) = 1 + a_1z + \cdots + a_pz^p$  dont toutes les racines sont de module strictement supérieur à 1, la fonction  $A(z)^{-1}$  se développe au voisinage de zéro en

$$(32) \quad \frac{1}{A(z)} = \frac{1}{1 + a_1z + \cdots + a_pz^p} = 1 + \alpha_1z + \alpha_2z^2 + \cdots + \alpha_kz^k + \cdots,$$

où la série de terme général  $\alpha_k$  converge absolument (démontrez le).

**DÉFINITION 6.1.** Étant donné un polynôme  $A(z) = 1 + a_1z + \cdots + a_pz^p$  de degré exactement  $p$  (c'est à dire que  $a_p \neq 0$ ) dont toutes les racines sont de module strictement supérieur à 1, et  $(\varepsilon_n)_{n \in \mathbb{Z}}$  un bruit blanc centré dont la variance notée  $\sigma^2$  est supposée non nulle, le processus défini par

$$(33) \quad X_n = \varepsilon_n + \alpha_1\varepsilon_{n-1} + \cdots + \alpha_k\varepsilon_{n-k} + \cdots \quad \forall n$$

est appelé processus autorégressif d'ordre  $p$  (ou encore processus  $AR_p$ ) de polynôme caractéristique  $A$  et de bruit d'innovation  $(\varepsilon_n)$ .

Avant d'examiner les diverses propriétés de ces processus, faisons quelques remarques.

**REMARQUE 6.1.** La série à termes aléatoires qui définit  $X_n$  dans (33) converge à la fois presque sûrement et en moyenne quadratique. Démontrez le en utilisant les propriétés asymptotiques des coefficients  $\alpha_k$ .

**REMARQUE 6.2.** Le polynôme  $A$  se décompose en  $A(z) = (1 - z_1 z) \cdots (1 - z_p z)$ , où les  $z_j$  sont les inverses des racines de  $A$  et sont, de ce fait, de module strictement inférieur à 1 (et non nécessairement réelles, évidemment). Du développement de chaque facteur  $(1 - z_j z)^{-1} = 1 + z_j z + z_j^2 z^2 + \cdots$  on peut déduire celui de  $A(z)^{-1}$ , ce qui permet le calcul des coefficients  $\alpha_j$ , comme on le voit dans l'exercice qui suit.

**EXERCICE 8.** Expliciter l'expression du processus autorégressif de polynôme caractéristique  $A$  dans chacun des cas suivants :

- $A(z) = 1 - z + \frac{1}{2}z^2$ .
- $A(z) = 1 - \frac{5}{6}z + \frac{1}{6}z^2$ .
- $A(z) = 1 - \frac{1}{2}z^2$ . △△

### 1.2. Stationnarité.

De la représentation en série (33), on déduit que  $\mathbb{E}(X_n) = 0$  quel que soit  $n$ . Quant à l'auto-covariance, elle s'écrit

$$\begin{aligned} \text{Cov}(X_n, X_{n+h}) &= \text{Cov}(\varepsilon_n + \cdots + \alpha_k \varepsilon_{n-k} + \cdots, \varepsilon_{n+h} + \cdots + \alpha_k \varepsilon_{n+h-k} + \cdots) \\ &= \sigma^2(\alpha_h + \alpha_{h+1}\alpha_1 + \cdots + \alpha_{h+k}\alpha_k + \cdots), \end{aligned}$$

puisque  $\text{Cov}(\varepsilon_j, \varepsilon_k)$  est nul si  $j \neq k$  et est égal à  $\sigma^2$  si  $j = k$ .

L'expression obtenue est loin d'être explicite. Cependant elle montre à l'évidence que la covariance ne dépend que de  $h$ . Donc *le processus proposé en (33) est stationnaire*. Nous verrons plus loin une façon plus commode de calculer cette covariance.

### 1.3. Expression autorégressive.

De même que le processus autorégressif d'ordre 1 vérifie l'équation de récurrence (18), il est assez facile de voir que le processus défini par (33) vérifie l'équation de récurrence linéaire (à second membre aléatoire)

$$(34) \quad X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} = \varepsilon_n \quad \forall n.$$

En effet

$$\begin{aligned} X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} &= \varepsilon_n + \alpha_1 \varepsilon_{n-1} + \alpha_2 \varepsilon_{n-2} + \cdots \\ &\quad + a_1(\varepsilon_{n-1} + \alpha_1 \varepsilon_{n-2} + \alpha_2 \varepsilon_{n-3} + \cdots) \\ &\quad + a_2(\varepsilon_{n-2} + \alpha_1 \varepsilon_{n-3} + \alpha_2 \varepsilon_{n-4} + \cdots) \\ &\quad + \cdots \\ &\quad + a_p(\varepsilon_{n-p} + \alpha_1 \varepsilon_{n-p-1} + \alpha_2 \varepsilon_{n-p-2} + \cdots) \end{aligned}$$

Soit

$$\begin{aligned} X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} &= \varepsilon_n + \varepsilon_{n-1}(\alpha_1 + a_1) + \varepsilon_{n-2}(\alpha_2 + a_1\alpha_1 + a_2) \\ &\quad + \cdots \\ &\quad + \varepsilon_{n-p}(\alpha_p + \alpha_{p-1}a_1 + \alpha_{p-2}a_2 + \cdots + \cdots + a_p) \\ &\quad + \cdots \end{aligned}$$

Or, les coefficients  $\alpha_1 + a_1$ ,  $\alpha_2 + a_1\alpha_1 + a_2$ , etc... sont tous nuls, car les  $a_j$  et les  $\alpha_j$  sont reliés par la relation

$$(1 + a_1z + \cdots + a_pz^p)(1 + \alpha_1z + \alpha_2z^2 + \cdots) = 1 \quad \forall z,$$

qui nécessite que le coefficient de  $z$  dans le produit soit nul (soit  $\alpha_1 + a_1 = 0$ ), ainsi que le coefficient de  $z^2$  (soit  $\alpha_2 + a_1\alpha_1 + a_2 = 0$ ), et ainsi de suite.

La représentation (34), qui fait apparaître les coefficients du polynôme caractéristique, est bien sûr préférée à la représentation en série infinie (33). C'est elle qui justifie le nom de *processus autorégressif* : en effet (34) peut se lire formellement comme un modèle de régression où l'observation est  $X_n$ , où les variables sur lesquelles on régresse sont  $X_{n-1}, \dots, X_{n-p}$ , les coefficients étant  $-a_1, \dots, -a_p$ , et où le bruit résiduel est  $\varepsilon_n$ .

#### 1.4. L'innovation.

De la représentation en série (33) on déduit que, quelque soit  $n$  et pour tout  $h > 0$ ,

$$\text{Cov}(\varepsilon_{n+h}, X_n) = \mathbb{E}(\varepsilon_{n+h}X_n) = \mathbb{E}(\varepsilon_{n+h}\varepsilon_n) + \alpha_1 \mathbb{E}(\varepsilon_{n+h}\varepsilon_{n-1}) + \cdots = 0,$$

parce que  $\varepsilon_n$  est un bruit blanc.

**PROPOSITION 6.1.** *Dans le modèle AR<sub>p</sub> (33), le bruit d'innovation est tel que, pour tout  $n$ ,  $\varepsilon_n$  est orthogonal à toutes les variables  $X_j$  dont l'indice  $j$  est strictement inférieur à  $n$ .*

En réalité, de la définition (33) et de la représentation (34) on déduit en outre que les espaces de Hilbert engendrés respectivement par les  $(X_j)_{j \leq n}$  et les  $(\varepsilon_j)_{j \leq n}$  sont égaux. Ce que l'on peut énoncer de façon plus imagée : *à tout instant, le passé du processus et le passé de son bruit d'innovation sont les mêmes*, ou, avec les notations introduites au chapitre 5,

$$(35) \quad [X]_{-\infty}^n = [\varepsilon]_{-\infty}^n \quad \forall n.$$

La proposition 6.1 est une conséquence directe de ce résultat.

L'expression *bruit d'innovation* vient d'ailleurs de là : le bruit représente ce que  $X_{n+1}$  apporte de vraiment nouveau à ce qui s'est passé jusqu'à l'instant  $n$ . Comme dans l'exercice 2, cette propriété importante va permettre de trouver la fonction d'auto-covariance du processus.

#### 1.5. Auto-covariance des processus AR<sub>p</sub>.

On part de la récurrence (34). Pour tout  $h$  et pour tout  $n$ , on a

$$(36) \quad \text{Cov}(X_{n+h} + a_1X_{n+h-1} + \cdots + a_pX_{n+h-p}, X_n) = \text{Cov}(\varepsilon_{n+h}, X_n).$$

Si  $h$  est strictement positif, le second membre est nul, ce qui conduit à

**PROPOSITION 6.2.** *La fonction d'auto-covariance d'un processus AR<sub>p</sub> de polynôme caractéristique*

$$1 + a_1z + \cdots + a_pz^p$$

*satisfait à l'équation de récurrence linéaire sans second membre*

$$(37) \quad \sigma(h) + a_1\sigma(h-1) + \cdots + a_p\sigma(h-p) = 0 \quad \forall h > 0.$$

Évidemment, puisque l'auto-corrélation est proportionnelle à l'auto-covariance, cette proposition est aussi valable pour  $\rho(h)$ , c'est à dire que

$$(38) \quad \rho(h) + a_1\rho(h-1) + \cdots + a_p\rho(h-p) = 0 \quad \forall h > 0.$$

Supposons que l'on ait à déterminer la suite des auto-covariances ou des auto-corrélations. La récurrence ci-dessus implique que chacune de ces deux suites est connue dès que sont connues ses  $p$  premières valeurs  $\sigma(0), \dots, \sigma(p-1)$  ou  $\rho(0) = 1, \rho(1), \dots, \rho(p-1)$ . Par ailleurs, supposons que toutes les racines du polynôme caractéristique  $A$  (voir la Définition 6.1) sont simples (ce sont les  $\frac{1}{z_j}$ , cf. la Remarque 6.2). On sait que les suites  $u_n$  solutions de la récurrence

$$u_h + a_1u_{h-1} + \cdots + a_pu_{h-p} = 0 \quad \forall h > 0.$$

sont toutes de la forme

$$u_h = \lambda_1 z_1^h + \cdots + \lambda_p z_p^h.$$

Donc la suite des auto-covariances est aussi de cette forme, ainsi que celle des auto-corrélations. Pour chacune de ces suites, les coefficients  $\lambda_j$  se déterminent à partir des  $p$  valeurs initiales de la suite (remarquer que les calculs sont plus faciles pour l'auto-corrélation puisque  $\rho(0) = 1$  n'a pas à être déterminé).

Le fait important est que les deux suites  $\rho(h)$  et  $\sigma(h)$  tendent vers zéro lorsque  $h$  tend vers l'infini. De plus cette convergence est assez rapide puisqu'elle se fait à la vitesse  $|z_m|^h$ , (où  $|z_m|$  est le plus grand des modules des  $z_j$ ) c'est à dire à la vitesse d'une suite géométrique. Par exemple, dans l'exercice 10 ci-dessous, la vitesse est  $2^{-h/2}$ .

**EXERCICE 9.** On considère le processus autorégressif d'ordre deux, de représentation

$$X_n + a_1 X_{n-1} + a_2 X_{n-2} = \varepsilon_n \quad \forall n.$$

Montrer que

$$\rho(1) = \frac{-a_1}{1+a_2},$$

et que

$$\rho(2) = -a_1\rho(1) - a_2.$$

△△

**EXERCICE 10.** Quelle est la fonction d'auto-covariance du processus autorégressif de représentation

$$X_n - \frac{1}{2}X_{n-2} = \varepsilon_n.$$

△△

**EXERCICE 11.** On considère le processus autorégressif de représentation

$$X_n - \frac{5}{6}X_{n-1} + \frac{1}{6}X_{n-2} = \varepsilon_n.$$

Donner les valeurs de  $\sigma(0)$  et  $\sigma(1)$ . Trouver les racines du polynôme  $1 - \frac{5}{6}z + \frac{1}{6}z^2$  et en déduire l'expression de  $\sigma(h)$  pour  $h \geq 1$ . △△

**EXERCICE 12.** On considère le processus autorégressif de représentation

$$X_n - X_{n-1} + \frac{1}{2}X_{n-2} = \varepsilon_n.$$

Donner les expressions de  $\sigma(0)$  et  $\sigma(1)$ . Trouver les racines du polynôme  $1 - z + \frac{1}{2}z^2$  et en déduire que

$$\rho(h) = 2^{-h/2} \left( \cos\left(\frac{h\pi}{4}\right) + \frac{1}{3} \sin\left(\frac{h\pi}{4}\right) \right) \quad \forall h \geq 0.$$

△△

### 1.6. Auto-corrélation partielle d'un processus autorégressif.

Dans l'exercice 5 on a vu que, pour un processus  $AR_1$ , le coefficient  $r(h)$  est nul pour  $h \geq 2$ .

Plus généralement, considérons un processus  $AR_p$  de représentation (34). Soit  $h \geq 1$ . Le projeté de  $X_{p+h+1}$  sur l'espace  $[X]_2^{p+h}$  est  $-a_1X_{p+h} - a_2X_{p+h-1} - \cdots - a_pX_{h+1}$ , puisque cette combinaison linéaire est dans l'espace sur lequel se fait la projection et puisque la différence  $X_{p+h+1} - (-a_1X_{p+h} - a_2X_{p+h-1} - \cdots - a_pX_{h+1}) = \varepsilon_{p+h+1}$  est orthogonale aux variables  $X_2, \dots, X_{p+h}$ . En conséquence,

$$\begin{aligned} & \text{Cov} \left( X_{p+h+1} - P_{[X]_2^{p+h}}(X_{p+h+1}), X_1 - P_{[X]_2^{p+h}}(X_1) \right) \\ &= \text{Cov} \left( \varepsilon_{p+h+1}, X_1 - P_{[X]_2^{p+h}}(X_1) \right) = 0, \end{aligned}$$

ce qui prouve que le coefficient d'auto-corrélation partielle  $r(h)$  est nul pour  $h = p+1, p+2, \dots$

De plus, on déduit directement de la proposition 5.2 que  $r(p) = -a_p$ .

Pour résumer, on peut énoncer

**PROPOSITION 6.3.** *La fonction d'auto-corrélation partielle  $r(h)$  d'un processus autorégressif d'ordre  $p$  est nulle pour  $h > p$ . De plus, si la représentation du processus est  $X_n + a_1X_{n-1} + \cdots + a_pX_{n-p} = \varepsilon_n$ , on a  $r(p) = -a_p$ .*

**EXERCICE 13.** Les figures 5.3, 5.4 et 5.7 représentent, pour trois processus autorégressifs différents (deux  $AR_1$  et un  $AR_2$ ) une trajectoire simulée de longueur 100 (en fait ce sont les 100 premiers termes d'une trajectoire de longueur 500, sur laquelle sont calculés les indices empiriques), l'auto-corrélation empirique et l'auto-corrélation partielle empirique. Commentez soigneusement ces figures et retrouvez (approximativement !) les valeurs des paramètres des processus qui ont servi de modèles aux simulations. △△

## 2. Les processus moyennes mobiles et leurs indices caractéristiques

**DÉFINITION 6.2.** *Étant donné un polynôme  $B(z) = 1 + b_1z + \cdots + b_qz^q$  de degré exactement  $q$  ( $b_q \neq 0$ ) et dont toutes les racines sont de module strictement supérieur à 1, et étant donné un bruit blanc centré  $(\varepsilon_n)_{n=\pm 1, \dots}$ , on appelle processus moyenne mobile d'ordre  $q$  (on*

*dit aussi  $MA_q$ , "MA" correspondant à "moving average") de polynôme caractéristique  $B$  et d'innovation  $\varepsilon_n$ , le processus défini par*

$$(39) \quad X_n = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q}, \quad \forall n.$$

Quelle que soit la position des racines de  $B$ , le processus (39) est stationnaire, centré et on a

$$\mathbb{E}(X_n \varepsilon_{n+h}) = 0 \quad \forall h > 0.$$

En réalité, l'hypothèse faite sur les racines du polynôme permet de montrer (on ne le fera pas ici) que

$$(40) \quad \varepsilon_n = X_n + \beta_1 X_{n-1} + \dots \quad \forall n$$

où les  $\beta_j$  sont les coefficients du développement en série de  $B(z)^{-1}$  autour de zéro. Ceci montre que, comme pour les processus autorégressifs, on a égalité des passés (35) ce qui justifie l'utilisation du terme *innovation* pour le bruit blanc.

De plus, (voir section 2.2 du chapitre 5)

$$\sigma(h) = 0 \quad \forall h > q.$$

On voit donc que l'auto-corrélation d'un processus moyenne mobile a le même type de comportement (annulation systématique au dessus d'un certain écart  $h$ ) que celui de l'auto-corrélation partielle d'un processus autorégressif.

L'auto-corrélation partielle d'un processus moyenne mobile est compliquée à calculer, on ne le fera pas ici. Par exemple le processus moyenne mobile d'ordre 1

$$X_n = \varepsilon_n + b \varepsilon_{n-1}$$

a pour fonction d'auto-corrélation partielle

$$r(h) = \frac{(-b)^h (1 - b^2)}{1 - b^{2(h+1)}} \quad \forall h \geq 1.$$

Comme  $-1 < \frac{-b}{1+b^2} < 1$ , on voit que cette suite tend vers zéro à vitesse exponentielle lorsque  $h \rightarrow \infty$  (en oscillant ou pas selon le signe de  $b$ ), comme le fait  $\rho(h)$  pour un processus autorégressif.

On admettra que ceci est toujours vrai, c'est à dire que l'auto-corrélation partielle d'un processus moyenne mobile a le même type de comportement que celui de l'auto-corrélation d'un processus autorégressif (décroissance exponentiellement rapide de la valeur absolue avec ou sans oscillations).

Les graphiques 5.5 et 5.6 représentent les 100 premières valeurs de deux séries de longueur 500 simulées à partir de modèles de moyennes mobiles, ainsi que les auto-corrélations et les auto-corrélations partielles correspondantes.

## 2.1. Un peu de reconnaissance de formes.

En se basant sur ce qu'on sait des comportements comparés des auto-corrélations et des auto-corrélations partielles des *AR* et des *MA*, il arrive qu'on puisse se faire une idée de la modélisation adéquate aux données simplement en regardant les auto-corrélations empiriques et les auto-corrélations partielles empiriques. Les graphiques 6.1 montrent ces

deux indices empiriques pour 5 séries simulées. Arrivez vous à dire quel type de modèle convient à chaque série ?

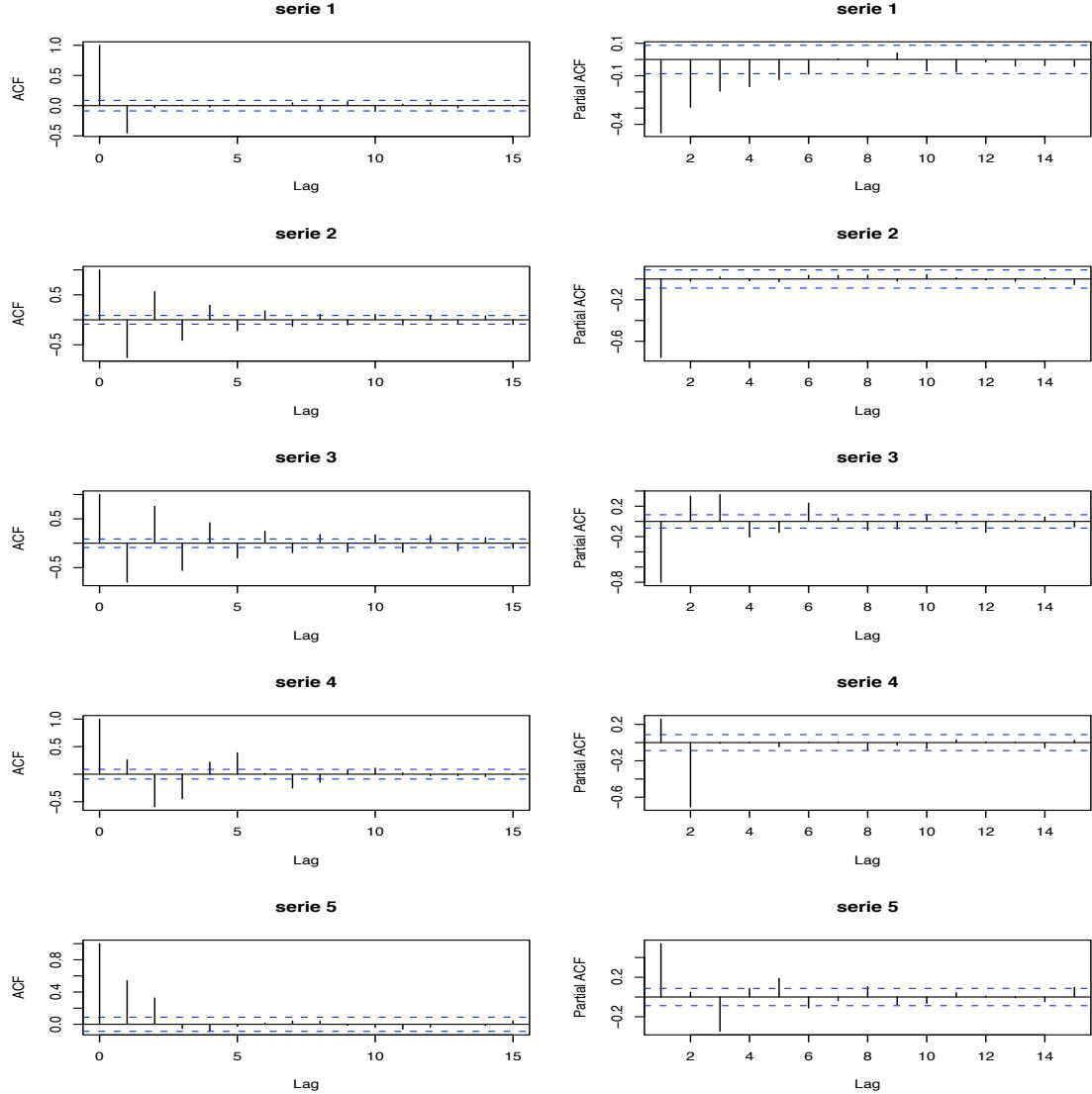


FIG. 6.1. Auto-corrélation et auto-corrélation partielle de quelques *ARMA*

### 3. Les ARMA mixtes

Ce sont des processus qui se représentent sous la forme d'une récurrence auto-régressive à second membre de type moyenne mobile (d'où le sigle *ARMA*)

$$(41) \quad X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q}, \quad \forall n.$$

où  $(\varepsilon_n)_{n=0,\pm 1,\dots}$  est un bruit blanc centré. Il y a donc deux polynômes caractéristiques  $A(z) = 1 + a_1z + \dots + a_pz^p$  et  $B(z) = 1 + b_1z + \dots + b_qz^q$ . On suppose  $pq \neq 0$  et on dit qu'on a affaire à un processus  $ARMA_{p,q}$ . Les racines de  $A$  et de  $B$  sont toutes de module strictement supérieur à 1 et ces deux polynômes n'ont pas de racine commune. L'hypothèse sur les racines de  $A$  permet de développer  $BA^{-1}$  en série

$$\frac{B(z)}{A(z)} = 1 + \gamma_1 z + \gamma_2 z^2 + \dots,$$

et de donner du processus une représentation explicite en forme de moyenne mobile de longueur infinie

$$(42) \quad X_n = \varepsilon_n + \gamma_1 \varepsilon_{n-1} + \gamma_2 \varepsilon_{n-2} + \dots,$$

qui permet de voir que le processus est centré, stationnaire, et que, pour tout  $n$ ,  $X_n$  est orthogonal aux  $\varepsilon_j$  d'indice strictement supérieur à  $n$ . L'hypothèse sur les racines de  $B$  permet de développer  $AB^{-1}$  en série

$$\frac{A(z)}{B(z)} = 1 + \delta_1 z + \delta_2 z^2 + \dots,$$

et de donner du processus une représentation explicite en forme d'une autorégression de longueur infinie

$$(43) \quad \varepsilon_n = X_n + \delta_1 X_{n-1} + \delta_2 X_{n-2} + \dots.$$

Ainsi, on a encore l'égalité des espaces (35) et  $(\varepsilon_n)_{n=0,\pm 1,\dots}$  est le bruit d'innovation du processus.

Ces propriétés sont donc communes à tous les processus  $ARMA$ .

Demander que  $A$  et  $B$  n'aient pas de racine commune revient à s'assurer qu'il n'y a pas une représentation plus courte. Par exemple le processus qui vérifie

$$X_n + aX_{n-1} = \varepsilon_n + a\varepsilon_{n-1}$$

vérifie aussi  $X_n = \varepsilon_n$ . C'est donc un bruit blanc, et utiliser la récurrence  $ARMA_{1,1}$  serait très maladroit car cela demanderait l'estimation d'un paramètre  $a$  qui en fait n'existe pas.

*Question de vocabulaire.* Les processus  $ARMA$  sont ceux qui vérifient la récurrence (41), avec les conditions restrictives qui l'accompagnent. Parmi ces processus il y a les *AR*, ceux pour lesquels les  $b_j$  sont nuls, et les *MA*, ceux pour lesquels les  $a_j$  sont nuls. Pour les distinguer de ces deux cas extrêmes simples, les autres seront appelés par la suite *ARMA mixtes*.

### 3.1. Auto-covariance, auto-corrélation et auto-corrélation partielle des processus $ARMA$ mixtes.

En reprenant (41), on trouve que

$$\begin{aligned} \text{Cov}(X_{n+h} + \sum_{j=1}^p a_j X_{n+h-j}, X_n) &= \sigma(h) + \sum_{j=1}^p a_j \sigma(h-j) \\ &= \text{Cov}(\varepsilon_{n+h} + \sum_{j=1}^q b_j \varepsilon_{n+h-j}, X_n), \end{aligned}$$

qui est nul dès que  $h$  dépasse  $q$ . Donc on peut énoncer

$$(44) \quad \sigma(h) + a_1 \sigma(h-1) + \cdots + a_p \sigma(h-p) = 0 \quad \text{si } h > q.$$

Ceci signifie que l'auto-covariance (et l'auto-corrélation) d'un processus *ARMA* mixte évolue selon la même récurrence linéaire que celle du processus autorégressif dont la représentation s'obtient en annulant les  $b_j$ . La différence est que la récurrence est valable dès que  $h = 1$  pour l'autorégressif alors qu'il faut attendre  $h = q + 1$  pour l' $\text{ARMA}_{p,q}$ . Mais à cette différence près, les deux auto-covariances se comportent de façon analogue si  $h$  est grand.

### 3.2. Les $\text{ARMA}_{1,1}$ .

On se contentera ici d'étudier en détail le cas le plus simple  $p = q = 1$ .

**EXERCICE 14.** On étudie ici les processus  $\text{ARMA}_{1,1}$

$$X_n + aX_{n-1} = \varepsilon_n + b\varepsilon_{n-1}$$

où  $|a| < 1$ ,  $|b| < 1$ , et où  $(\varepsilon_n)$  est un bruit blanc.

- Écrire le développement en moyenne mobile infinie de ces processus.
- Calculer  $\sigma(0), \sigma(1), \dots$  à partir de l'expression précédente.
- Directement, sans utiliser l'expression en moyenne mobile infinie, calculer  $\sigma(0)$ , montrer que  $\sigma(1) + a\sigma(0) = b\sigma^2$ , et que

$$\sigma(h) + a\sigma(h-1) = 0 \quad \forall h > 1.$$

En déduire de nouveau les valeurs des covariances.

- Donner l'expression de  $\rho(1)$ .
- La figure 5.8 montre les 100 premières valeurs d'une trajectoire de longueur 500 obtenue par simulation d'un processus  $\text{ARMA}_{1,1}$ , ainsi que les auto-corrélations et auto-corrélations partielles empiriques. Comment pourriez vous reconstituer la valeur du paramètre  $a$  à partir de ces graphiques ?

△△

## CHAPITRE 7

### Comment prévoir dans un processus ARMA ?

#### 1. Principe général

On cherche à prévoir la valeur que va prendre la variable aléatoire  $X_{n+h}$ , et pour cela on s'appuie sur les variables  $X_n, X_{n-1}, \dots$  que l'on suppose avoir été observées. On dit que  $h$  est l'horizon de la prévision. Comme au chapitre 4, on note  $\hat{X}_{n,h}$  le prédicteur que l'on va choisir. Ce prédicteur est donc, ici, une variable aléatoire fonction des  $X_j$  d'indices au plus  $n$ . En fait, dans le cadre de la modélisation *ARMA* on choisit de prendre pour prédicteur une combinaison linéaire de ces  $X_j$ . Autrement dit

$$\hat{X}_{n,h} = c_{1,h}X_1 + \dots + c_{n,h}X_n.$$

Dans le cadre choisi il est bien naturel de prendre celle de ces combinaisons linéaires qui est "la plus proche" de la variable aléatoire  $X_{n+h}$ . La projection de  $X_{n+h}$  sur l'espace engendré  $[X]_1^n$  est toute indiquée car c'est la combinaison linéaire qui minimise

$$\mathbb{E}(c_1X_1 + \dots + c_nX_n - X_{n+h})^2.$$

On prendra donc

$$\hat{X}_{n,h} = P_{[X]_1^n}(X_{n+h}),$$

et le reste du chapitre est consacré aux méthodes de calcul (ou de calcul approché) de cette projection lorsque le processus  $(X_n)_{n \in \mathbb{Z}}$  est un *ARMA*. Le calcul est très simple pour un processus autorégressif, un peu plus compliqué pour les moyennes mobiles et les *ARMA* mixtes (dans ce cas, on fera un calcul approché).

#### 2. Prévision à horizon 1

##### 2.1. Prévision dans un processus autorégressif.

Comme  $\varepsilon_{n+1}$  est orthogonal aux  $X_j$  pour  $j = n, n-1, \dots$ , et comme

$$X_{n+1} = -a_1X_n - \dots - a_pX_{n+1-p} + \varepsilon_{n+1},$$

on voit directement que

$$\hat{X}_{n,h} = P_{[X]_1^n}(X_{n+1}) = -a_1X_n - \dots - a_pX_{n+1-p}.$$

De ce fait, l'erreur de prévision, c'est à dire la différence entre  $X_{n+1}$  et son prédicteur, est

$$X_{n+1} - \hat{X}_{n,1} = X_{n+1} + a_1 X_n + \cdots + a_p X_{n+1-p} - (\hat{X}_{n,1} + a_1 \hat{X}_{n,1} + \cdots + a_p \hat{X}_{n,1}) = \varepsilon_{n+1}.$$

Donc le bruit d'innovation est aussi l'erreur de prévision à horizon 1.

On déduit de ce qui précède que cette erreur est une variable aléatoire centrée et de variance  $\sigma^2$ . Si on suppose en plus que le bruit  $\varepsilon_n$  est gaussien, on sera capable de donner un intervalle de confiance pour la prévision (cette question sera abordée au paragraphe 5 de ce chapitre).

## 2.2. Prévision dans les processus moyenne mobile et dans les processus mixtes.

Pour ces processus, il est plus facile de projeter non pas sur  $[X]_1^n$  mais sur  $[X]_{-\infty}^n$ , espace engendré par *tous* les  $X_j$  d'indice au plus  $n$ . Cet espace est en quelque sorte le *passé infini linéaire* du processus. Projeter sur cet espace est évidemment peu réaliste sur le plan pratique car  $X_0, X_{-1}, \dots$  ne sont pas accessibles à l'observateur, ce qui rend incalculable toute expression qui les utilise. On reviendra sur cette question plus loin.

### 2.2.1. Préliminaires.

On va d'abord montrer que les deux espaces  $[X]_{-\infty}^n$  et  $[\varepsilon]_{-\infty}^n$  sont égaux. Tout d'abord, pour un processus moyenne mobile, l'inclusion  $[X]_{-\infty}^n \subset [\varepsilon]_{-\infty}^n$  vient de la définition elle-même. Pour un processus autorégressif ou pour un *ARMA* mixte, elle découle des développements (33) et (43). Dans l'autre sens, l'inclusion  $[\varepsilon]_{-\infty}^n \subset [X]_{-\infty}^n$  vient, pour un *AR*, de son expression autorégressive. Dans les autres cas on utilise la proposition suivante, qui sera admise, et dont la preuve repose sur le fait que les racines de  $B$  sont de module strictement plus grand que 1.

**PROPOSITION 7.1.** *Les processus moyenne mobile et les processus mixtes admettent une expression de type autorégressif de longueur infinie*

$$(45) \quad \varepsilon_n = X_n + \beta_1 X_{n-1} + \beta_2 X_{n-2} + \cdots,$$

dont les coefficients  $\beta_j$  sont ceux du développement

$$\frac{A(z)}{B(z)} = \frac{1 + a_1 z + \cdots + a_p z^p}{1 + b_1 z + \cdots + b_q z^q} = 1 + \beta_1 z + \beta_2 z^2 + \cdots$$

Ici encore, la convergence de la série de terme général  $\beta_j X_{n-j}$  sort des limites du cours.

**EXERCICE 15.** Déterminer le développement autorégressif infini d'un processus *MA*<sub>1</sub> et d'un processus *ARMA*<sub>1,1</sub>.  $\triangle\triangle$

### 2.2.2. Prévision et erreur de prévision.

On peut maintenant résumer ce qui vient d'être dit et en tirer les conséquences. Pour simplifier l'écriture, dans la proposition qui suit, on note  $X_n = \varepsilon_n + \alpha_1 \varepsilon_{n-1} + \alpha_2 \varepsilon_{n-2} + \cdots$  les représentations en moyenne mobile (finie, si le processus est un *MA* <sub>$q$</sub> , infinies dans les autres cas) des processus *ARMA*.

## PROPOSITION 7.2.

1) Les passés linéaires d'un processus ARMA et de son bruit d'innovation sont égaux :

$$[X]_{-\infty}^n = [\varepsilon]_{-\infty}^n \quad \forall n,$$

2) Pour tout  $n$ , le prédicteur à horizon 1,  $\hat{X}_{n,1}$  est

$$(46) \quad \hat{X}_{n,1} = P_{[X]_{-\infty}^n}(X_{n+1}) = P_{[\varepsilon]_{-\infty}^n}(X_{n+1}) = \alpha_1 \varepsilon_n + \alpha_2 \varepsilon_{n-1} + \dots$$

3) l'erreur de prévision à horizon 1 est

$$X_{n+1} - \hat{X}_{n,1} = X_{n+1} - \alpha_1 \varepsilon_n - \alpha_2 \varepsilon_{n-1} - \dots = \varepsilon_{n+1}.$$

4) l'expression du prédicteur en fonction des  $X_j$  s'obtient en remplaçant, dans (46), les  $\varepsilon_j$  par leur développement (45).

DÉMONSTRATION. Le 1) a déjà été prouvé, le 2) vient du fait que la différence  $X_{n+1} - (-\alpha_1 \varepsilon_n - \alpha_2 \varepsilon_{n-1} - \dots)$ , qui est  $\varepsilon_{n+1}$  (voir le 3)) est orthogonale à  $[X]_{-\infty}^n$ .  $\square$

Cette proposition établit entre autres que l'erreur de prévision à horizon 1 est le bruit d'innovation. C'est le résultat qui avait déjà été obtenu pour la prévision dans un processus autorégressif.

Pour bien comprendre cette proposition, le mieux est de la tester sur des exemples simples.

EXERCICE 16. Exprimer  $\hat{X}_{n,1}$  pour un processus  $MA_1$  et pour un  $ARMA_{1,1}$   $\triangle\triangle$

Comme il a été dit plus haut, pour les processus mixtes ou pour les moyennes mobiles, on obtient une expression du prédicteur qui dépend de *tous* les  $X_j$  d'indices au plus  $n$ . De ce fait ce prédicteur n'est pas calculable par le statisticien puisque celui-ci ne connaît que les valeurs de  $X_n, \dots, X_1$ . La solution adoptée consiste à poser  $X_0 = X_{-1} = \dots = 0$  dans l'expression fournie par la proposition. On obtient ainsi un prédicteur approché,  $\tilde{X}_{n,1}$ , qui n'est ni  $P_{[X]_{-\infty}^n}(X_{n+1})$  ni  $P_{[X]_1^n}(X_{n+1})$ , mais qui n'est pas loin de  $P_{[X]_{-\infty}^n}(X_{n+1})$  si  $n$  est grand.

EXERCICE 17. Faire le calcul de  $\tilde{X}_{n,1}$  pour un processus  $MA_1$  et pour un processus  $ARMA_{1,1}$ . Montrer que la différence  $\tilde{X}_{n,1} - \hat{X}_{n,1}$  entre ce prédicteur approché et celui trouvé dans l'exercice précédent tend vers zéro quand  $n \rightarrow \infty$ .  $\triangle\triangle$

### 3. Prévision à horizon supérieur à 1

La situation est un peu particulière pour les processus moyennes mobiles. Nous commençons par eux.

**3.1. Les processus moyenne mobile.** Pour ceux-là, on voit tout de suite, à partir de la représentation

$$X_{n+h} = \varepsilon_{n+h} + b_1 \varepsilon_{n+h-1} + \dots + b_q \varepsilon_{n+h-q}, \quad \forall n,$$

que le prédicteur est nul dès que l'horizon  $h$  dépasse  $q$ . En effet, la différence

$$X_{n+h} - 0 = \varepsilon_{n+h} + b_1 \varepsilon_{n+h-1} + \dots + b_q \varepsilon_{n+h-q}$$

est orthogonale à l'espace  $[X]_{-\infty}^n = [\varepsilon]_{-\infty}^n$  puisque  $n + h - q > n$ . L'erreur est alors

$$X_{n+h} - \hat{X}_{n,h} = X_{n+h},$$

et la variance de l'erreur est  $\sigma(0)$ .

### 3.2. Les processus autorégressifs et les ARMA mixtes.

La projection sur un espace donné est, comme on l'a vu, une opération linéaire. Donc, si  $X_n$  est un processus ARMA de représentation (41), on a

$$(47) \quad \begin{aligned} & P_{[X]_{-\infty}^n}(X_{n+h}) + a_1 P_{[X]_{-\infty}^n}(X_{n+h-1}) + \cdots + a_p P_{[X]_{-\infty}^n}(X_{n+h-p}) \\ &= P_{[X]_{-\infty}^n}(\varepsilon_{n+h}) + \cdots + b_q P_{[X]_{-\infty}^n}(\varepsilon_{n+h-q}). \end{aligned}$$

Mais on sait que  $P_{[X]_{-\infty}^n}(Y) = Y$  si  $Y \in [X]_1^n$  et que  $P_{[X]_{-\infty}^n}(Y) = 0$  si  $Y$  est orthogonale à  $[X]_{-\infty}^n$  c'est à dire si  $\mathbb{E}(YX_n) = \mathbb{E}(YX_{n-1}) = \cdots = 0$ . Il en résulte que, dans (47),

$$\begin{aligned} P_{[X]_{-\infty}^n}(X_{n+h-l}) &= X_{n+h-l} \quad \text{et} \quad P_{[X]_{-\infty}^n}(\varepsilon_{n+h-l}) = \varepsilon_{n+h-l} \quad \text{si} \quad l \geq h, \\ &\quad \text{tandis que} \quad P_{[X]_{-\infty}^n}(\varepsilon_{n+h-l}) = 0 \quad \text{si} \quad h > l. \end{aligned}$$

Notamment, si  $h > q$ , le second membre de (47) est nul et on a

$$\hat{X}_{n,h} + a_1 \hat{X}_{n,h-1} + \cdots + a_p \hat{X}_{n,h-p} = 0.$$

Tout d'abord, cela donne la possibilité de calculer la prévision à horizon  $h$  lorsqu'on a calculé toutes les prévisions à horizon plus petit.

Ensuite, on note que la prévision  $\hat{X}_{n,h}$  évolue ( $n$  étant fixé,  $h$  croissant) selon la même récurrence que la fonction d'auto-covariance. Donc, comme elle, elle tend vers zéro à vitesse exponentielle quand  $h \rightarrow \infty$ . Il est donc inutile de prévoir à horizon trop grand.

**EXERCICE 18.** Faites les calculs des prévisions à horizon 2 et 3 pour un processus  $AR_1$ , un processus  $AR_2$ , un processus  $MA_1$ , et un processus mixte  $ARMA_{1,1}$ .  $\triangle\triangle$

## 4. Comportement de l'erreur de prévision quand l'horizon croît

Le problème abordé est le suivant : les variables observées étant celles d'indice au plus  $n$ , comment se comporte l'erreur de prévision à horizon  $h$  quand  $h$  croît.

### 4.1. Évolution de l'erreur.

On reprend l'expression en moyenne mobile infinie, écrite au rang  $n + h$

$$X_{n+h} = \varepsilon_{n+h} + \alpha_1 \varepsilon_{n+h-1} + \cdots + \alpha_h \varepsilon_n + \cdots$$

On a vu que le prédicteur est

$$\hat{X}_{n,h} = \alpha_h \varepsilon_n + \alpha_{h+1} \varepsilon_{n-1} + \cdots$$

Il s'ensuit que l'erreur de prévision est

$$X_{n+h} - \hat{X}_{n,h} = \varepsilon_{n+h} + \alpha_1 \varepsilon_{n+h-1} + \cdots + \alpha_{h-1} \varepsilon_{n+1},$$

variable aléatoire centrée et de variance  $\text{Var}_h = \sigma^2(1 + \alpha_1^2 + \cdots + \alpha_{h-1}^2)$ .

#### 4.2. Évolution de la variance d'erreur pour les AR ou les ARMA mixtes.

On voit clairement que la variance d'erreur de prévision croît avec  $h$ . La valeur  $\sigma^2$  obtenue pour  $h = 1$  est la plus petite. Lorsque  $h \rightarrow \infty$ , la variance d'erreur tend vers

$$\text{Var}_\infty = \sigma^2 \lim_{h \rightarrow \infty} (1 + \alpha_1^2 + \cdots + \alpha_{h-1}^2) = \sigma^2 (1 + \sum_{j=1}^{\infty} \alpha_j^2) = \sigma(0),$$

ce qui est bien normal puisque, étant donné que le prédicteur tend vers zéro quand l'horizon tend vers l'infini, l'erreur de prévision finit par ressembler à  $X_{n+h}$  (pour un processus moyenne mobile, cette valeur est atteinte dès que l'horizon dépasse l'ordre).

Pour résumer

**PROPOSITION 7.3.** *La variance d'erreur de prévision à horizon  $h$  dans un processus ARMA croît depuis la variance du bruit d'innovation (valeur prise pour  $h = 1$ ) jusqu'à la variance du processus lui-même.*

### 5. Intervalles de confiance

Supposons maintenant que le bruit est gaussien. Autrement dit les  $\varepsilon_j$  sont des variables aléatoires gaussiennes, centrées, indépendantes et de variance  $\sigma^2$ . Il en résulte que toutes les variables considérées ici (ce sont des combinaisons linéaires finies ou infinies de ces  $\varepsilon_j$ ) sont aussi gaussiennes. Notamment l'erreur de prévision

$$X_{n+h} - \hat{X}_{n,h} = \varepsilon_{n+h} + \alpha_1 \varepsilon_{n+h-1} + \cdots + \alpha_{h-1} \varepsilon_{n+1}$$

est gaussienne, centrée, et sa variance est, comme on l'a vu  $\text{Var}_h = \sigma^2 (1 + \alpha_1^2 + \cdots + \alpha_{h-1}^2)$ .

Il en résulte que l'intervalle aléatoire

$$[\hat{X}_{n,h} - z_{\alpha/2} \sigma \sqrt{1 + \alpha_1^2 + \cdots + \alpha_{h-1}^2}, \hat{X}_{n,h} + z_{\alpha/2} \sigma \sqrt{1 + \alpha_1^2 + \cdots + \alpha_{h-1}^2}]$$

a une probabilité  $1 - \alpha$  de contenir  $X_{n+h}$ . Cet intervalle est donc un intervalle de confiance de niveau  $1 - \alpha$  pour prévoir  $X_{n+h}$  lorsqu'on a observé les  $X_j$  d'indice au plus  $n$ . Et on remarque, comme on s'y attendait, que la largeur de cet intervalle croît avec  $h$ .

### 6. Mise à jour

Le problème est le suivant. On a, en se basant sur les observations antérieures à  $n$ , prévu  $X_{n+1}, X_{n+2}, \dots, X_{n+h}$ . On observe maintenant  $X_{n+1}$ . Non seulement on n'a plus à le prévoir, mais encore on connaît maintenant la valeur de l'erreur de prévision  $X_{n+1} - \hat{X}_{n,1}$ . De plus, on va pouvoir modifier les prévisions de  $X_{n+2}, \dots, X_{n+h}$ . De quelle façon ?

On écrit le développement en moyenne mobile

$$X_{n+h} = \varepsilon_{n+h} + \alpha_1 \varepsilon_{n+h-1} + \cdots + \alpha_{h-1} \varepsilon_{n+1} + \alpha_h \varepsilon_n + \cdots.$$

Le prédicteur est

$$\begin{aligned}\hat{X}_{n+1,h-1} &= P_{[X]_{-\infty}^{n+1}}(X_{n+h}) = \alpha_{h-1}\varepsilon_{n+1} + \alpha_h\varepsilon_n + \dots \\ &= \alpha_{h-1}\varepsilon_{n+1} + P_{[X]_{-\infty}^n}(X_{n+h}) = \alpha_{h-1}\varepsilon_{n+1} + \hat{X}_{n,h}.\end{aligned}$$

Donc la prévision de  $X_{n+h}$  s'obtient en ajoutant  $\alpha_{h-1}\varepsilon_{n+1}$  à la prévision précédemment obtenue. Or on connaît la valeur de ce terme complémentaire. En effet,  $\varepsilon_{n+1}$  est, on l'a vu, l'erreur de prévision  $X_{n+1} - \hat{X}_{n,1}$ . On a donc le résultat suivant

**PROPOSITION 7.4.** *Lorsque l'observation  $X_{n+1}$  vient s'ajouter aux précédentes, la mise à jour des prévisions se fait de la façon suivante :*

$$\hat{X}_{n+1,h-1} = \hat{X}_{n,h} + \alpha_{h-1}(X_{n+1} - \hat{X}_{n,1}).$$

Autrement dit, on ajoute au prédicteur précédent de  $X_{n+h}$  une correction proportionnelle à l'erreur que l'on avait faite en prédisant, avant de l'avoir observée, la donnée que l'on vient de recueillir.

## CHAPITRE 8

### Identification

Le statisticien à qui on demande de faire la prévision d'un certain phénomène dispose des données  $x_1, \dots, x_n$ . Par contre il n'a pas idée, en général, d'un modèle de processus bien adapté à ces données. Il va chercher dans l'ensemble des modèles *ARMA* celui qui convient le mieux. On dit qu'il *identifie* le modèle. Pour cela il doit tout d'abord choisir les ordres  $p$  et  $q$  du modèle *ARMA*, puis, lorsque ces ordres sont choisis, il doit estimer les paramètres  $a_1, \dots, a_p, b_1, \dots, b_q$ , et la variance  $\sigma^2$  du bruit d'innovation. En réalité il (ou plutôt le logiciel dont il se sert) procède en sens inverse. Il choisit deux valeurs à ne pas dépasser  $p^*$  et  $q^*$  pour les ordres. Puis, pour chaque couple  $p \leq p^*, q \leq q^*$ , il estime les paramètres  $a_1, \dots, a_p, b_1, \dots, b_q$ , et  $\sigma^2$  du modèle  $ARMA_{p,q}$ . Ensuite seulement il choisit le meilleur parmi tous ces modèles.

#### 1. Estimation des paramètres d'un modèle ARMA

##### 1.1. Estimation préliminaire.

###### 1.1.1. Exemple.

Considérons le processus  $AR_2$  de l'exercice 7. Montrons que les paramètres  $a_1, a_2, \sigma^2$  sont fonction de  $\sigma(2), \sigma(1), \sigma(0)$ .

On a vu que

$$\sigma(1) = \frac{-a_1}{1 + a_2} \sigma(0), \quad \sigma(2) = -a_1 \sigma(1) - a_2 \sigma(0).$$

De ces deux équations on peut déduire l'expression des paramètres  $a_1$  et  $a_2$  en fonction de  $\sigma(2), \sigma(1)$  et  $\sigma(0)$ . On trouve

$$a_1 = \sigma(1) \frac{\sigma(2) - \sigma(0)}{\sigma(0)^2 - \sigma(1)^2}, \quad a_2 = \sigma(1) \frac{\sigma(1) - \sigma(2)}{\sigma(0)^2 - \sigma(1)^2}.$$

Par ailleurs, en calculant la variance des deux membres de l'équation

$$X_n = -a_1 X_{n-1} - a_2 X_{n-2} + \varepsilon_n,$$

on obtient

$$\sigma(0)(1 - a_1^2 - a_2^2) = 2a_1 a_2 \sigma(1) + \sigma^2,$$

qui permet d'exprimer la variance du bruit en fonction de  $\sigma(2), \sigma(1)$  et  $\sigma(0)$ .

Comme on dispose d'estimateurs convergents de ces auto-covariances (voir chapitre 5, paragraphe 4), on peut estimer  $a_1, a_2$  et  $\sigma^2$  de façon convergente en remplaçant les  $\sigma(j)$  par les  $\hat{\sigma}_n(j)$  dans les trois expressions précédentes.

### 1.1.2. Cas général.

On admettra que, de la même façon, les paramètres  $a_1, \dots, a_p, b_1, \dots, b_q$  et  $\sigma^2$  d'un  $ARMA_{p,q}$  s'expriment en fonction de  $\sigma(0), \dots, \sigma(p+q)$ . De là on déduit la construction d'estimateurs convergents des paramètres.

**EXERCICE 19.** Faites le calcul pour un processus  $MA_1$ .

△△

## 1.2. Estimation du maximum de vraisemblance.

En réalité les estimateurs précédents peuvent être améliorés.

Pour cela on utilise la méthode du maximum de vraisemblance. Cette vraisemblance  $V(X_1, \dots, X_n, a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2)$  a en général une expression trop compliquée pour qu'on puisse en trouver le maximum par le calcul exact. On se contente d'une méthode numérique. Par exemple, on calcule la vraisemblance pour les valeurs des paramètres  $(a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2)$  formant une grille suffisamment fine et on prend comme estimateur le point de la grille qui donne la plus grande valeur de la vraisemblance. L'estimateur préliminaire proposé au paragraphe 1.1 sert alors simplement à bien placer la grille sur laquelle on fait les calculs.

Nous ne donnerons pas plus de détails.

## 2. Choix du modèle

Après la procédure décrite plus haut, on a obtenu, pour tout choix des ordres  $p$  et  $q$ , un modèle  $ARMA_{p,q}$  estimé. Il reste à choisir le meilleur de ces modèles.

### 2.1. Critères d'information.

Toutes les méthodes de choix commencent par l'optimisation d'un certain critère (AIC, BIC, etc...). Ces critères sont tous basés sur les deux idées suivantes :

- Étant donné que  $\sigma^2$  est la variance d'erreur de prévision à horizon 1, on aimerait choisir, parmi les modèles estimés, celui qui fournit la plus petite valeur de  $\hat{\sigma}_n^2$ .
- On ne peut pas, pour des raisons statistiques, choisir un modèle présentant un grand nombre de paramètres ( $p$  et  $q$  grands). Pour un tel modèle, la qualité de l'estimation risque d'être mauvaise si la série n'est pas assez longue.

Or, ces deux exigences, pouvoir prédictif et réalisme statistique, sont antagonistes. On prédit mieux avec un  $ARMA$  long, mais on l'estime moins bien.

Les critères utilisés permettent d'équilibrer ces deux exigences. Pour les détails, consulter par exemple [1] et [5].

### 2.2. Vérification.

L'optimisation du critère conduit à la sélection de deux ordres  $p$  et  $q$  et à l'estimation des  $p+q+1$  paramètres correspondants. On obtient donc à partir des données  $x_1, \dots, x_n$

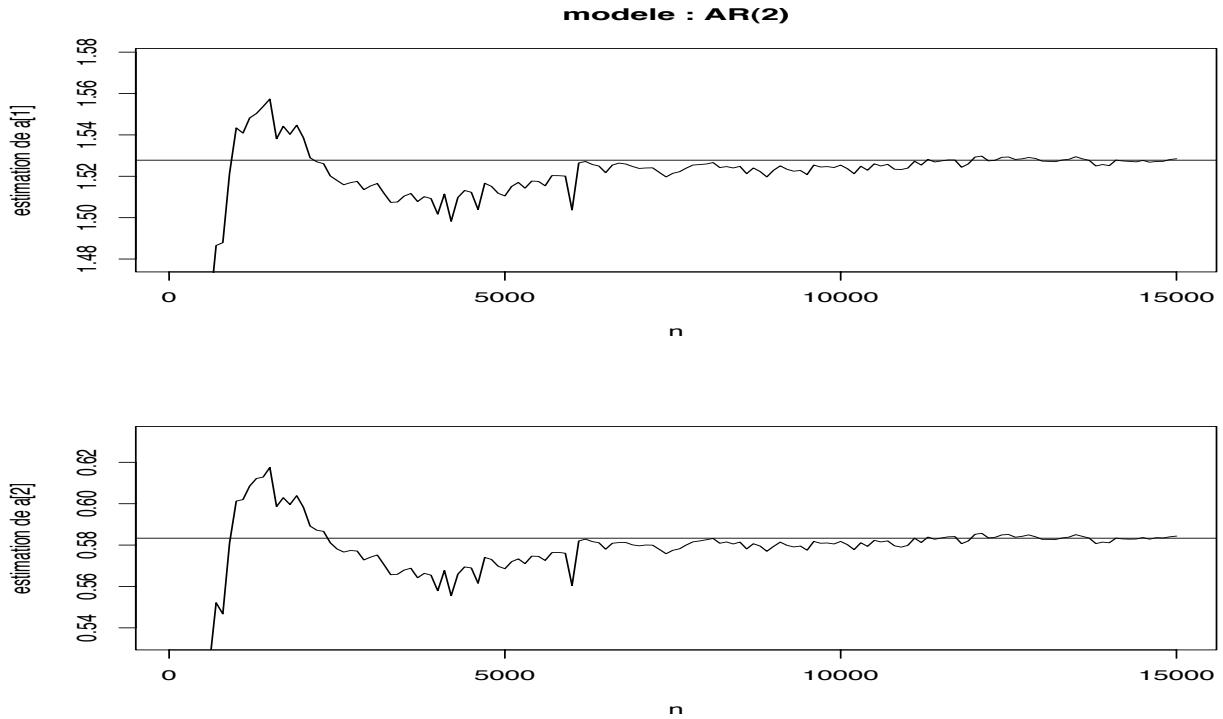


FIG. 8.1. Convergence des estimateurs de  $a_1$  et  $a_2$  dans un modèle  $AR_2$

un modèle estimé

$$X_j + \hat{a}_{1,n}X_{j-1} + \cdots + \hat{a}_{p,n}X_{j-p} = \varepsilon_j + \hat{b}_{1,n}\varepsilon_{j-1} + \cdots + \hat{b}_{q,n}\varepsilon_{j-q}$$

où la variance du bruit est  $\hat{\sigma}_n^2$ .

Il est alors tout à fait possible d'estimer les  $\varepsilon_j$ . Par exemple, si le modèle est autorégressif ( $q = 0$ ), il est naturel de prendre

$$\hat{\varepsilon}_j = x_j + \hat{a}_{1,n}x_{j-1} + \cdots + \hat{a}_{p,n}x_{j-p}, \quad j = p+1, \dots, n,$$

où les  $x_j$  sont les observations. Pour un modèle mixte ou pour un  $MA_q$ , il faut faire appel au développement autorégressif infini donné au chapitre 6 (Proposition 7.1), en posant  $x_j = 0$  si  $j \leq 0$ .

On vérifie alors que le modèle choisi est convenable en testant que le bruit ainsi estimé est bien blanc, c'est à dire que ses auto-corrélations empiriques des résidus sont statistiquement nulles.

**EXERCICE 20.** De quelle façon estimeriez vous le bruit si le modèle ajusté est un  $MA_1$  ? Même question pour un  $ARMA_{1,1}$ .  $\triangle\triangle$

### 3. Identification et prévision

Dans ce paragraphe on se contente d'illustrer ce qui précède sur trois exemples. Les résultats sont représentés dans les figures 8.2 et 8.3.

- Figure 8.2.

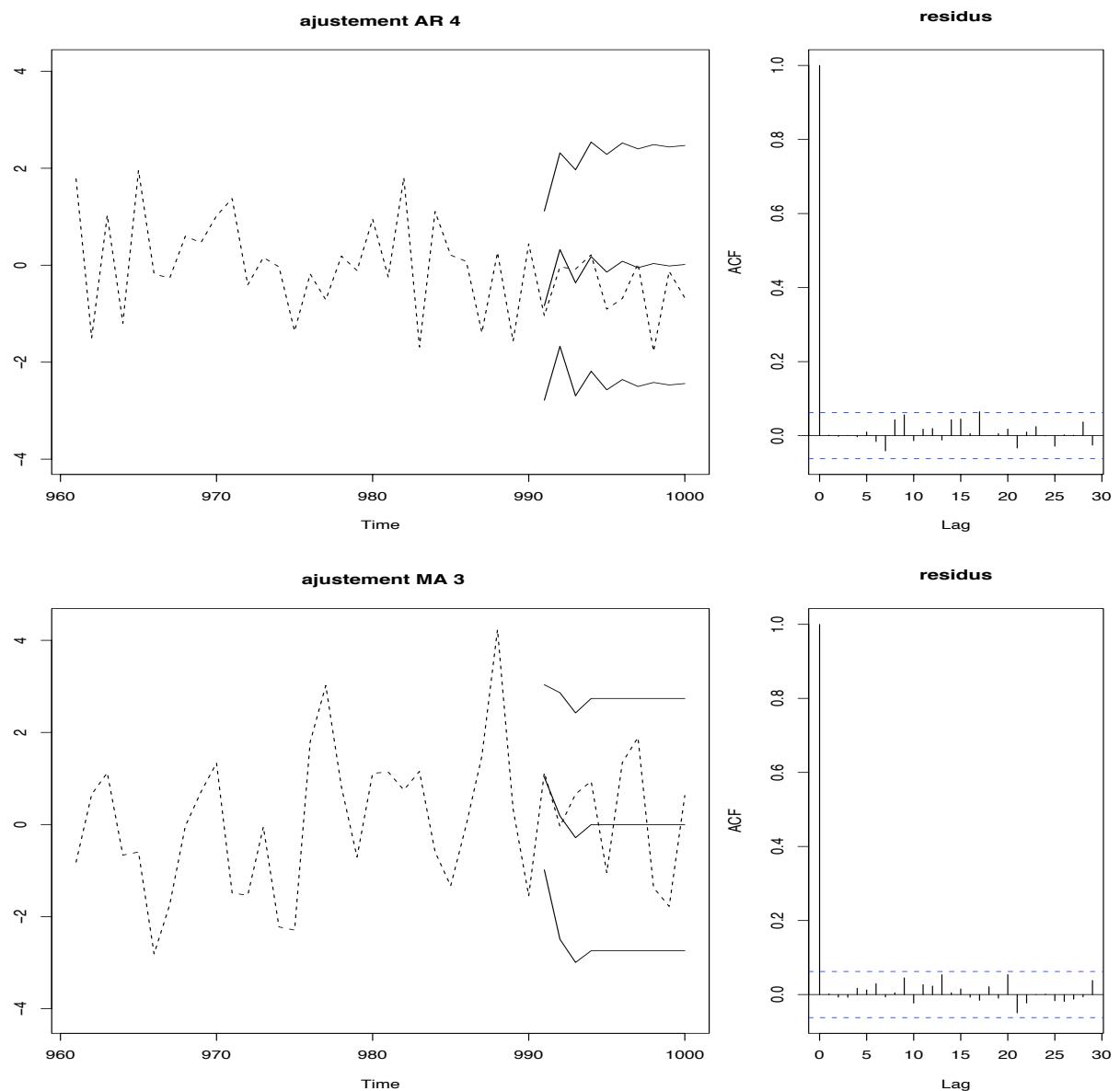
Sur des séries simulées (un  $AR_4$  en haut et un  $MA_3$  en bas) de longueur 1000 on a ajusté un modèle, puis effectué les prévisions à horizon  $1, 2, \dots, 10$ . Le graphique de gauche représente la série, les prévisions faites à partir de  $n = 990$ , et les intervalles de confiance sur la prévision. Le graphique de droite représente l'auto-corrélation des résidus.

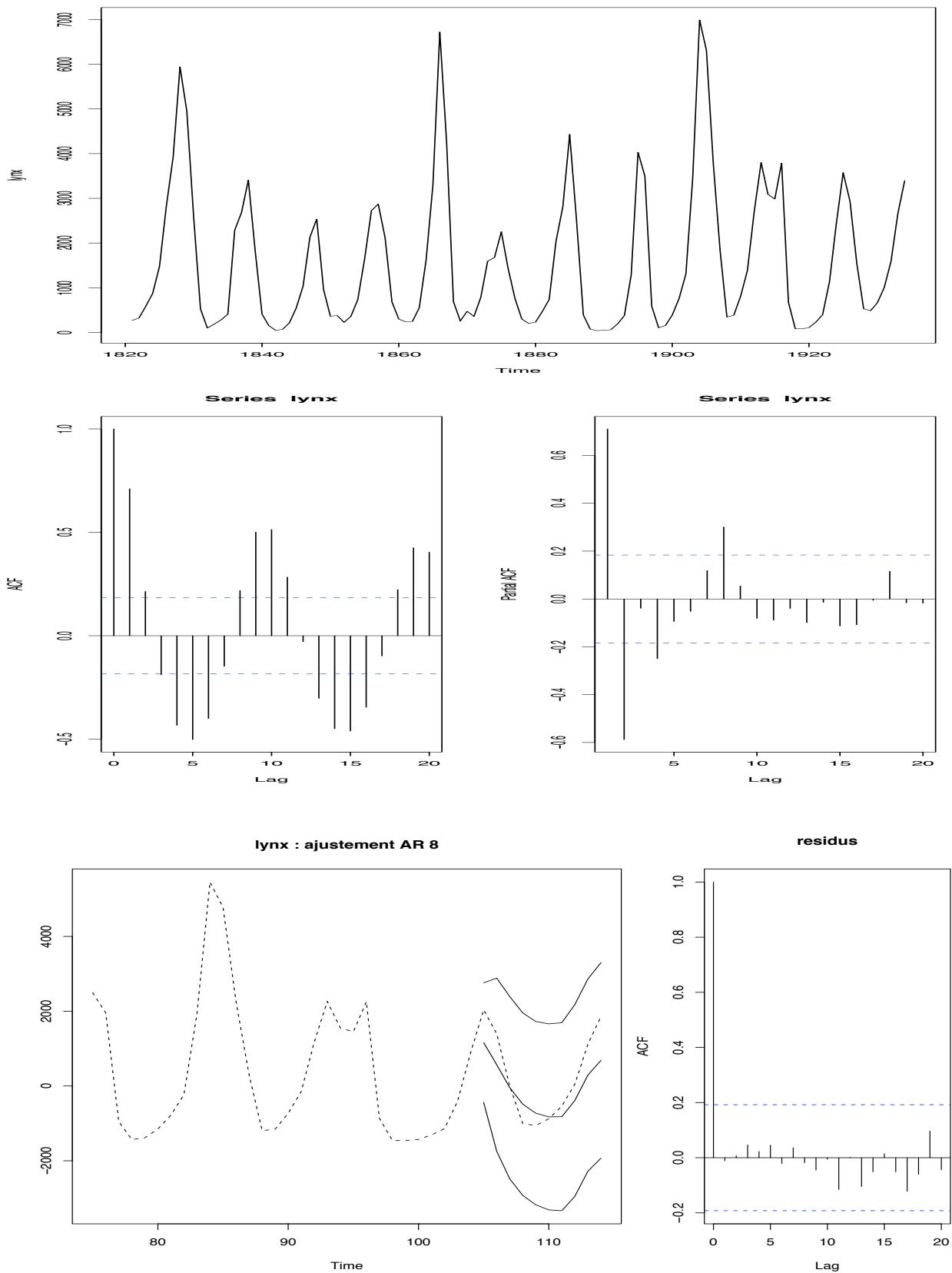
- Figure 8.3.

Il s'agit ici d'une série réellement observée : le nombre annuel de lynx capturés par piégeage au Canada entre 1820 et 1935. Cette série, représentée dans le graphique du haut, a une allure périodique (son auto-corrélation et son auto-corrélation partielle sont représentés dans les graphiques du milieu).

On pourrait penser à estimer la période et ensuite à supprimer la composante périodique. En fait, il est difficile de travailler avec une période estimée, ne serait-ce que parce qu'elle n'est pas forcément entière (comment désaisonnaliser, alors?). On choisit donc de lancer un programme d'identification  $ARMA$ . Le modèle qui optimise le critère AIC est un  $AR_8$ .

Les graphiques du bas représentent l'auto-corrélation des résidus (à droite), et les prévisions (et intervalles de prévision) à horizon allant de 1 à 10 calculés sur la fin de la série.

FIG. 8.2. Identification et prévision dans un  $AR_4$  et un  $MA_3$  simulés

FIG. 8.3. La série des lynx. Ajustement d'un  $AR_8$  et prévision

## CHAPITRE 9

### **Récapitulons : les prévisions ARIMA et SARIMA**

Reprendons la procédure depuis le début. Un statisticien est chargé de faire la prévision d'un certain phénomène. L'unique information dont il dispose est la série temporelle observée  $x_1, \dots, x_n$ .

#### **1. Résumé des étapes déjà vues**

*Étape 1.* Suppression des tendances et désaisonnalisation.

Par exemple, si la série n'est pas centrée et présente une période de longueur 12, on la transforme en

$$y_j = x_j - x_{j-12}, \quad j = 13, \dots, n$$

*Étape 2.* Ajustement d'un modèle ARMA adapté à la série  $y_j$  (voir chapitre 8).

*Étape 3.* Prévision de  $y_{n+1}, \dots, y_{n+h}$  en s'appuyant sur le modèle identifié à l'étape 2 (voir chapitre 7).

*Étape 4.* Retour à la série  $x_j$  d'origine.

C'est cette dernière étape qu'il nous reste à détailler.

#### **2. Filtrage inverse : modélisation ARIMA et SARIMA**

##### **2.1. Exemple.**

Reprendons l'exemple donné à l'étape 1. Pour tout  $j \in \{13, \dots, n\}$ , on a  $x_j = y_j + x_{j-12}$ . D'où  $x_j = y_j + y_{j-12} + x_{j-24}$ , puis  $x_j = y_j + y_{j-12} + y_{j-24} + x_{j-36}$  et ainsi de suite jusqu'à obtenir

$$x_j = y_j + y_{j-12} + y_{j-24} + \dots + y_{r(j)+12} + x_{r(j)}$$

où l'indice  $r(j)$  n'est autre que le reste de la division de  $j$  par 12.

Puisque le statisticien dispose des  $x_j$ ,  $j \leq n$  ainsi que des prévisions de  $y_{n+1}, \dots, y_{n+h}$ , on en déduit les prévisions cherchées. Par exemple, pour un horizon  $h = 1$ ,

$$\hat{x}_{n,1} = \hat{y}_{n,1} + y_{n-11} + y_{n-23} + \dots + y_{r(n+1)+12} + x_{r(n+1)}.$$

**EXERCICE 21.** Reprendre ce qui précède pour une série  $x_j$  qui présente une tendance linéaire sans saisonnalité.  $\triangle\triangle$

## 2.2. Les modèles ARIMA et SARIMA.

Il s'agit simplement de formaliser ce qui vient d'être décrit sur un exemple. Pour simplifier les écritures, on notera  $\Delta$  l'opération de différenciation associant  $(X_n - X_{n-1})_{n \in \mathbb{Z}}$  au processus  $(X_n)_{n \in \mathbb{Z}}$ . Ainsi le processus  $X_n - 2X_{n-1} + X_{n-2}$ , obtenu en différenciant deux fois de suite, sera noté  $\Delta^2 X_n$ . On définit de même  $\Delta^d$ , l'opération qui consiste à différencier  $d$  fois de suite.

L'opération qui fait passer de  $(X_n)$  à  $(X_n - X_{n-D})$  est notée  $\Delta_D$  (à ne pas confondre avec  $\Delta^D$ !).

Vous savez que  $\Delta^2$  annule les tendances linéaires et que  $\Delta_D$  annule les suites  $D$ -périodiques et remet la moyenne à zéro.

**DÉFINITION 9.1.** *On dit que  $(X_n)$  est un processus ARIMA $_{p,d,q}$  si le processus*

$$Y_n = \Delta^d X_n$$

*est un processus ARMA $_{p,q}$ .*

Ce type de modèle convient bien aux séries chronologiques présentant une tendance polynomiale de degré  $d - 1$ .

**DÉFINITION 9.2.** *On dit que  $(X_n)$  est un processus SARIMA $_{p,d,D,q}$  si le processus*

$$Y_n = \Delta_D \circ \Delta^d X_n$$

*est un processus ARMA $_{p,q}$ . (Ici,  $\Delta_D \circ \Delta^d$  est l'opération qui consiste à différencier  $d$  fois, suivie de  $\Delta_D$ , ce qui, d'ailleurs, peut se faire dans l'ordre inverse).*

Ce type de modèle convient bien aux séries chronologiques présentant une période de longueur  $D$  superposée à une tendance polynomiale de degré  $d$ .

Il existe évidemment des modèles plus compliqués, tenant compte notamment de la présence éventuelle de plusieurs périodes différentes. On trouvera une description complète des modèles SARIMA dans [5].

## 2.3. Que font les logiciels de prévision, que fait le statisticien ?

- Le statisticien décide du pré-traitement qu'il veut faire subir à la série : il décide par exemple d'appliquer une transformation logarithme (exemple des ventes de champagne), et il choisit  $d$  et  $D$ , c'est à dire le type de désaisonnalisation et le nombre de différenciations et successives qu'il désire faire subir à la série.
- Le logiciel différencie et désaisonnalise comme on le lui a demandé. Sur la série ainsi modifiée il ajuste un modèle ARMA (voir chapitre 8). Le statisticien peut encore intervenir dans le choix du modèle car il arrive que le meilleur modèle au sens du critère en cours dans le logiciel (AIC, BIC, etc...) ne soit pas celui qui donne le résidu le plus "blanc".
- S'appuyant sur le modèle estimé, le logiciel fournit la prévision de la série modifiée (et l'intervalle de prévision) à l'horizon demandé (voir chapitre 7).
- Enfin il revient à la prévision sur la série d'origine en effectuant les filtrages inverses comme on l'a vu sur deux exemples au début du présent paragraphe.

La figure 9.1 illustre ce qui précède sur la série du trafic aérien (graphique du haut) et la série des ventes de champagne (deuxième ligne). Les graphiques du bas représentent les résidus des ajustements (à gauche, le trafic aérien).

Pour les ventes de champagne, on a identifié un  $SARIMA_{0,0,12,13}$ . Plus précisément, on choisit de modéliser  $\Delta_{12}X_n$  par un processus moyenne mobile d'ordre 13 de la forme

$$Y_n = \varepsilon_n + b_1\varepsilon_{n-1} + b_2\varepsilon_{n-2} + \dots + b_{13}\varepsilon_{n-13}.$$

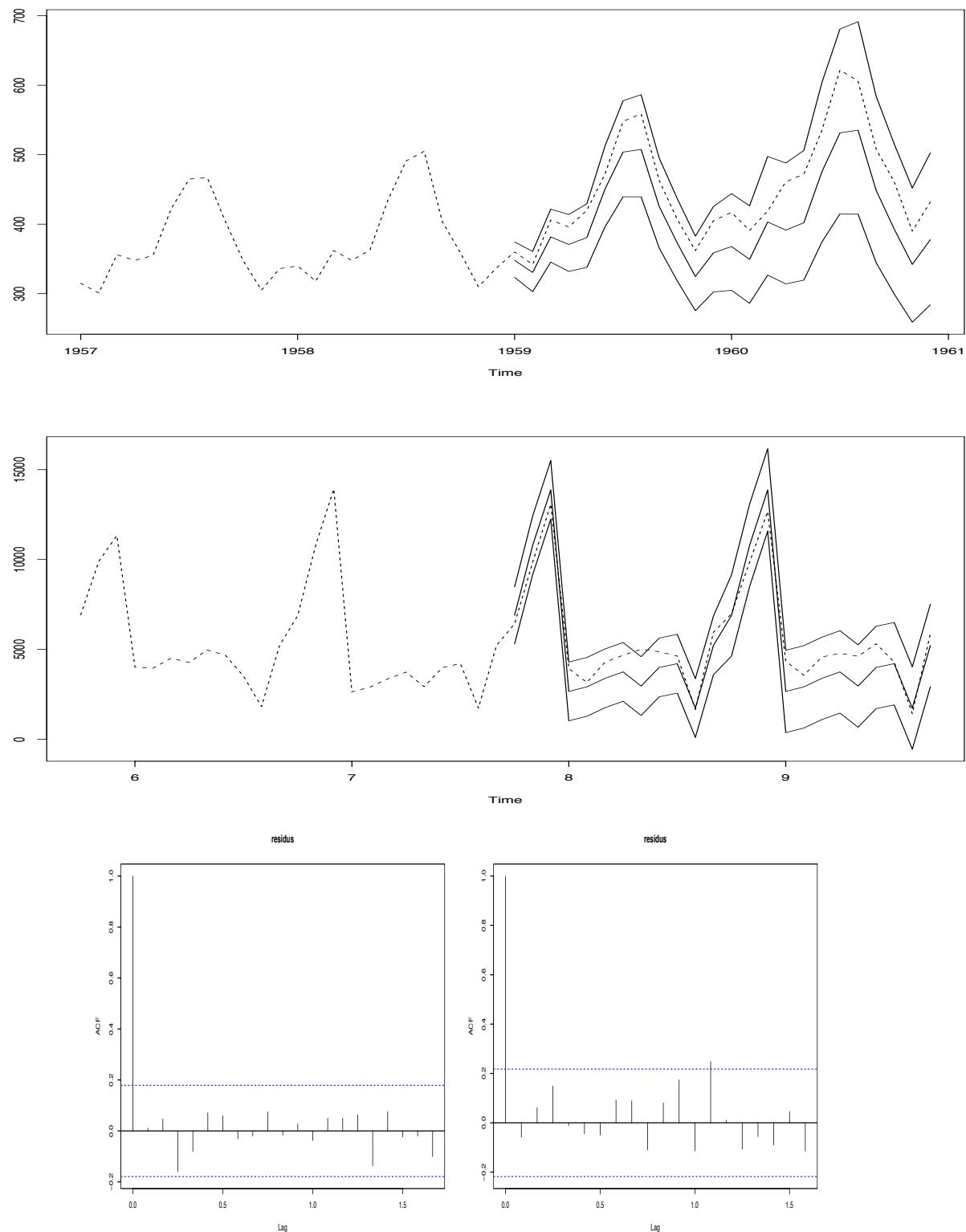


FIG. 9.1. Série du trafic aérien et série des ventes de champagne : ajustement d'un *SARIMA* et prévision

## CHAPITRE 10

### Initiation à la prévision non paramétrique

Supposons, comme pour les processus autorégressifs, que seule une partie de longueur  $r$  fixé du passé serve à construire le présent, à un bruit près. Mais ne supposons plus que la partie du passé qui intervient s'exprime linéairement. Plus précisément, pour tout  $n$ ,

$$(48) \quad X_n = g(X_{n-1}, \dots, X_{n-r}) + \varepsilon_n$$

où  $\varepsilon$  est un bruit blanc centré, tel que, pour tout  $n$ ,  $\varepsilon_n$  soit indépendant de  $(X_{n-1}, \dots, X_{n-r})$ . La fonction  $g$  est inconnue. Pour  $r = 1$ , vous reconnaissiez une dépendance Markovienne.

On cherche à prédire la valeur de  $X_{n+1}$ , en se basant sur l'observation de  $X_n, \dots, X_1$ . Supposant que pour tout  $j$   $\mathbb{E}|X_j| < \infty$ , on a évidemment

$$\mathbb{E}(X_{n+1}|X_n, \dots, X_{n-r+1}) = g(X_n, \dots, X_{n-r+1}),$$

qui est donc (au sens de  $L^2$ ) le meilleur prédicteur de  $X_n$  lorsqu'on a observé le passé. Il s'agit d'en proposer un estimateur puisqu'on ne connaît pas  $g$ . Lorsqu'on sait que  $g$  est linéaire, on est ramené la prédiction des coefficients d'un autorégressif dont on connaît l'ordre. Que faire si on ne sait rien sur  $g$  ?

#### 1. Idée de base

Pour simplifier prenons  $r = 1$  :

$$X_n = g(X_{n-1}) + \varepsilon_n \quad \forall n.$$

Supposons que le processus  $X$  est stationnaire et que la loi des grands nombres est valide. Il faut poser des conditions sur  $g$  pour que l'équation (48) admette une solution qui vérifie ces conditions, mais nous n'aborderons pas cette question ici. Très artificiellement, supposons aussi que les variables  $X_n$  ne prennent qu'une infinité dénombrable de valeurs,  $(z_1, z_2, \dots)$ . Ceci va nous permettre quelques remarques de bon sens qui aideront par la suite.

Oublions momentanément le problème de la prévision pour nous contenter de l'estimation de  $g(z_k)$ , pour un  $z_k$  fixé. Considérons

$$\frac{\sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{\{z_k\}}(X_j)}{\sum_{j=1}^{n-1} \mathbb{I}_{\{z_k\}}(X_j)} = \frac{n^{-1} \sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{\{z_k\}}(X_j)}{n^{-1} \sum_{j=1}^{n-1} \mathbb{I}_{\{z_k\}}(X_j)}.$$

Lorsque  $n \rightarrow \infty$ , le numérateur converge presque sûrement vers  $\mathbb{E}(X_1 \mathbb{I}_{\{z_k\}}(X_0))$  et le dénominateur vers  $P(X_0 = z_k)$ . Donc le quotient converge vers  $\mathbb{E}(X_1|X_0 = z_k) = g(z_k)$ .

Evidemment ce petit raisonnement ne tient plus si les variables  $X_n$  ne sont pas discrètes car alors à la fois  $\mathbb{E}(X_1 \mathbb{I}_{\{z_k\}}(X_0))$  et  $P(X_0 = z_k)$  peuvent être nuls.

Il est alors intéressant de considérer,  $x$  étant fixé,

$$\frac{\sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{[x-v, x+v]}(X_j)}{\sum_{j=1}^{n-1} \mathbb{I}_{[x-v, x+v]}(X_j)} = \frac{n^{-1} \sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{[x-v, x+v]}(X_j)}{n^{-1} \sum_{j=1}^{n-1} \mathbb{I}_{[x-v, x+hv]}(X_j)},$$

ce qui consiste à faire la moyenne des  $X_{j+1}$  dont l'antécédent est dans un intervalle de longueur  $v$  autour de  $x$ .

Pourvu que  $P(X_0 \in [x-v, x+v])$  ne soit pas nul, ceci tend vers  $\mathbb{E}(X_1 | X_0 \in [x-v, x+v])$ , qui n'est certes pas  $g(x)$ , mais qui tend vers  $g(x)$  quand  $v \rightarrow 0$ , du moins sous certaines conditions de régularité.

La bonne idée semble donc être de considérer une suite  $v_n$  qui tend vers zéro quand  $n \rightarrow \infty$  et de prendre comme estimateur de  $g(x)$

$$\hat{g}_n(x) = \frac{\sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)}{\sum_{j=1}^{n-1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)} = \frac{n^{-1} \sum_{j=1}^{n-1} X_{j+1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)}{n^{-1} \sum_{j=1}^{n-1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)} = \hat{g}_n(x).$$

## 2. Quelques mots sur l'asymptotique

Avant d'affiner l'outil que l'on vient d'ébaucher, disons quelques mots sur ce qui se passe lorsque  $n \rightarrow \infty$ . On décompose l'écart entre l'estimateur et  $g(x)$  en

$$\hat{g}_n(x) - g(x) = \frac{\sum_{j=1}^{n-1} (g(X_j) - g(x)) \mathbb{I}_{[x-v_n, x+v_n]}(X_j)}{\sum_{j=1}^{n-1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)} + \frac{\sum_{j=1}^{n-1} \varepsilon_{j+1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)}{\sum_{j=1}^{n-1} \mathbb{I}_{[x-v_n, x+v_n]}(X_j)}.$$

Si la fonction  $g$  est uniformément continue, le premier terme est majoré en valeur absolue par  $\sup\{|g(x) - g(y)| / |x - y| \leq v_n\}$  qui tend vers zéro si  $v_n \rightarrow 0$ . Quant au deuxième terme on montre qu'il tend vers zéro à condition que son dénominateur tends vers l'infini. Ce qu'il convient de retenir de tout ceci est que  $v_n$ , qu'on appelle souvent « largeur de la fenêtre » (pourquoi ?), doit tendre vers zéro, mais pas trop vite. D'où l'importance d'un bon réglage.

## 3. Autre écriture et généralisation à d'autres noyaux

**3.1. Autre écriture.** Il est bien instructif de constater que  $\hat{g}_n(x)$  n'est autre qu'une combinaison linéaire des  $X_j$

$$\hat{g}_n(x) = \sum_{j=1}^n X_{j+1} W_{n,j}(x)$$

où le poids  $W_{n,j}(x)$  mesure la proximité de  $x$  à  $X_j$ . Les poids sont de somme 1.

**3.2. Généralisation.** Une fois ce constat fait, il est facile de se libérer de la forme particulière des poids. On avait

$$W_{n,j}(x) = \frac{\mathbb{I}_{[x-h, x+h]}(X_j)}{\sum_{l=1}^{n-1} \mathbb{I}_{[x-h, x+h]}(X_l)}.$$

On va le remplacer par

$$W_{n,j}(x) = \frac{K\left(\frac{x-X_j}{v_n}\right)}{\sum_{l=1}^{n-1} K\left(\frac{x-X_l}{v_n}\right)},$$

ce qui revient à remplacer l'indicatrice de  $[-1, 1]$  par une fonction  $K$  plus régulière (par exemple la discontinuité de l'indicatrice en  $\pm 1$  peut créer quelques effets désagréables), en conservant toutefois quelques propriétés comme la positivité, la symétrie par rapport à zéro, la décroissance sur  $\mathbb{R}^+$  et l'intégrabilité. La fonction  $K$  est appelée « noyau ». C'est un indicateur de la proximité de l'origine. On peut notamment choisir la fonction triangulaire

$$K(x) = (1 - |x|)\mathbb{I}_{[-1,1]}(x)$$

ou encore le noyau gaussien

$$K(x) = \exp -\frac{x^2}{2},$$

ce qui donne

$$\hat{g}_n(x) = \frac{\sum_{j=1}^{n-1} X_{j+1} \exp -\frac{(x-X_j)^2}{2v_n^2}}{\sum_{j=1}^{n-1} \exp -\frac{(x-X_j)^2}{2v_n^2}}.$$

#### 4. Généralisation à d'autres mémoires

Revenant à l'équation autorégressive non linéaire (48), on voit qu'il convient de remplacer l'écart entre  $X_j$  et  $x$  à celui entre  $(X_j, \dots, X_{j-r+1})$  et  $(x_r, \dots, x_1)$  si l'on veut estimer  $g(x_1, \dots, x_r)$ . Il suffit pour cela d'utiliser comme noyau une fonction convenable de  $r$  variables, par exemple

$$K(x_1, \dots, x_r) =_{not.} K(x_1^r) = \exp -\frac{\sum_{l=1}^r x_l^2}{2},$$

ce qui donne,

$$\begin{aligned} \hat{g}_n(x_1, \dots, x_r) &= \frac{\sum_{j=r}^{n-1} X_{j+1} K\left(\frac{x_1^r - X_{1+j-r}^j}{v_n}\right)}{\sum_{j=r}^{n-1} K\left(\frac{x_1^r - X_{1+j-r}^j}{v_n}\right)} \\ &= \frac{\sum_{j=r}^{n-1} X_{j+1} \exp -\frac{\sum_{l=1}^r (x_l - X_{l+j-r})^2}{2v_n^2}}{\sum_{j=r}^{n-1} \exp -\frac{\sum_{l=1}^r (x_l - X_{l+j-r})^2}{2v_n^2}} \\ &= \sum_{j=r}^{n-1} X_{j+1} W_{n,j}(x_1^r). \end{aligned}$$

## 5. Retour à la prévision

Pour prévoir  $X_{n+1}$  en s'appuyant sur le passé observé  $(X_n, \dots, X_1)$  l'idée la plus naturelle est de remplacer le vecteur  $(x_1, \dots, x_r)$  par  $(X_{n-r+1}, \dots, X_n)$  dans l'estimateur de  $g(x)$  donné au paragraphe 4. Soit

$$(49) \quad \hat{X}_{n+1|n} = \frac{\sum_{j=r}^{n-1} X_{j+1} K\left(\frac{X_{n-r+1}^n - X_{j-r+1}^j}{v_n}\right)}{\sum_{j=r}^{n-1} K\left(\frac{X_{n-r+1}^n - X_{1+j-r}^j}{v_n}\right)}$$

$$= \frac{\sum_{j=r}^{n-1} X_{j+1} \exp -\frac{\sum_{l=1}^r (X_{l+n-r} - X_{l+j-r})^2}{2v_n^2}}{\sum_{j=r}^{n-1} \exp -\frac{\sum_{l=1}^r (X_{l+n-r} - X_{l+j-r})^2}{2v_n^2}},$$

$$(50) \quad = \sum_{j=r}^{n-1} X_{j+1} W_{n,j}(X_{n-r+1}^n).$$

ou toute expression analogue obtenue à l'aide d'un autre noyau.

**5.1. Interprétation.** On voit, en examinant de près (49), que le prédicteur est construit sur le principe suivant. On réalise une combinaison linéaire des  $X_l$  d'indices allant de  $r+1$  à  $n$ . Chaque  $X_l$  est pondéré par un coefficient qui est d'autant plus grand que son passé immédiat de longueur  $r$  (soit  $(X_{l-r}, \dots, X_{l-1})$ ) ressemble plus au passé immédiat de même longueur de  $X_n$ . De plus la somme des coefficients est égale à 1. Autrement dit on a un « bloc témoin »  $X_{n-r+1}^n$  de longueur  $n$  auquel on va comparer tous les blocs de même longueur dans le passé de la série. Ces blocs sont plus ou moins ressemblants au bloc témoin. Le degré de ressemblance du bloc qui se termine à  $X_j$  va chiffrer la contribution de  $X_{j+1}$  dans la combinaison linéaire finale.

**5.2. Changement d'horizon.** On pourrait comme on l'a fait dans la modélisation ARMA, vouloir prévoir  $X_{n+h}$  pour  $h > 1$ . Il suffit pour cela de remplacer  $X_{j+1}$  par  $X_{j+h}$  dans l'expression (49), soit

$$(51) \quad \hat{X}_{n+1|n} = \frac{\sum_{j=r}^{n-h} X_{j+h} K\left(\frac{X_{n-r+1}^n - X_{j-r+1}^j}{v_n}\right)}{\sum_{j=r}^{n-h} K\left(\frac{X_{n-r+1}^n - X_{1+j-r}^j}{v_n}\right)}.$$

**EXERCICE 22.** Simuler une série ARMA(p,q) de longueur 100. Sur les 90 premières valeurs, mettre en oeuvre des prédicteurs non paramétriques pour des horizons allant de 1 à 10. Faire varier la longueur du bloc témoin, prendre trois noyaux différents (l'indicatrice de  $[-1, 1]$ , la fonction triangulaire et le noyau gaussien) et faire varier la vitesse de fermeture de fenêtre  $v_n$ . Confronter avec les 10 dernières valeurs de la série.  $\triangle\triangle$

**5.3. Prise en compte d'instationnarités.** Certains auteurs préconisent la procédure suivante lorsque la série présente une tendance et éventuellement une dispersion variable. Pour chaque bloc  $X_{j-r+1}^j = (X_{j-r+1}, \dots, X_j)$  on calcule sa moyenne  $m_{j-r+1}^j$  et son écart-type  $s_{j-r+1}^j$ . Puis, dans l'expression du prédicteur (51) on remplace les blocs par leurs homologues centré et réduits, et on recifie chaque  $X_{j+h}$  de façon à prendre en compte la différence des moyennes et des écarts types entre le bloc témoin et le bloc attaché à  $X_j$ . Avec la notation

$$\tilde{X}_{j-r+1}^j = \frac{X_{j-r+1}^j - m_{j-r+1}^j \delta_r}{s_{j-r+1}^j},$$

où  $\delta_r$  est le vecteur de longueur  $r$  dont toutes les coordonnées sont égales à 1 on obtient

$$(52) \quad \hat{X}_{n+1|n} = \frac{\sum_{j=r}^{n-h} (X_{j+h} + m_{n-r+1}^n - m_{j-r+1}^j) \frac{s_{n-r+1}^n}{s_{j-r+1}^j} K\left(\frac{\tilde{X}_{n-r+1}^n - \tilde{X}_{j-r+1}^j}{v_n}\right)}{\sum_{j=r}^{n-h} K\left(\frac{\tilde{X}_{n-r+1}^n - \tilde{X}_{1+j-r}^j}{v_n}\right)}.$$

#### 5.4. Estimation de la loi conditionnelle et intervalles de prévision.

## Bibliographie

- [1] Brockwell P. J. , Davis R. A. (1991) *Time Series : Theory and Methods*. Springer.
- [2] Brockwell P. J. , Davis R. A. (1996) *Introduction to Time Series and Forecasting*. Springer.
- [3] Gardner E.S Jr. Exponential Smoothing : The State of the Art. *Journal of Forcasting*, 4(1) :1–28, 1985.
- [4] Gourioux C.- Monfort A. *Cours de series temporelles*. Economica, 1983.
- [5] Mélard G. (1990) *Méthodes de prévision à court terme*. Ed. Ellipses.
- [6] Shumway H., Stoffer D.S. (1999) *Time Series Analysis and its Applications, with R-examples*. Springer.