

# Séries chronologiques

October 1, 2007

# Chapter 1

## Introduction

nombre d'observations  $N = ms = 40$   
nombre d'observations  $N = ms = 40$

On ne peut pas éviter les séries temporelles. Exemples :

- Nombre de chômeurs en France
- Température dans le coeur d'un réacteur nucléaire.
- Précipitation quotidienne
- Fuel consommation trimestrielle

### 1.1 Décomposition d'une série temporelle

Les séries temporelles sont souvent composées de 3 parties

1. Tendence
2. Cycle saisonnier
3. Perturbation aléatoire

$$X(t) = T(t) + S(t) + R(t)$$

**Identification d'un système :** Identifier les composants  $T(t)$  et  $S(t)$ . On examine dans la suite un exemple concret

**Fuel consommation exemple** La table 1.1 représente la quantité (en  $m^3$ ) de fuel consommée dans une ville pendant 5 ans. La figure 1.2 trace le *trajectoire* (observation), on remarque facilement que il présente une période et

année	1	2	3	4
1965	874	679	616	816
1966	866	700	603	814
1967	843	719	594	819
1968	906	703	634	844
1969	952	745	635	871
moyenne	888.2	709.2	616.4	835.8

Table 1.1: Fuel consommation

une tendance et il est perturbée par une variable aléatoire.

### 1.2 Identification du système, première méthode

On cherche à estimer les 3 composants  $T(t)$ ,  $S(t)$  et  $R(t)$ .

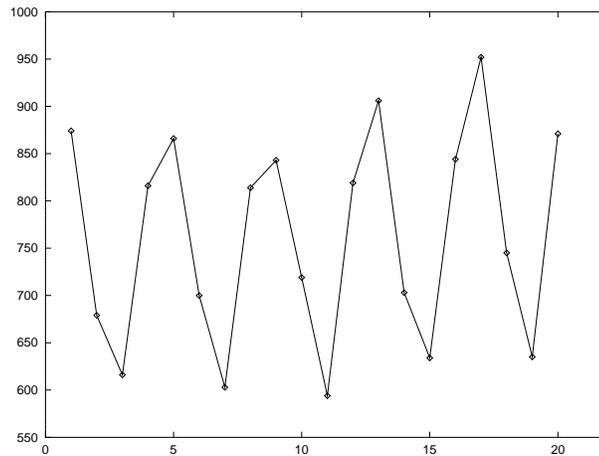


Figure 1.1: Fuel consommation

### 1.2.1 Identification du composant saisonnier

On désigne

nombre de cycle  $m = 5$ , période  $s = 4$ ,  
 nombre d'observations  $N = ms = 40$

la moyenne totale  $M = \frac{1}{N} \sum_{t=1}^N X(t) = 761.65$

**Première méthode** Une méthode très simple pour trouver  $S(1), S(2), S(3)$  et  $S(4)$  est de calculer la différence entre moyenne totale et la moyenne partielle

$$S(1) = 882.2 - 761.65 = 126.55, \quad S(2) = 709.2 - 761.65 = -52.4$$

$$S(3) = 616.4 - 761.65 = -145.25, \quad S(4) = 832.8 - 761.65 = 71.15$$

Soit  $t = k + sj$

$$X(k + sj) = T(k + sj) + S(k + sj) + R(k + sj) \quad j = 0, 1, 2, 3, 4 \quad k = 1, 2, 3, 4$$

En soustrayant le composant  $S(k + sj)$  on obtient la figure 1.3. Cette méthode marche bien en générale quand le composant tendanciel est une fonction linéaire. Soit  $T(t) = a + bt$  ( $b$ : gradient *pen*te, ici très faible,  $a, b$  coefficients inconnus). On examine maintenant le défaut de cette méthode, en ignorant le composant aléatoire, la moyenne totale est (Attention : en effet, la moyenne sur  $R(t)$  est proche de 0)

$$\begin{aligned} \bar{M} &= \frac{1}{N} \sum_{j=1}^N (T(t) + S(t)) = \frac{1}{N} (Na + b \frac{N(N+1)}{2}) + \frac{1}{N} \sum_{j=1}^N S(t) \\ &= a + \frac{N+1}{2} b \end{aligned}$$

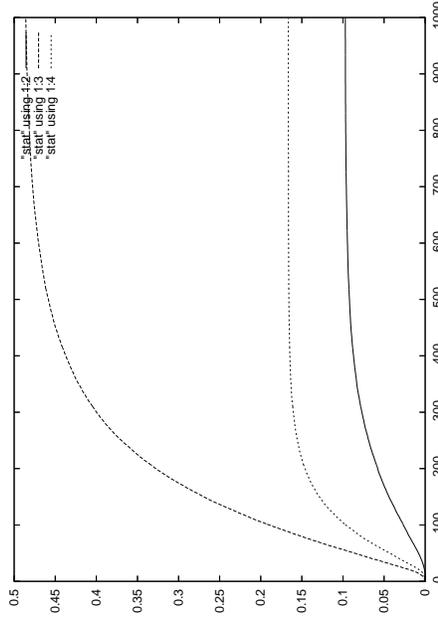


Figure 1.2: Fuel consommation

$\sum_{j=1}^N S(j) = 0$ . La moyenne partielle correspond à  $k$  est

$$\begin{aligned} \bar{M}_k &= \frac{1}{m} \sum_{j=0}^{m-1} (T(k + sj) + S(k + sj)) \\ &= \frac{1}{m} \sum_{j=0}^{m-1} (a + (k + sj)b + S(k + sj)) \\ &= a + bk + bs \frac{m-1}{2} + \frac{1}{m} \sum_{j=0}^{m-1} S(k + sj) \end{aligned}$$

La méthode précédente consiste donc de soustraire

$$\begin{aligned} \bar{S}(k) &= \bar{M}(k) - \bar{M} = a + bk + bs \frac{m-1}{2} + \frac{1}{m} \sum_{j=0}^{m-1} S(k + sj) - (a + \frac{ms+1}{2}b) \\ &= bk - b \frac{s+1}{2} + \frac{1}{m} \sum_{j=0}^{m-1} S(k + sj) \end{aligned}$$

Donc si la pente  $b$  est petite, alors la méthode marche bien.

**Deuxième méthode** On effectue des calculs successifs pour obtenir la table 1.2. Grâce à la dernière colonne (moyenne mobile centrée) on peut estimer le composant saisonnier par

$$S(3) = \frac{1}{4}((616 - 745.25) + (603 - 741.75) + (594 - 750.) + (634 - 777.7)) = -142.00$$

et  $S(1) = 128.594$   $S(2) = -46.487$ ,  $S(4) = 65.00$ . On remarque que  $S(1) + S(2) + S(3) + S(4) = 5.107$  (non zéro). ceci peut être réajusté pour que  $\sum S(i) = 0$

$$S(1) = 128.594 - \frac{5.107}{4} = 127.74 \quad S(2) = -46.487 - \frac{5.107}{4} = -47.763$$

$$S(3) = -142.00 - \frac{5.107}{4} = -143,254 \quad S(4) = 65.00 - \frac{5.107}{4} = 63.746$$

Le résultat est tracé dans la figure 1.4. Cette méthode est facile à implementer. A noter que dans la 3ème colonne il manque des données CMA au début et à la fin de la série. Cette méthode est assez efficace dans le cas où  $b$  est petit.

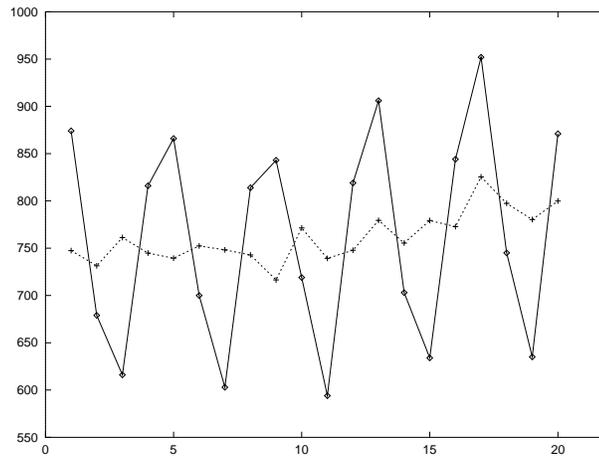


Figure 1.3: composant saisonnier éliminé

### 1.2.2 Identification du composant aléatoire

Nous avons dans cet exemple estimé le composant saisonnier ( $S(1), S(2), S(3), S(4)$ ) et la tendance (sans calculer les coefficients  $a$  et  $b$ ). On peut alors, grâce à l'observation, calculer le composant aléatoire. Le résultat obtenu est représenté par la figure 1.5. Dans les applications, il n'y a très peu d'intérêt à trouver le composant aléatoire, car les perturbations aléatoires changent constamment, seuls  $T(t)$  et  $S(t)$  sont importants. On trace maintenant ces trois composants sur la même figure 1.6.

### 1.2.3 Identification du composant tendanciel

On suppose que la tendance est linéaire

$$T(t) = a + bt$$

Par une méthode d'approximation 1.7 classique on obtient l'estimation

$$\hat{T} = 713.38 + 3.647t$$

## 1.3 Identification du système, deuxième méthode

On présente ici une méthode alternative, mais sans entrer en détails. Soit

$$X_t = \beta_0 + \beta_1 t + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \beta_5 d_4 + \varepsilon(t)$$

où  $\beta_0 + \beta_1 t$  représente le composant tendanciel,  $\beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \beta_5 d_4$  le composant saisonnier avec

$$d_i = \begin{cases} 1 & \text{ième saison} \\ 0 & \text{sinon} \end{cases}$$

En appliquant la méthode de régression, on obtient la solution suivante

$$X_t = 3.38t + 858.83d_1 + 675.45d_2 + 579.28d_3 + 792.30d_4 + \varepsilon(t)$$

Dans la figure 1.8 on compare cette méthode avec l'observation.

saisons	$X(t)$	MA	CMA
1	874		
2	679		
3	616	746.250	745.250
4	816	744.250	746.875
5	866	749.500	747.875
6	700	746.250	746.000
7	603	745.750	741.750
8	814	737.750	740.125
9	834	742.500	741.375
10	719	740.250	740.875
11	594	741.500	750.500
12	819	759.500	757.500
13	906	755.500	760.500
14	703	765.500	768.625
15	634	771.750	777.500
16	844	783.250	788.500
17	952	793.750	793.875
18	745	794.000	797.375
19	635	800.750	
20	871		

Table 1.2: calcul de moyenne mobile

## 1.4 Prédiction

Identification du système, pourquoi ? Une application concrète est la prédiction. En fait, d'après les calculs précédents on a :

$$X_t = T(t) + S(t) + R(t)$$

et

$$\hat{T}(t) = 713.38 + 3.647t$$

$$\hat{S}(1) = 127.34, \hat{S}(2) = -46.763, \hat{S}(3) = -143.254, \hat{S}(4) = 63.746$$

Question : quelle est la quantité consommée pour l'année 1970 ? Pour répondre à cette question, on peut naturellement utiliser la prédiction suivante, le résultat est dans table 1.3

$$\hat{X}_t = \hat{T}(t) + \hat{S}(t), \quad t = 21, 22, 23, 24$$

La qualité de cette prédiction est tracée dans la figure 1.9.

période	tendance	saisonnier	prédiction	observation
$t$	$T$	$S$	$\hat{X}$	$X$
21	789.963	127.340	917.303	981
22	793.610	-47.763	745.847	759
23	797.257	-143.254	654.003	674
24	800.904	63.746	864.650	900

Table 1.3: Prédiction

## 1.5 Études préliminaire d'une série

### 1.5.1 Les données

$X_t, \quad t = 1, 2, \dots, n \quad n$  : taille échantillon

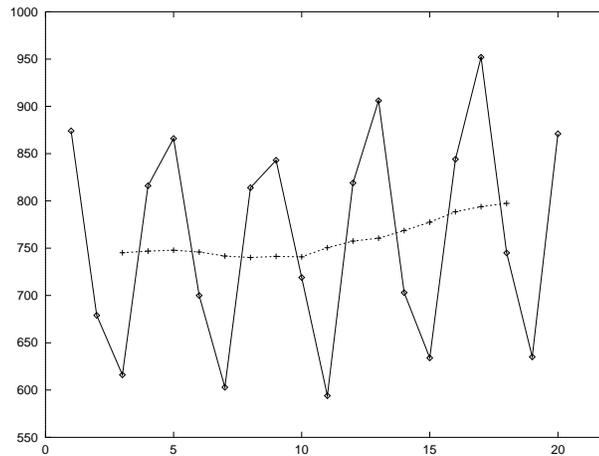


Figure 1.4: Fuel consommation et Moyenne mobile centrée

série observée. Ces observations sont faites à intervalle de temps régulière. A noter que cet *échantillon* n'est pas issu d'un n-uplet de variables aléatoires indépendantes. Voici quelques exemples des séries temporelles

- Relevés mensuels d'un compte bancaire
- longueurs journalière de la tige d'une plante
- Évaluation d'heur en heur du stock pour chaque article d'un hypermarché.

### 1.5.2 Observation de graphique

On n'hésitera pas, avant et pendant l'analyse, à user de graphique. Par simple observation, on pourra s'apercevoir par exemple

- Existence de valeurs manquantes
- Valeurs accidentelles
- Mouvement tendanciel
- mouvement périodique
- ruptures de comportement
- analogie de comportement de plusieurs séries représentées simultanément.

On peut aussi observer les données transformées.

- $\frac{X_t^{\lambda+a}}{b}$  où  $a \in R$ ,  $b \in R$ ,  $\lambda \in R - \{0\}$  (Box-Cox)
- $\log X_t$  choisit pour atténuer une croissance exponentielle.
- $\frac{X_{t+1}-X_t}{X_t}$  taux de croissance

### 1.5.3 Corrections des données

Modifier les données brutes avant de se lancer dans une étude. Quelques taches possibles.

- Évaluation préliminaire de données manquantes, remplacement de données accidentelles.
- Découpage en sous séries : rupture de comportement
- Standardisation(harmonisation), calcul de moyenne pour se ramener à un intervalle de temps d'amplitude fixe entre les relevés successifs. Par exemple le mot *mois*) doit être normalisé selon le contexte.

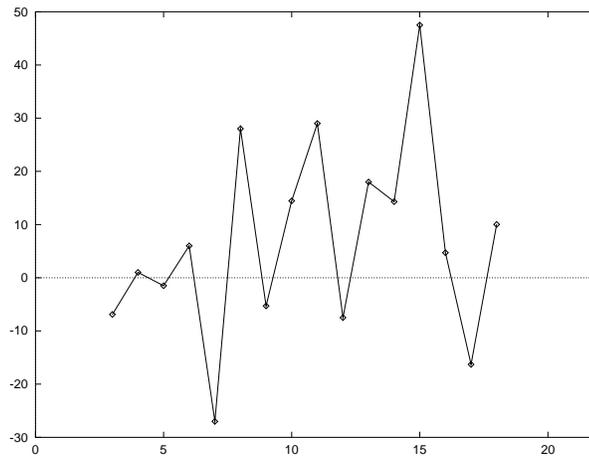


Figure 1.5: Composant irrégulier

- \* en consommation  $365/12$  jours
- \* en production  $(365 - \text{jours fériés})/12$
- Transformations des données pour des raisons divers
  - \* Cartésienne vers polaire
  - \* log.
  - \* etc.

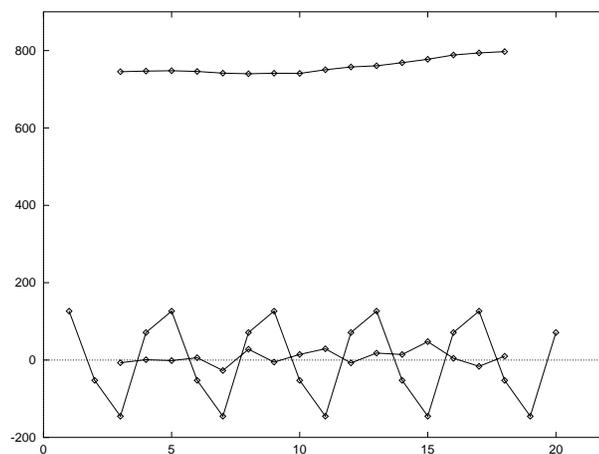


Figure 1.6: Décomposition

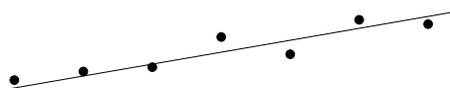


Figure 1.7: Estimation de la tendance

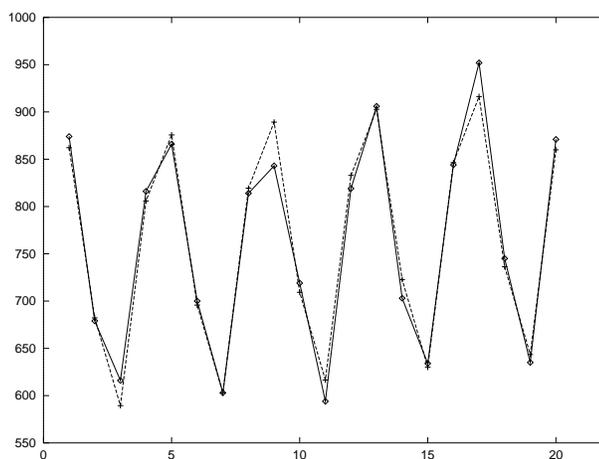


Figure 1.8: Modèle de régression linéaire et observation réelle

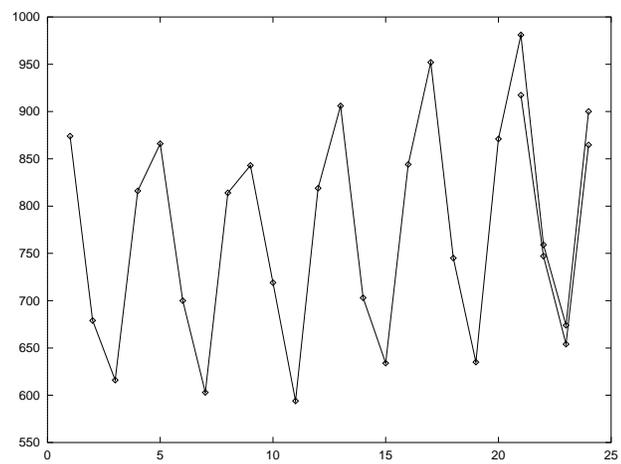


Figure 1.9: Prédiction et observation réelle

# Chapter 2

## Modèles des séries temporelles

La résolution des divers problèmes s'appuie sur des modèles décrivant la façon dont la dérive évolue. Il est commode de distinguer trois types de modèles: *les modèles d'ajustement, les modèles auto-projectifs et les modèles explicatifs*. Nous allons présenter brièvement ces trois modèles.

### 2.1 Modèles d'ajustement

D'une façon générale, les modèles d'ajustement sont sous la forme

$$X_t = f(t, \varepsilon_t)$$

où  $f$  est une fonction indexée par un nombre fini de paramètres inconnus et où  $\varepsilon_t$  est une v. a centrée sur la quelle diverses hypothèses additionnelles peuvent être faites. Il existe en pratique 3 modèles généraux dans cette catégorie de série temporelle, articulants les composants de façons différentes. Notations:

$X_t$	: Chronique donnée observation
$T_t$	: tendance à l'instant $t$
$V_t$	: chronique de variation saisonnière
$\varepsilon(t)$	: chronique de variation accidentelle( $R(t)$ )
$S_t$	: cycle saisonnière.

#### 2.1.1 Modèle Additif

Le modèle additif consiste à poser

$$X_t = T_t + S_t + \varepsilon_t$$

Le modèle le plus simple suppose que le saisonnier de période  $p$  est constant

$$S_{t+p} = S_t, \quad \forall t$$

Il est alors déterminé par  $p$  coefficients  $c_1, c_2, \dots, c_p$ . On impose la contrainte suivante :

$$\sum_{i=1}^p c_i = 0$$

**Exemple 2.1.1** *Soit*

$$T_t = 0.05t + 100$$

$$S_1 = -1, \quad S_2 = 2, \quad S_3 = -2, \quad S_4 = 1$$

$$\varepsilon_t \sim N(0, 0.5^2)$$

Voici la simulation de cet exemple présentée dans la figure 2.1.

De toutes façon, il y a une infinité de solutions au problème de décomposition, correspondant d'abord à l'infinité de choix possible de tendance

$X_t - T_t$  = est la série détendancialisée, elle est égale à saisonnier + résidu

$X_t - S_t$  = est la série désaisonnalisée, elle est égale à tendance + résidu.  
on la nomme aussi *série corrigée des variations saisonnières*

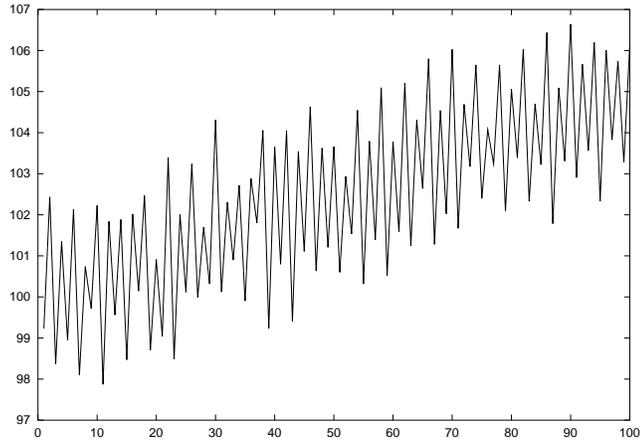


Figure 2.1: Modèle additif

### 2.1.2 Premier modèle multiplicatif

le deuxième modèle est multiplicatif : il suppose que les variations saisonnières et le facteur accidentel ont un effet multiplicatif sur la taille des observations. On le note

$$\begin{aligned} X_t &= T_t(1 + S_t)(1 + R_t) \\ &= T_t + T_t S_t + (T_t + T_t S_t)R_t = T_t + V_t + \varepsilon_t \end{aligned}$$

Dans la mesure où les termes sont positifs, on peut calculer les logarithmes

$$\ln X_t = \ln T_t + \ln(1 + S_t) + \ln(1 + R_t)$$

Nous prenons comme exemple le modèle suivant

$$X_t = e^{a+bt}(1 + S_t)(1 + R_t)$$

avec

$$a = \ln 100, \quad b = 0.05$$

$$S_1 = 0.06, \quad S_2 = -0.06, \quad S_3 = -0.04, \quad S_4 = 0.04$$

$$R_t \sim N(0, 20^2)$$

Le résultat de simulation est présenté dans la figure 2.2. On distingue encore nettement la périodicité des variations saisonnières, mais la différence  $X_t - T_t$  s'accroît avec le temps. On trace maintenant cet exemple dans la figure 2.3

### 2.1.3 Deuxième modèle multiplicatif

Le troisième modèle est

$$X_t = T_t(1 + S_t) + \varepsilon_t$$

Par rapport au modèle précédent, seule la variation saisonnière  $T_t S_t$  est proportionnelle à la tendance, on peut alors écrire

$$X_t = T_t(1 + S_t) + \varepsilon_t$$

Voici un exemple dont le résultat simulé est présenté dans la figure 2.4

$$T_t = 0.05t + \log(100)$$

$$S_1 = 0.06, \quad S_2 = -0.06, \quad S_3 = -0.04, \quad S_4 = 0.04$$

$$\varepsilon_t \sim N(0, 0.1^2)$$

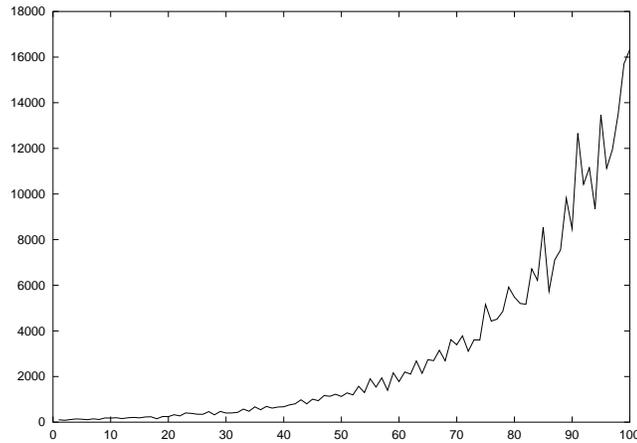


Figure 2.2: Premier modèle multiplicatif

*Est-ce que la variation accidentelle est proportionnelle ou non à la tendance ?* Il est difficile de donner une réponse à cette question uniquement à partir d'une simple représentation graphique. En effet on distingue entre les deux modèles multiplicatifs suivant les besoins du modèle et de la technique de description employée. Comment alors définir un modèle ?

1. Graphique pour déterminer les facteurs explicatifs de son évolution.
2. Méthode numérique (chapitres suivants)

Ces deux méthodes sont deux approches complémentaires.

## 2.2 Modèles auto-projectifs

Dans un modèle auto-projectif, on suppose que  $X_t$  est une fonction des ses valeurs passées et d'une perturbation aléatoire  $\varepsilon_t$

$$X_t = f(X_{t-1}, X_{t-2}, \dots, \varepsilon_t)$$

Une classe de tels modèles, particulièrement utiles pour la prévision, sont les modèles ARIMA. Comme on verra, cette classe permet d'atteindre à l'aide d'un nombre de paramètres relativement limité une gamme de modèles très variés. En outre, il est possible de proposer des méthodes dites méthodes d'identification, permettant de choisir dans cet ensemble de modèles celui qui semble le mieux adapté aux données dont on dispose. Une fois ce modèle choisi, on peut estimer les paramètres et déterminer les prévisions optimales.

Parmi les méthodes de prévision ainsi mises en évidence on trouve, en particulier, les méthodes de lissage exponentiel, l'avantage essentiel de la méthode (*identification, estimation, prévision*) fondée sur les modèles ARIMA, appelée aussi méthode de BOX et Jenkins, est qu'elle permet de sélectionner la méthode de prévision optimale dans un vaste ensemble possibilités. tandis que les méthodes classiques (ajustement et lissage) comportent une part plus grande d'arbitraire dans le choix de la fonction utilisée pour l'ajustement.

## 2.3 Modèles explicatifs

Dans cette dernière catégorie de modèles la variable  $X_t$  est exprimée en fonction d'un vecteur de variables observable, dite exogènes,  $Y_t$  et d'une perturbation aléatoire  $\varepsilon_t$ :

$$X_y = f(Y_t, \varepsilon_t)$$

Où  $Y_t$  est soit déterministe, soit aléatoire; dans ce dernier cas, les processus  $\{Y_t\}$  et  $\{u_t\}$  ont certaines propriétés d'indépendance ou de non corrélation.

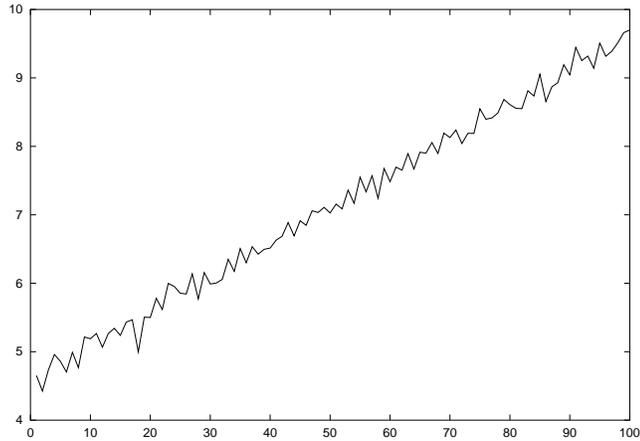


Figure 2.3: Log du premier modèle multiplicatif

Ces modèles sont les méthodes de base de l'économétrie et nous les considérons essentiellement pour faire le lien entre eux et les modèles auto-projectifs.

### 2.3.1 Modèle explicatif statique

Comme son nom indique, dans un modèle explicatif statique les variables  $Y_t$  ne contiennent pas de valeurs passées de  $X_t$  et les  $\varepsilon_t$  sont indépendants entre eux. Par exemple

$$X_t = a + bY_t + \varepsilon_t, \quad t = 1, \dots, T$$

où les  $Y_t$  sont indépendantes de tous les  $\varepsilon_t$  et où les  $\varepsilon_t$  sont centrés et indépendants entre eux.

### 2.3.2 Modèle explicatif dynamique

Un modèle explicatif peut être dynamique soit parce que les  $\varepsilon_t$  sont auto-corrélées, soit parce que  $Y_t$  contient des variables passées de  $X_t$ , c'est à dire des variables dite *endogènes retardées*. On distingue donc deux classes différentes

- Perturbation auto-corrélées
- Variables endogènes retardées

### 2.3.3 TP: simulation des chroniques

Dans cette section, on simule quelques séries chronologiques en utilisant le langage C++.

#### TP: simulation

Les séries temporelles peuvent être simulées par ordinateur.

**Exercice 2.3.1** Soient  $X \sim U[0, 1]$  et  $Y \sim U[0, 1]$  deux variables aléatoires indépendantes. Calculer la loi de la variable aléatoire  $Z = X - Y$ . Montrer qu'il s'agit d'une v. a. centrée.

**Exercice 2.3.2** Variable aléatoire  $N(0, 1)$

Soit  $X_1, X_2, \dots, X_n, \dots$  des variables aléatoires i.i.d. et  $X_k \sim U(0, 1)$  (loi uniforme sur  $(0, 1)$ ). Définissons

$$Y_{2n-1} = (-2 \ln X_{2n-1})^{\frac{1}{2}} \cos 2\pi X_{2n},$$

$$Y_{2n} = (-2 \ln X_{2n-1})^{\frac{1}{2}} \sin 2\pi X_{2n}.$$

Montrer que  $Y_1, Y_2, \dots, Y_n, \dots$  sont i.i.d. et  $Y_k \sim N(0, 1)$ .

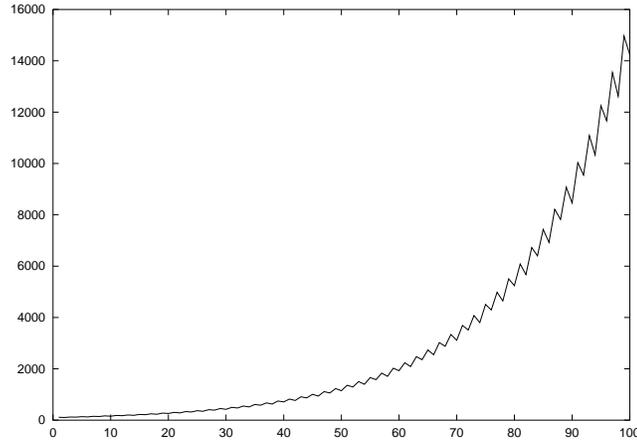


Figure 2.4: Deuxième modèle multiplicatif

Solution : Le résultat de cette exercice nous donne une méthode pour simuler les suites des variables aléatoires indépendantes gaussiennes  $N(0, 1)$ . Il est donc très utile pour la simulation numérique. On cherche donc à montrer : Soit  $X$  et  $Y$  indépendantes de loi  $U[0, 1]$ , Alors le couple  $(U, V)$  fabriqué par

$$U = (-2 \ln X)^{\frac{1}{2}} \cos 2\pi Y,$$

$$V = (-2 \ln X)^{\frac{1}{2}} \sin 2\pi Y.$$

est indépendantes, et suivent la même loi  $N(0, 1)$  PROOF. On calcule d'abord la loi de  $(U, V)$ . Soit  $h(u, v)$  une fonction mesurable. On introduit un changement de variable

$$\begin{cases} u = (-2 \ln x)^{\frac{1}{2}} \cos 2\pi y, \\ v = (-2 \ln x)^{\frac{1}{2}} \sin 2\pi y. \end{cases} \iff \begin{cases} x = \exp\left\{-\frac{u^2 + v^2}{2}\right\} \\ y = \frac{1}{2\pi} \arctan\left(\frac{v}{u}\right) \end{cases}$$

On calcule Jacobien

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} \end{vmatrix} = \begin{vmatrix} -u \exp\left\{-\frac{u^2 + v^2}{2}\right\} & -v \exp\left\{-\frac{u^2 + v^2}{2}\right\} \\ -\frac{1}{2\pi} \frac{v}{u^2 + v^2} & \frac{1}{2\pi} \frac{u}{u^2 + v^2} \end{vmatrix} = -\frac{1}{2\pi} \exp\left\{-\frac{u^2 + v^2}{2}\right\}$$

Donc

$$\begin{aligned} \mathbf{E} h(U, V) &= \mathbf{E} h\left((-2 \ln X)^{\frac{1}{2}} \cos 2\pi Y, (-2 \ln X)^{\frac{1}{2}} \sin 2\pi Y\right) \\ &= \int_0^1 \int_0^1 h\left((-2 \ln x)^{\frac{1}{2}} \cos 2\pi y, (-2 \ln x)^{\frac{1}{2}} \sin 2\pi y\right) dx dy \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} h(u, v) \frac{1}{2\pi} \exp\left\{-\frac{u^2 + v^2}{2}\right\} du dv \end{aligned}$$

Donc

$$\begin{aligned} p(u, v) &= \frac{1}{2\pi} \exp\left\{-\frac{u^2 + v^2}{2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\}\right) \times \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{v^2}{2}\right\}\right) \end{aligned}$$

On vérifie facilement que  $\left(\frac{1}{\sqrt{2\pi}} \exp\{-\frac{u^2}{2}\}\right)$  et  $\left(\frac{1}{\sqrt{2\pi}} \exp\{-\frac{v^2}{2}\}\right)$  sont respectivement la densité de  $U$  et  $V$ . En prenant dans la démonstration une fonction  $h(u, v)$  qui dépend uniquement de  $u$ . En effet

$$\begin{aligned} \mathbf{E} h(U) &= \mathbf{E} h\left((-2 \ln X)^{\frac{1}{2}} \cos 2\pi Y\right) \\ &= \int_0^1 \int_0^1 h\left((-2 \ln x)^{\frac{1}{2}} \cos 2\pi y\right) dx dy \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} h(u) \frac{1}{2\pi} \exp\left\{-\frac{u^2 + v^2}{2}\right\} dudv \\ &= \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{v^2}{2}\right\} dv \int_{\mathbf{R}} h(u) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du \\ &= \int_{\mathbf{R}} h(u) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du \end{aligned}$$

**Exercice 2.3.3** TP : simulation des chroniques. On peut simuler à l'aide de ordinateur les série suivantes

$$X_t = 0.05t + 100 + S_t + \varepsilon_t \text{ où } S_t = \begin{cases} -1 \\ 2 \\ -2 \\ 1 \end{cases} \quad \varepsilon_t \sim N(0, 0.5^2) \quad (2.1)$$

$$X_t = e^{a+bt}(1 + S_t)(1 + R_t) \text{ où } S_t = \begin{cases} 0.3 \\ -0.3 \\ 0.2 \\ -0.2 \end{cases} \quad R_t \sim N(0, 20^2) \quad (2.2)$$

avec  $a = \ln 100$  et  $b = 0.05$

Tracer  $\ln X_t$ .

$$X_t = (0.05t + 100)(1 + S_t) + \varepsilon_t \text{ où } S_t = \begin{cases} -1 \\ 2 \\ -2 \\ 1 \end{cases} \quad \varepsilon_t \sim N(0, 1) \quad (2.3)$$

### 2.3.4 Implémentation

On construit d'abord une classe BBG:

```
/* ----- BBG.h -----*/
#ifndef BBG_H_IS_DEFINE
#define BBG_H_IS_DEFINE
class BBG {
    double W1, W2;
    int seesaw;
public:
    double gaussien();
    BBG(int);
};
#endif
/* ----- fin BBG.h -----*/
```

Voici les membres fonctions pour BBG.h

```
/* ----- BBG.C -----*/
#include <stdlib.h> /* drand48(), srand48(), sin(), cos() */
#include <math.h> /* M_PI, log() */
#include <iostream.h> /* cout */
#include "BBG.h"

BBG::BBG(int seed) {
    seesaw = 0;
    srand48(seed);
}

double BBG::gaussien()
{
    double c1,c2;

    if (seesaw==0)
    {
        c1=2.0*M_PI*drand48();
        c2=-2.0*log(drand48());
        c2=sqrt(c2);
        W1 = c2*cos(c1);
        W2 = c2*sin(c1);
        seesaw=1;
        // cout <<"seesaw = "<<seesaw<<" "<<W1<<"\n ";
        return(W1);
    }
    else
    {
        seesaw=0;
        // cout <<"seesaw = "<<seesaw<<" "<<W2<<" \n ";
        return(W2);
    }
}
/* ---- fin BBG.C -----*/
```

Pour tester le générateur de nombre v.a.  $N(0,1)$ , on utilise le subroutine testBBG.C

```
/* ---- testBBG.C ----- */
#include<iostream.h>
#include"BBG.h"

void testBBG(int seed, int n) {
    BBG a(seed);
    double sum =0.0;
    double *X = new double[n];
    for(int i = 0; i<n; ++i) {
```

```

    X[i] = a.gaussien();
    sum += X[i];
}
sum = sum/n;
double var = 0.0;
for(int i = 0; i<n; ++i)
    var += (X[i] - sum)*(X[i] - sum);
var = var/(n-1);
cout << " sum = " << sum/n << "\n";
cout << " var = " << var << "\n";
}
/* ----- Fin testBBG.C ----- */

```

En C ou C++, tous les applications contiennent un seul main.C

```

/* ----- main.C ----- */
#include<math.h>          /* M_PI */
#include<stdlib.h>
/* #include<unistd.h> */ /* getpid */
#include<iostream.h>
#include"BBG.h"

extern void testBBG(int,int);

main()
{
    int seed=getpid();
    // int seed=100;
    /* srand48(getpid()); */
    int n = 100;
    testBBG(seed,n);
}

```

## Remarques

1. Dans les fichiers \*.h, on peut uniquement mettre

Type de finition	struct BBG {int seed; };	
Template déclarations	template<class T> class Z;	
Fonction déclarations	extern int strlen(const char*);	
Inline fonction définition	inline double dmax(double x, double y) {...};	
Data déclaration	extern int a;	
constant définition	const double pi = 3.14159;	
Enumeration	enum coulour {red, yellow, green};	Il faut jamais met-
Name déclaration	class Matrix;	
Include	#include <std.h>	
Macros définition	#define VERSION 12	
conditional compilation directives	#ifdef _cplusplus	
comments	/* tres important */	

tre  
 Ordinary function definition double dmax (doubé x, double y)..  
 Data definition int a;

2. Il faut connaître les options standartes de g++, par exemple -o -g -Wall -C -O -E lire attentivement le manuelle man g++
3. Un bon *Makefile* est très important aussi, voir le fichier joint.
4. Comprendre les options -I -L

Voici le fichier Makefile

```

# ----- ZHANG Huilong Makefile for C++
#
# ----- C++ compiler [g++] options on Unix (Sparc Linux HP)
#

```

```

#         -g : produces additional symbol table information for debugging
#         -c : suppresses loading and produces .o files
#         -C : checks that subscripts are within declared array bounds
#         -o : names the final output file
#         -O : optimizes
#         -S : stops after the stage of compilation proper
#         -p : prepares object files for profiling (prof)
#         -pg : prepares object files for profiling (gprof)
#         -Wall : issues warnings for conditions

# ----- directories

# ----- libraries

LIBMATH = -lm

# ----- include
INC = -I
# ----- macros definition
SRC =

#
# ----- object files
OBJS      = main.o\
           BBG.o \
           testBBG.o \
# ----- compilateur
COMPG++ = g++
COMPGCC = gcc
# ----- options
RULE = -o -g $@ $?
CLEAN = @ /bin/rm $?
MESS = @ echo "compilation ... $@ done
#     OPT = -g
#     OPT = -g -Wall
#     OPT = -O
# -----
# OPT = -g compiler avec cette option vous permet de debugger
# a l'aide de xgdb ou gdb
# OPT = -O programme compile avec cette option sera optimise, et
# tourne plus vite, mais on ne pourra pas debugger.
# -----

essai : $(OBJS)
        $(COMPG++) $(OPT) $(OBJS) -o essai -lm
        $(MESS)

clean:
        @rm -f $(OBJS)
# ----- fin Makefile for C++

```

## 2.4 Recherche et estimation de la tendance

Le mot *lissage* : exprime le mouvement tendanciel d'une succession de points  $(t, Z_t)$  situés au milieu d'un nuage de points  $(t, X_t)$ .

Si on désire exprimer la tendance avec une courbe très lisse, une méthode consiste à proposer une expression analytique. on calcule souvent cette expression à partir de données lissées jugée convenable (non à partir des données initiales). Alors c'est quoi une *bonne* série lissée ?

Il est difficile de définir. les motivations sont souvent très variées. Il ne s'agit pas le plus souvent d'atteindre le cas limite : la droite, ni de s'ajuster au point de passer chaque point observé. Lisser le moins fort possible tout en atteignant un objectif précis

- Se débarrasser des résidus.
- Se débarrasser des résidus et du saisonnier

de telle sorte que ces résidus (ou résidus + saisonnière) fluctuent autour de la courbe lisse.

### 2.4.1 Moyennes par intervalles

Pour simplifier la présentation, on considère le problème sans variations saisonnières

$$X_t = T_t + \varepsilon_t$$

Le but consiste à mettre en évidence la tendance  $T_t$ . La méthodes par intervalle est utilisée pour résumer les séries chronologiques comportant un grand nombre d'observations. La démarche consiste à approcher la chronique par une fonction en escalier ou par une fonction affine par morceaux. Plus précisément, soient  $I_1, I_2, \dots, I_j, \dots$  intervalles de temps (*les semaines, les mois, les années, etc*) on calcule les moyennes des observations appartenant à ces intervalles pour obtenir la série  $m_j$ , où  $j$  est une indice centrée, pas toujours une indice coïncidant avec  $t$  dans  $X_t$ . Cette méthode très simple présente certains avantages :

- Mettre en évidence la tendance de la série
- La variance de la variation accidentelle est divisée par le nombre de terms considérés
- Les valeurs successives sont indépendantes si les  $\varepsilon_t$  le sont.

l'inconvénient de cette méthode est la diminution importante du nombre de valeurs, elle est appliquée uniquement dans des séries présentant un nombre important d'observations.

**exemple 2.4.1** On considère la série suivante. La figure 2.4.1 compare sa simulation et la moyenne par intervalle, la longueur de la moyenne est de 4.

$$T_t = 0.05t + 100$$

$$S_1 = -1, \quad S_2 = 2, \quad S_3 = -2, \quad S_4 = 1$$

$$\varepsilon_t \sim N(0, 0.5^2)$$

### 2.4.2 Moyennes mobiles

En pratique, le nombre d'observations est insuffisant pour que l'on puisse se satisfaire de moyennes par intervalle. Deux cas se présentent

- le nombre  $q$  d'observations considérées est impaire

$$m_t = \frac{1}{q} \sum_{i=-k}^{i=+k} X_{t+i}, \quad k \text{ vérifie } q = 2k + 1$$

Par exemple

$$m_t = \frac{1}{5}(X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2})$$

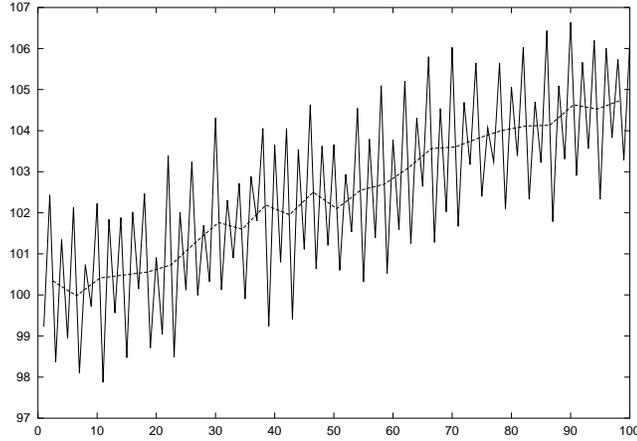


Figure 2.5: Moyenne par intervalle

- Le nombre  $q$  d'observations considérées est paire, la formule est

$$m_t = \frac{1}{q} \left[ \sum_{i=-k+1}^{i=k-1} X_{t+i} + \frac{(X_{t-k} + X_{t+k})}{2} \right] \quad k \text{ vérifie } q = 2k$$

Par exemple, soit  $q = 4$

$$m_t = \frac{1}{4} \left( X_{t-1} + X_t + X_{t+1} + \frac{1}{2}(X_{t-2} + X_{t+2}) \right)$$

On remarque facilement que dans l'exemple de *Fuel consommation*, chapitre 1, table 1.2, on a utilisé cette moyenne mobile avec  $q = 4$ .

**Remarque 2.4.2** *On note que*

1. Il s'agit d'une moyenne pondérée.
2. Pour une suite d'observations  $(t = 1, \dots, T)$ , les instants où on peut calculer  $m_t$  sont  $(k + 1, k + 2, \dots, T - k)$ , soit seulement  $t - 2k$  instants.
3. Non indépendance entre les  $m_t$ . Par exemple  $m_t$  et  $m_{t+1}$  ne sont pas indépendants, car

$$\begin{aligned} X_{t-k}, \quad \dots \quad X_{t+k} &\implies m_t \\ X_{t-k+1}, \quad \dots \quad X_{t+k+1} &\implies m_{t+1} \end{aligned}$$

$X_{t-k+1}, \dots, X_{t+k}$  sont prises en compte à la fois dans le calcul de  $m_t$  et dans le calcul de  $m_{t+1}$ .

Nous étudions dans la suite de cette section l'effet du filtre de moyenne mobile sur la tendance d'une chronique.

**Chronique de modèle**  $X_t = a + \varepsilon_t$

Il s'agit d'un modèle d'ajustement simplifié : *constant + perturbation aléatoire*. On suppose  $\mathbf{E} \varepsilon_t = 0$ ,  $\mathbf{Var} \varepsilon_t = \sigma^2$  et les  $\{\varepsilon_t\}$  sont indépendantes identiques distribuées (*Bruit Blanc*). Soit  $q = 2k + 1$  on a alors  $\forall h \geq 1$

$$\begin{aligned} m_t &= \frac{1}{q} \sum_{i=-k}^{i=k} \varepsilon_{t+i} + a \\ m_{t+h} &= \frac{1}{q} \sum_{i=-k}^{i=k} \varepsilon_{t+h+i} + a \end{aligned}$$

**Proposition 2.4.3** La Série  $m_t$  ne possède pas de tendance et elle est de variance  $\frac{\sigma^2}{q}$

Preuve:  $\mathbf{E} m_t = a$ ,  $\mathbf{Var} m_t = \frac{\sigma^2}{q}$

**Remarque 2.4.4** On remarque que

1. On dit que la série ne possède pas de composante tendancielle (où que la tendance est nulle) dans la mesure où sa moyenne est constante.
2. La variance est divisée par  $q$  par rapport à la série initiale, les  $m_t$  sont soumises à une variation accidentelle d'amplitude beaucoup plus petite (lisse).

**Chronique de modèle**  $X_t = at + b + \varepsilon_t$

Ce modèle fait partie aussi des chroniques à tendance polynomiale. On suppose que la longueur de moyenne mobile  $q$  est impaire

$$\begin{aligned} m_t &= \frac{1}{q} \sum_{i=-k}^{i=k} [a(t+i) + b + \varepsilon_{t+i}] = \frac{1}{q} \sum_{i=-k}^{i=k} [at + ai + b + \varepsilon_{t+i}] \\ &= at + b + \frac{1}{q} \left[ a \sum_{i=0}^{i=k} i + a \sum_{i=-k}^0 i + \sum_{i=-k}^{i=k} \varepsilon_{t+i} \right] \\ &= at + b + \frac{1}{q} \sum_{i=-k}^{i=k} \varepsilon_{t+i} \end{aligned}$$

**Proposition 2.4.5** Si  $X_t$  est une chronique dont la composante tendancielle est linéaire, alors la série  $m_t$  des moyennes mobiles concervent la même tendance et a pour composante accidentelle la moyenne mobile des composantes accidentelle de la série initiale.

On admet (en exercice) le même résultat pour une moyenne mobile de longueur  $q =$  paire.

**Chronique de modèle**  $X_t = at^2 + bt + c + \varepsilon_t$

Nous pouvons nous limiter à l'étude du terme de seconde degré. (Pourquoi ? A réfléchir...). Soit\*  $q = 2k + 1$ , impaire

$$\begin{aligned} m_t &= \frac{1}{q} \sum_{i=-k}^{i=k} [a(t+i)^2] = \frac{1}{q} \sum_{i=-k}^{i=k} [a(t^2 + 2it + i^2)] \\ &= at^2 + \frac{a}{q} \left[ 2t \sum_{i=-k}^k i + \sum_{i=-k}^k i^2 \right] = at^2 + \frac{a}{q} 2 \sum_{i=1}^k i^2 \\ &= at^2 + a \frac{k(k+1)}{3} \end{aligned}$$

**Proposition 2.4.6** Soit  $X_t$  une chronique dont la composante tendancielle est un polynôme de 2ième degré  $at^2 + bt + c$ , alors la composante tendancielle de la série des moyennes mobiles est égale à

$$T_t = at^2 + bt + c + \frac{ak(k+1)}{3}$$

**Remarque 2.4.7** Il est évident que

1. l'estimateur de la tendance par le filtre des moyennes mobiles est biaisé, le biais est constant.
2. On montre que dans le cas d'une tendance du 3ième degré, c'est un biais linéaire en  $t$  et dans le cas d'une tendance du 4ième degré, c'est un biais en  $t^2$ , etc.

Il faut donc se montrer prudent dans l'utilisation des moyennes mobiles et ramener le problème autant que possible à une fonction linéaire en  $t$ . La figure 2.6 illustre un exemple d'une série à tendance polynomiale de 2ième degré et ces moyennes mobiles.

---

\*  $\sum_{i=1}^k i^2 = \frac{k(k+1)(2k+1)}{6}$

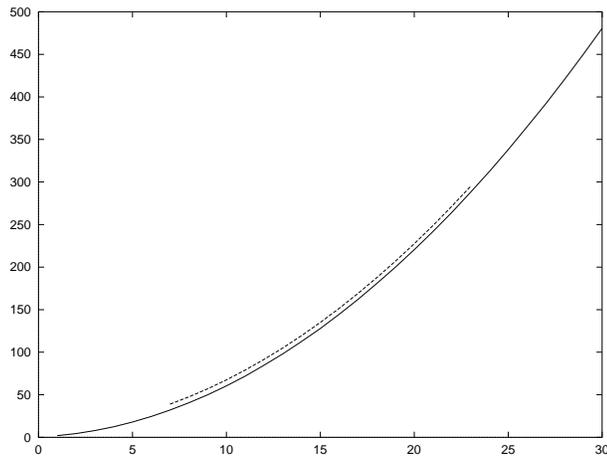


Figure 2.6: Moyenne mobile de longueur 12 pour une chronique de degré 2

### Présence de variations saisonnières

Soit

$$X_t = T_t + S_t + \varepsilon_t$$

avec  $S_{t+p} = S_t \quad \forall t$  où  $p$  est la période de  $S_t$ . Nous nous limitons à l'étude de la série  $S_t$ , car nous savons que la moyenne mobile d'une somme est la somme des moyennes mobiles. On considère une moyenne mobile de longueur  $q = 2k + 1$ .

$$m_t = \frac{1}{q} \sum_{i=-k}^k S_{t+i}$$

**Remarque 2.4.8** *On remarque que*

1. *Il est donc évident que si la longueur de la moyenne mobile  $q$  et la période de variation saisonnière sont égales ( $p = q = 2k + 1 = p$ ) alors la moyenne mobile calculée sur  $S_t$  est constante.*
2. *Il est évident également que la moyenne mobile est constante si sa longueur est un multiple de la période  $q = rp$*

**Proposition 2.4.9** *Dans le cadre du modèle additif, le filtre de la moyenne mobile de longueur  $q$  élimine les variations saisonnières dont la période  $p$  divise  $q$ .*

En figure 2.7, nous donnons la représentation simultanée d'une série et sa moyenne mobile de longueur 4.

### Quelques remarques supplémentaires sur les moyennes mobiles

Il existe d'autres moyennes mobiles plus générales, non symétriques, elles sont utilisées pour la prévision à court terme pour conserver des tendances ou pour éliminer des variations saisonnières particulières. Par exemple, les moyennes de Spencer

- Moyennes de Spencer de longueur 15

$$\frac{1}{320}(-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3)$$

Réduction de variance résiduelle est de 0.19.

- Moyennes de Spencer de longueur 21

$$\frac{1}{350}(-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, 60, 57, \dots, -1)$$

Réduction de variance résiduelle est de 0.15.

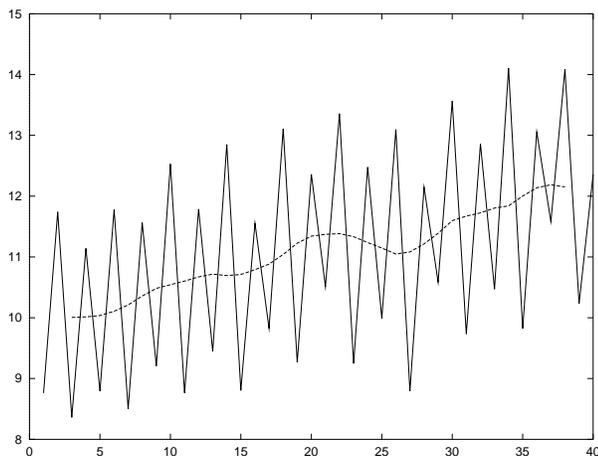


Figure 2.7: Moyenne mobile de longueur 4 pour une chronique de période 4

Les moyennes Spencer annulent les saisonnalités de période 4 en conservant les tendances polynomiales de degré  $\leq 3$  (sans biais).

**Exercice 2.4.10** Soit la moyenne

$$m_t = \frac{1}{4}(X_{t-1} + X_t + X_{t+1} + \frac{1}{2}(X_{t-2} + X_{t+2}))$$

Montrer que cette moyenne mobile élimine les variations saisonnières de période 4.

*Solution*  $q = 4$  paire. On sait que  $S_{t-2} = S_1, S_{t-1} = S_2, S_t = S_3, S_{t+1} = S_4, S_{t+2} = S_5 = S_1$ .

$$\begin{aligned} & \frac{1}{4} \left[ S_{t-1} + S_t + S_{t+1} + \frac{1}{2}(S_{t-2} + S_{t+2}) \right] \\ &= \frac{1}{4} \left[ S_2 + S_3 + S_4 + \frac{1}{2}(S_1 + S_1) \right] = \frac{1}{4}(S_1 + S_2 + S_3 + S_4) = 0 \end{aligned}$$

**Exercice 2.4.11** Soit  $S_t$  est évolutive linéaire de période 4. Montrer que la moyenne mobile pondérée

$$m_t = \frac{1}{16} (X_{t-3} + 2X_{t-2} + 3X_{t-1} + 4X_t + 3X_{t+1} + 2X_{t+2} + X_{t+3})$$

élimine la variation saisonnière

*Solution* Soit  $S_1, S_2, S_3, S_4$  coefficients saisonniers. On pose

$$S_{t-3} = S_1, S_{t-2} = S_2, S_{t-1} = S_3, S_t = S_4$$

$$S_{t+1} = S_1, S_{t+2} = S_2, S_{t+3} = S_3$$

On a alors

$$\begin{aligned} & \frac{1}{16} (S_{t-3} + 2S_{t-2} + 3S_{t-1} + 4S_t + 3S_{t+1} + 2S_{t+2} + S_{t+3}) \\ &= \frac{1}{16} (S_1 + 2S_2 + 3S_3 + 4S_4 + 3S_1 + 2S_2 + S_3) = \frac{1}{16} (4S_1 + 4S_2 + 4S_3 + 4S_4) = 0 \end{aligned}$$

### 2.4.3 Différences successives

Les différences successives sont utilisées pour étudier la tendance et pour obtenir une chronique stationnaire (la notion *stationnaire* sera étudiée dans les chapitres suivants).

### Définition

Les différences successives d'une chronique sont obtenues par transformation linéaire des observations initiales.

$$DX_t = X_{t+1} - X_t, \quad D^2X_t = DX_{t+1} - DX_t = D(DX_t)$$

d'une façon générale

$$D^k X_t = D^{k-1} X_{t+1} - D^{k-1} X_t = D(D^{k-1} X_t)$$

Donc la série  $D^k X_t$  est définie de  $t = 1$  à  $t = T - k$ . On note que le filtre des différences successives conserve la périodicité des variations saisonnière éventuelles, car  $S_{t+1} - S_t$  est périodique et de période  $p$ . On remarque également.

$$DX_t = X_{t+1} - X_t = T_{t+1} - T_t + S_{t+1} - S_t + \varepsilon_{t+1} - \varepsilon_t$$

C'est une série de modèle additif. On étudie dans la suite l'effet des différences successives sur les modèles de chronique

### Effet sur bruit blanc

Soit  $X_t = \varepsilon_t$  alors  $DX_t = \varepsilon_{t+1} - \varepsilon_t$ , donc le filtre linéaires des différences premières n'est pas un lissage puisque la variance de  $DX_t$  est multipliée par 2. On obtient une suite de v.a. non indépendantes ( $DX_t$  et  $DX_{t+1}$  sont dépendantes). En figure 2.8 illustre les différences successives d'un bruit blanc. On remarque l'amplitude des variations augmente avec l'ordre des différences successives.

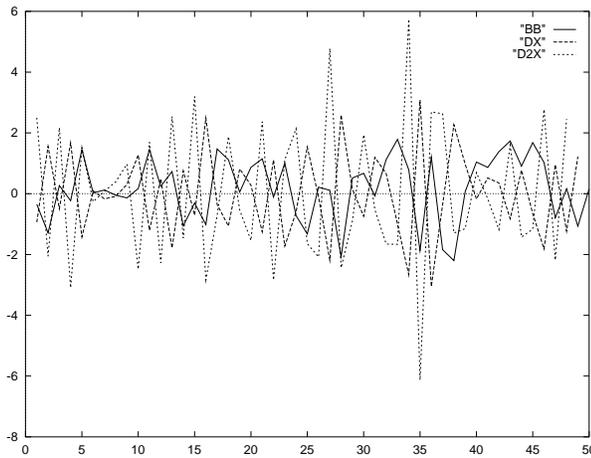


Figure 2.8: Représentation simultanée d'un BBG,  $DX_t$  et  $D^2X_t$

**Proposition 2.4.12** *La série des différences premières d'une chronique de modèle  $X_t = a + \varepsilon_t$  où  $\varepsilon_t$  est un bruit blanc est une suite de v. a. de moyenne nulle, de variance  $2\sigma^2$ .*

### Effet sur le modèle polynomial

On considère une chronique de modèle  $X_t = P(t) + \varepsilon_t$  où  $P(t)$  est un polynôme en  $t$  de degré  $n$ .  $\varepsilon_t$  un bruit blanc de variance  $\sigma^2$ .

- Soit  $X_t = at + b + \varepsilon_t$  alors  $DX_t = X_{t+1} - X_t = a + \varepsilon_{t+1} - \varepsilon_t$  On a donc

$$\mathbf{E} DX_t = a, \quad \mathbf{Var} DX_t = 2\sigma^2$$

c'est à dire  $DX_t$  est une suite de v. a. de moyennes  $a$  et de variance  $2\sigma^2$ .

- Soit  $X_t = at^2 + bt + c + \varepsilon_t$  alors

$$DX_t = 2at + a + b + \varepsilon_{t+1} - \varepsilon_t$$

chronique composée d'une tendance linéaire et d'une variation accidentelle.

$$\begin{aligned} D^2 X_t &= 2a + (\varepsilon_{t+2} - \varepsilon_{t+1}) - (\varepsilon_{t+1} - \varepsilon_t) \\ &= 2a + \varepsilon_{t+2} - 2\varepsilon_{t+1} + \varepsilon_t \end{aligned}$$

chronique composée d'une tendance constante et d'une variation accidentelle. On calcule facilement que

$$\mathbf{E} D^2 X_t = 2a \quad \mathbf{Var} D^2 X_t = 6\sigma^2$$

La figure 2.9 donne la représentation d'une chronique simulée dont la tendance est un polynôme de degré 2, le  $DX_t$  et  $D^2 X_t$ . Les différences premières présentent une tendance linéaire et les différences secondes une tendance constante, on constate également que l'amplitude des variations augmente avec l'ordre des différences.

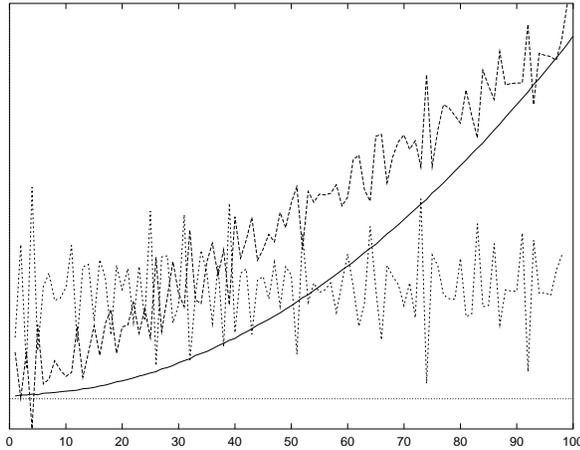


Figure 2.9: Représentation simultanée d'une chronique (échelles différentes), d'une chronique de terme  $X_t = at^2 + bt + c + \varepsilon_t$ ,  $DX_t$  et  $D^2 X_t$

**Proposition 2.4.13** Soit  $X_t$  une chronique dont le terme général de la tendance est défini par un polynôme de degré  $n$ . Alors la série des différences  $D^n X_t$  est de la forme  $a + \varepsilon_t$ , où  $\varepsilon_t$  est une suite de v. a. de moyenne nulle et de variance constante par rapport à  $t$ .

## 2.4.4 Autocorrélation et correlogramme

Nous avons vu précédemment que les moyennes mobiles d'une chronique sont corrélées.

### Définition et propriété

**Définition 2.4.14** On appelle coefficient d'autocorrélation de rang  $k$  de la chronique  $X_t$ ,  $t = 1, \dots, T$  le coefficient  $r_k$  :

$$r_k = \frac{\sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2} \quad \bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$$

<sup>†</sup>rappel: Soient  $\xi, \eta$  deux v. a. alors le coefficient de corrélation est défini par

$$r = \frac{\mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)}{\sqrt{\mathbf{Var}\xi}\sqrt{\mathbf{Var}\eta}}$$

ce coefficient peut être estimé par

$$\bar{r} = \frac{\sum_{i=1}^n (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\sqrt{\sum_{i=1}^n (\xi_i - \bar{\xi})^2} \sqrt{\sum_{i=1}^n (\eta_i - \bar{\eta})^2}}$$

où  $\xi_i$  et  $\eta_i$  sont deux échantillons de taille  $n$ . Il y a donc une légère différence avec la définition 2.4.14, ce choix est dicté par le fait que  $\varepsilon_t$  est de moyenne  $\mu$  et de variance  $\sigma^2$ , indépendantes de  $t$ .

**Définition 2.4.15** On appelle *corrélogramme* la représentation graphique de l'application:  $k \rightarrow r_k$ .

Le coefficient  $r_k$  est défini pour mesurer la dépendance entre  $X_t$  et  $X_{t+k}$ .

**Théorème 2.4.16** On considère un bruit blanc  $\varepsilon_t$ . les coefficient d'autocorrélation  $r_k$  de rang supérieur ou égale à 1 suivent approximativement une loi normale de moyenne  $-\frac{1}{T}$  et de variance  $\frac{1}{T}$ .

ce théorème permet de juger de la taille des coefficients d'autocorrélation: en négligeant la moyenne  $-\frac{1}{T}$ , on peut dire que la probabilité qu'un coefficient soit supérieur en valeur absolue à  $\frac{1.96}{\sqrt{T}}$  est égale à 0.05.

$$P(|r_k| \geq \frac{1.96}{\sqrt{T}}) = 0.05$$

Pour illustrer le théorème, nous avons simulé un bruit blanc de 500 termes dont les coefficients d'autocorrélation sont présentés dans la table 2.1. (Voir le corrélogramme Fig. 2.10.)

$r_1 = -0.026720$	$r_2 = 0.052195$	$r_3 = -0.069710$	$r_4 = -0.097432$
$r_5 = 0.042616$	$r_6 = -0.021875$	$r_7 = 0.044295$	$r_8 = -0.023381$
$r_9 = 0.043714$	$r_{10} = -0.020600$	$r_{11} = -0.018972$	$r_{12} = -0.000013$
$r_{13} = -0.006135$	$r_{14} = 0.022235$	$r_{15} = 0.005640$	$r_{16} = -0.047127$
$r_{17} = 0.013550$	$r_{18} = -0.053561$	$r_{19} = 0.005232$	$r_{20} = 0.034764$

Table 2.1: coefficient d'autocorrélation  $k = 1, \dots, 20$  sur une série BBG

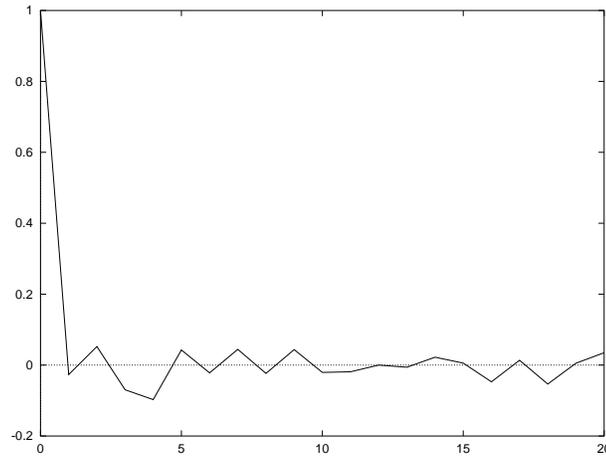


Figure 2.10: corrélogramme d'un bruit blanc simulé

Analyse des variation accidentelle obtenues par l'ajustement d'un modèle à des données observées est fondamentale puisqu'elle permet de vérifier l'adéquation du modèle choisi. Cette vérification est faite en deux étapes:

1. Calcul des résidus, définis par la différence entre les observations et les valeurs estimées, forment un bruit blanc (si le modèle est correct), à partir duquel on peut plus tirer d'information.
2. Calcul du corrélogramme, et vérifier l'hypothèse selon la quelle il s'agit bien un bruit blanc avec le théorème 2.4.16

### Corrélogramme de chronique de tendance non nulle

Le corrélogramme de chronique de tendance non nulle est en général caractéristique : les coefficients d'autocorrélation restent relativement élevés jusqu'à un rang important si la série est monotone. Lorsque la tendance change de sens au cours de la période étudiée, certains coefficients d'autocorrélation prennent des valeurs négatives, ceci correspond aux changement de régime. (ruptur de comportement). On peut simuler un exemple en prenant deux tendances différentes (monotone croissant suivi par monotone décroissant figure 2.11) en milieu de la série et tracer le corrélogramme associé, figure 2.12.

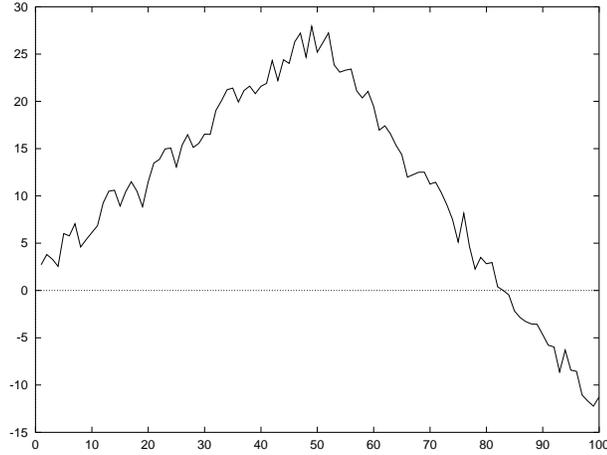


Figure 2.11: changement de régime : trajectoir

### Autocorrélation et filtre des moyennes mobiles

Le filtre des moyennes mobiles transforme la chronique en une serie fortement corrélée. Considérons les moyenne mobiles de longueur impaire ( $q = 2k + 1$ )  $m_t$  et  $m_{t+h}$ , appliquées sur un bruit blanc ,

$$m_t = \frac{1}{q} \sum_{i=-k}^k \varepsilon_{t+i} \quad m_{t+h} = \frac{1}{q} \sum_{j=-k}^k \varepsilon_{t+h+j}$$

On étudie maintenant les coefficients de corrélation théoriques entre  $m_t$  et  $m_{t+h}$ .

$$\mathbf{Cov}(m_t, m_{t+h}) = \left(\frac{1}{q}\right)^2 \sum_{i=-k}^k \sum_{j=-k}^k \mathbf{E}(\varepsilon_{t+i}, \varepsilon_{t+h+j})$$

Ces espérance sont tous nulles sauf si les indices sont les même

$$\mathbf{Cov}(m_t, m_{t+h}) = \left(\frac{1}{q}\right)^2 \sum_{i=t+h-k}^{t+k} \sigma^2 = \frac{2k - h + 1}{(2k + 1)^2} \sigma^2$$

où  $q = 2k + 1$  et  $h \leq 2k$ . Donc le coefficient théorique est

$$r_h = \frac{\mathbf{Cov}(m_t, m_{t+h})}{\sqrt{\mathbf{Var} m_t} \sqrt{\mathbf{Var} m_{t+h}}} = \frac{2k - h + 1}{2k + 1} = 1 - \frac{h}{2k + 1}$$

Ce coefficient est d'autant plus grand que la longueur de la moyenne mobile est grande, et d'autant plus petit que le rang  $h$  est grand. pour  $h > 2k$ , il est nul. La figure 2.13 est une simulation. On notera que le coefficient de corrélation est positive ou nul, cela signifie que les séries  $m_t$  et  $m_{t+h}$  varient dans le même sens.

### Autocorrélation et filtre des différences successives

Nous avons vu que la présence d'une tendance croissante ou décroissante dans la série étudiée se traduit par de fortes valeurs des coefficients d'autocorrélation est en lui même intéressant car on cherche souvent à éliminer la tendance de la chronique pour n'en conserver que la part aléatoire. On sait que les différences successives d'ordre  $n$  transforment les chroniques à tendance polynomiale de degré  $n$  en chronique à tendance constante. Au niveau des coefficients d'autocorrélation, l'effet n'est pas négligeable. On étudie dans la suite

### Autocorrélation des différences premières d'un modèle linéaire

Soit  $X_t = at + b + \varepsilon_t$ . On calcule facilement

$$\begin{aligned} DX_t &= a + \varepsilon_{t+1} - \varepsilon_t \\ \mathbf{Cov}(DX_{t+1}, DX_t) &= \mathbf{E}(\varepsilon_{t+2} - \varepsilon_{t+1})(\varepsilon_{t+1} - \varepsilon_t) = -E(\varepsilon_{t+1})^2 \end{aligned}$$

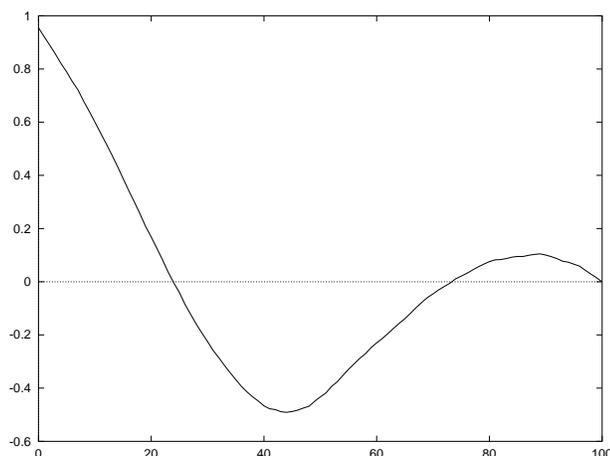


Figure 2.12: changement de régime : corrélogramme

Car  $\varepsilon_t$  est un bruit blanc. On obtient ainsi le coefficient d'autocorrélation d'ordre 1:

$$\begin{aligned}\delta_1 &= \frac{\text{Cov}(DX_{t+1}, DX_t)}{\sqrt{\text{Var} DX_{t+1}} \sqrt{\text{Var} DX_t}} \\ &= \frac{\sigma^2}{2\sigma^2} = -\frac{1}{2}\end{aligned}$$

**Proposition 2.4.17** *le coefficient d'autocorrélation de rang 1 des différences premières d'une chronique à tendance linéaire est égal à  $-0.5$  et les coefficients d'autocorrélation de rang supérieur sont nuls.*

### Autocorrélation des différences successives d'une chronique à tendance polynomiale

Étudions le cas où  $X_t = at^2 + bt + c + \varepsilon_t$ . On sait que les différences secondes sont

$$D^2X_t = 2a + \varepsilon_{t+2} - 2\varepsilon_{t+1} + \varepsilon_t = 2a + \mathcal{E}_t$$

Les  $\mathcal{E}_{t+h}$  et  $\mathcal{E}_t$  sont indépendantes pour  $h \geq 3$ , pour d'autre  $h$  on a

$$\begin{aligned}\mathbf{E}(D^2X_{t+1} - 2a)(D^2X_t - 2a) &= \mathbf{E}(\mathcal{E}_{t+1}\mathcal{E}_t) \\ &= \mathbf{E}(\varepsilon_{t+3} - 2\varepsilon_{t+2} + \varepsilon_{t+1})(\varepsilon_{t+2} - 2\varepsilon_{t+1} + \varepsilon_t) \\ &= -2\mathbf{E}\varepsilon_{t+2}^2 - 2\mathbf{E}\varepsilon_{t+1}^2 = -4\sigma^2\end{aligned}$$

On en déduit que le coefficient de corrélation  $\sigma_1 = -\frac{2}{3}$ . calculons maintenant  $\sigma_2$

$$\begin{aligned}\mathbf{E}(D^2X_{t+2} - 2a)(D^2X_t - 2a) &= \mathbf{E}(\mathcal{E}_{t+2}\mathcal{E}_t) \\ &= \mathbf{E}(\varepsilon_{t+4} - 2\varepsilon_{t+3} + \varepsilon_{t+2})(\varepsilon_{t+2} - 2\varepsilon_{t+1} + \varepsilon_t) \\ &= \mathbf{E}\varepsilon_{t+2}^2 = \sigma^2\end{aligned}$$

donc  $\delta_2 = \frac{1}{6}$

**Proposition 2.4.18** *Le coefficient d'autocorrélation de rang 1 des différences secondes d'une chronique à tendance polynomiale de degré 2 est égal à  $-\frac{2}{3}$  et le coefficient d'autocorrélation de rang 2 à  $\frac{1}{6}$ . Les coefficients d'autocorrélation de rang supérieur sont nuls.*

Voir en figure 2.14 On admet plus généralement

**Proposition 2.4.19** *Les coefficients d'autocorrélation de rang supérieur ou égal à  $n + 1$  des différences successives d'ordre  $n$  sont nuls pour une chronique de tendance polynomiale de degré  $n$ , sans saisonnalité.*

Une façon d'éliminer la tendance d'une chronique est donc de calculer les différences successives jusqu'à l'obtention de coefficients d'autocorrélation diminuant rapidement. On demande comme exercice la simulation et le calcul d'un exemple de  $X_t = at^2 + bt + c + \varepsilon_t$  afin de vérifier cette proposition (figure 2.15). Les coefficients d'autocorrélation

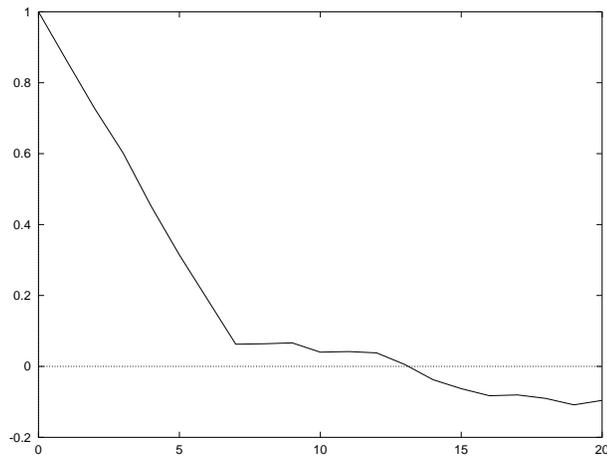


Figure 2.13: corrélogramme des moyennes mobiles de longueur 7 d'un bruit blanc simulé 600 termes

calculés sur la série initiale sont tous presque égaux à 1, ce qui dénote une tendance monotone. Les coefficients d'autocorrélation reste élevés sur la série des différences premières (compris entre 0.5 et 1), ce qui montre l'existence d'une tendance non nulle. Par contre, les coefficient d'autocorrélation sur les différences seconde mettent en évidence l'absence de tendance. Il a donc fallu aller jusqu'aux différence du seconde ordre pour annuler la tendance existante dans la série initiale.

### 2.4.5 Méthode des moindres carrés

Une autre méthode pour calculer la tendance est la méthode des moindres carrés, en régressant les observations par le temps.

#### Tendance polynômiale

Nous avons (dans la section 2.4.2) étudié le cas de tendance polynômiale à l'aide de moyennes mobiles, il paraît normal, lorsque la forme de la tendance est fixée, d'utiliser le critère des moindres carrés pour estimer les coefficients du polynôme supposé. Nous nous limitons ici au cas où la série ne subit pas de variations saisonnières.

Le critère des moindres carrés appliqué à un polynôme de degré 2 est donné par la formule suivante

$$S = \sum_{t=1}^T \{X_t - (at^2 + bt + c)\}^2$$

La méthode consiste à déterminer les coefficients  $a$ ,  $b$  et  $c$  tels que la quantité  $S$  soit minimale. Cette somme représente la somme des carrés des écarts entre les observations  $X_t$  et les valeurs du polynôme. On sait que le minimum est atteint pour les valeurs des coefficients qui annulent les dérivées partielles premières:

$$\begin{aligned} \frac{\partial S}{\partial a} &= - \sum 2\{X_t - (at^2 + bt + c)\}t^2 = 0 \\ \frac{\partial S}{\partial b} &= - \sum 2\{X_t - (at^2 + bt + c)\}t = 0 \\ \frac{\partial S}{\partial c} &= - \sum 2\{X_t - (at^2 + bt + c)\} = 0 \end{aligned}$$

On obtient donc le système de trois équations à trois inconnues suivant:

$$\begin{aligned} a \sum t^4 + b \sum t^3 + c \sum t^2 &= \sum X_t t^2 \\ a \sum t^3 + b \sum t^2 + c \sum t &= \sum X_t t \\ a \sum t^2 + b \sum t + cT &= \sum X_t \end{aligned}$$

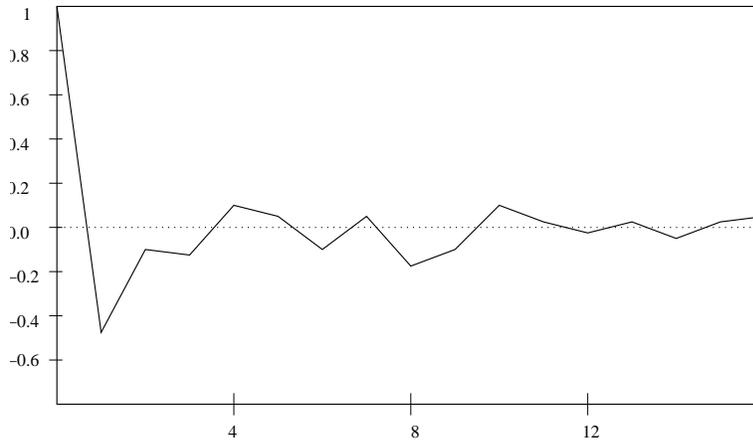


Figure 2.14: corrélogramme des différences premières d'une chronique à tendance linéaire.

Ces trois équations, appelées équations normales, se limitent lorsque le polynôme est de degré 1 (tendance linéaire), aux deux équations suivantes:

$$\begin{aligned} b \sum t^2 + c \sum t &= \sum X_t t \\ b \sum t + cT &= \sum X_t \end{aligned}$$

Donc on en déduit facilement les formules suivantes<sup>‡</sup>

$$b = \frac{\mathbf{Cov}(t, X_t)}{\mathbf{Var} t} \quad c = \mathbf{E} X_t - b \mathbf{E} t$$

Rappelons, pour faciliter es calculs:

$$\begin{aligned} \sum_{t=1}^T t &= \frac{T(T+1)}{2} \\ \sum_{t=1}^T t^2 &= \frac{T(T+1)(2T+1)}{6} \\ \sum_{t=1}^T t^3 &= \left(\frac{T(T+1)}{2}\right)^2 \\ \sum_{t=1}^T t^4 &= \frac{T(T+1)(2T+1)(3T^2+3T-1)}{30} \end{aligned}$$

Toutes les tendances ne sont évidemment pas linéaires ou polynomiales, la présence d'une limite vers laquelle la série converge rend impossible l'ajustement de la tendance par un polynôme. Nous présentons dans la suite que des modèles les plus courants.

<sup>‡</sup>Ici, on prolonge la définition d'espérance (variance) mathématique ainsi que son estimateur empirique à des fonctions déterministes:

$$\begin{aligned} \mathbf{E} f(t) &\sim \frac{1}{T} \sum_{t=1}^T f(t) \\ \mathbf{Var} f(t) &\sim \frac{1}{T} \sum_{t=1}^T (f(t) - \frac{1}{T} \sum_{t=1}^T f(t))^2 = \mathbf{E} f(t)^2 - (\mathbf{E} f(t))^2 \end{aligned}$$

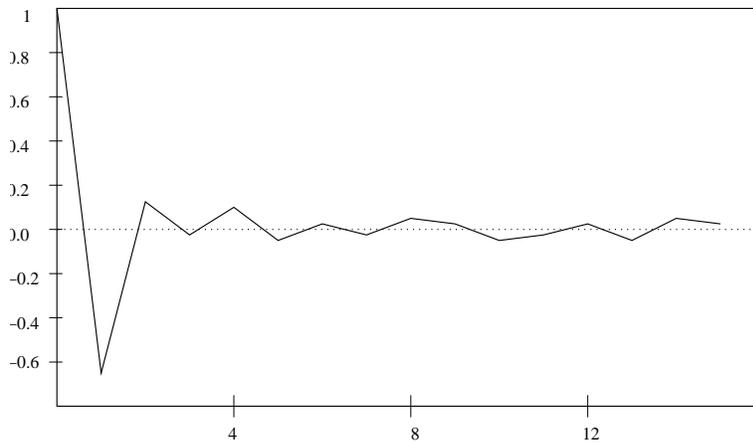


Figure 2.15: corrélogramme des différences successives d'une chronique à tendance polynômiale de degré 2

### Tendance exponentielle

La tendance exponentielle, souvent constatée dans les série d'observation économiques, s'exprime de la façon suivante :

$$T_t = ce^{bt} \quad \text{ou} \quad \log(T_t) = bt + \log(c)$$

On suppose  $c > 0$ , l'ajustement est réalisé en considérant le logarithme des observations et de la tendance : on cherche les valeurs  $b$  et  $c$  telles que la somme  $S$  soit minimale :

$$S = \sum_{t=1}^T \{(\log(X_t) - (bt + \log(c)))\}^2$$

On se ramène donc au cas l'ajustement linéaire, bien que le critère des moindres carrés, utilisé sur la série des logarithmes, ne soit pas équivalent au critère des moindres carrés appliqué aux observations.

### Modèle exponentiel à saturation

La tendance est sous la forme

$$T_t = a - be^{kt}$$

où  $a$  est un paramètre connu (limite de la tendance),  $b$  est positif et  $k$  négatif. La tendance est croissante vers  $a$  pour les grandes valeurs de  $t$  et l'accroissement marginal  $T_t$  est décroissant. L'ajustement est réalisé en se ramenant au cas précédent par le changement de variable suivant:

$$\mathcal{T}_t = T_t - T_{t-1} = b(1 - e^{-k})e^{kt}$$

## Modèle logistique

Le modèle logistique est défini de la façon suivante :

$$T_t = \frac{1}{a + be^{kt}}$$

où  $b > a > 0$  et  $k < 0$ . Cette fonction est encore croissante mais l'accroissement marginal n'est pas monotone : il a tendance à augmenter jusqu'à la valeur  $t = \frac{\log(a/b)}{k}$ , qui est mathématiquement l'abscisse du point d'inflexions, puis à diminuer. Ce modèle est particulièrement utilisé pour ajuster les taux d'équipement des ménages en biens de consommation durable. Pour effectuer l'ajustement, on se ramène au modèle précédent en posant :

$$\mathcal{T}_t = \frac{1}{T_t} = a + be^{kt}$$

Voir la figure 2.16<sup>§</sup>

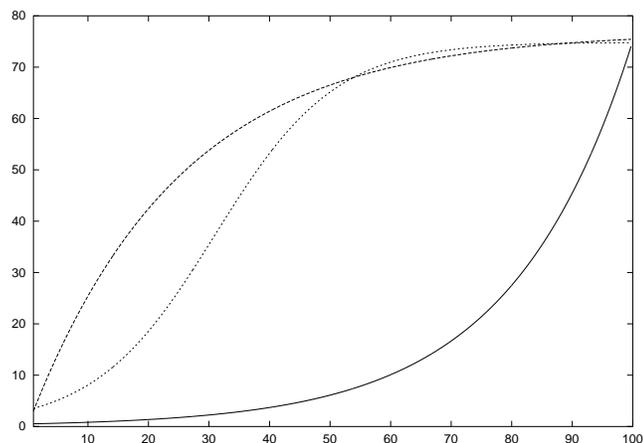


Figure 2.16: Tendence exponentielle, exponentielle à saturation, tendance logistique

<sup>§</sup>gnuplot plot 0.5 exp(0.05x), 77 - 77 exp(-0.04x), 1.0/((1.0/75) + 0.3 exp(-0.1x))

## 2.5 Estimation des variations saisonnières

Après la recherche du modèle et l'estimation de la tendance, nous étudions dans ce chapitre l'estimation des variations saisonnières

### 2.5.1 Estimation des coefficient saisonniers

Il s'agit maintenant d'estimer les variations saisonnières en supposant connue la tendance. Nous utilisons un tableau (Table de Buys Ballot 2.2). où  $p$  est la période, et  $n$  est le nombre de période considérée (bien sûr ici, on considère

	1	2	...	$j$	...	$p$	
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,j}$	...	$X_{1,p}$	$\overline{X}_{1.}$
2	$X_{2,1}$	$X_{2,2}$	...	$X_{2,j}$	...	$X_{2,p}$	$\overline{X}_{2.}$
...	...	...	...	...	...	...	...
$i$	$X_{i,1}$	$X_{i,2}$	...	$X_{i,j}$	...	$X_{i,p}$	$\overline{X}_{i.}$
...	...	...	...	...	...	...	...
$n$	$X_{n,1}$	$X_{n,2}$	...	$X_{n,j}$	...	$X_{n,p}$	$\overline{X}_{n.}$
	$\overline{X}_{.1}$	$\overline{X}_{.2}$	...	$\overline{X}_{.j}$	...	$\overline{X}_{.p}$	$\overline{\overline{X}}$

Table 2.2: Table de Buys Ballot

que  $p$  est connu). On a la relation suivante

$$t = (i - 1)p + j \quad 0 < j \leq p \quad i \geq 1$$

D'après le tableau on a aussi

$$\overline{X}_{i.} = \frac{1}{p} \sum_{j=1}^p X_{i,j} \quad \overline{X}_{.i} = \frac{1}{n} \sum_{i=1}^n X_{i,j}$$

et

$$\overline{\overline{X}} = \frac{1}{n} \sum_{i=1}^n \overline{X}_{i.} = \frac{1}{p} \sum_{j=1}^p \overline{X}_{.j}$$

Alors le modèle

$$X_t = T_t + S_t + \varepsilon_t$$

devient

$$X_{i,j} = T_{i,j} + S_j + \varepsilon_{i,j}$$

#### Estimation de coefficients saisonniers

On déduit de l'expression du modèle

$$\forall j = 1, \dots, p, \quad S_j = X_{i,j} - T_{i,j} - \varepsilon_{i,j} \quad \forall i = 1, \dots, n$$

Il est naturel de choisir pour estimateur  $S_j$  la moyenne

$$\overline{s}_j = \frac{1}{n} \sum_{i=1}^n (X_{i,j} - T_{i,j})$$

Quelques modifications sont utilisées

- Moyenne en éliminant de la somme ci-dessus les valeurs extrêmes des différences.
- Principe de la conservation des aires: La somme des coefficients saisonniers sur une période doit être nulle, pour cela, il suffit de considérer les valeurs  $\overline{s}_j - \overline{s}$  où  $\overline{s}$  est la moyenne de  $\overline{s}_j$  sur  $j$ .

$$\overline{s} = \frac{1}{p} \sum_{j=1}^p \overline{s}_j$$

A noter aussi que le filtre moyennes mobiles permet d'estimer la tendance sauf pour les premiers et les derniers termes, l'estimation des coefficients ne peut être effectuée que sur  $n - 1$  périodes dans le tableau Buys Ballot.

Les coefficients saisonniers sont utilisés pour *désaisonnaliser* les séries. L'objectif est d'éliminer des observations les variations saisonnières afin de pouvoir les comparer entre elles, et comparer la dernière observation connue à la précédente. On peut ainsi suivre l'évolution mois par mois du nombre de chômeurs, de l'indice des prix sans que la variation saisonnière soit prise en considération.

**Définition 2.5.1** Soit  $s_j$  l'estimateur des coefficients saisonniers, on appelle série désaisonnalisée (ou série corrigée des variations saisonnières) la série des observations après élimination des variations saisonnières

$$X'_{i,j} = X_{i,j} - \bar{s}_j$$

On obtient ainsi une série  $X'_t$ .

**Exemple 2.5.2 (TP).** On se donne dans un fichier obs une chronique  $X_t$ ,  $t = 1, \dots, 100$  de période  $p = 4$ . Calculer les coefficients saisonniers. On en utilise pour cela la moyenne mobile de longueur  $q = 4$ , pour estimer la tendance, et moyenne centrée pour estimer les coefficients saisonniers en éliminant la plus grande et la plus petite valeur dans chaque estimation.

## 2.5.2 Recherche simultanée de la tendance et des variations saisonnières

Dans les paragraphes précédents, nous avons proposé des méthodes d'estimation des variations saisonnières indépendantes de la procédure choisie pour ajuster la tendance. Il est clair qu'il est plus intéressant de lier les deux démarches

### Filtre de Buys Ballot

Le filtre de Buys Ballot consiste à ajuster, par le critère des moindres carrés, à la série  $X_{i,j}$  un modèle additif intégrant une tendance linéaire et des variations saisonnières constantes de période connue  $p$ . Il s'applique évidemment aux séries qui suivent le premier modèle multiplicatif. Soit

$$X_t = at + b + S_t + \varepsilon_t$$

en adoptant les notations de Buys Ballot,

$$X_{i,j} = a(p(i-1) + j) + b + S_j + \varepsilon_{i,j}$$

Les coefficients  $S_j$  ( $j = 1, \dots, p$ ) et les paramètres  $a$  et  $b$  de la tendance sont déterminés par minimisation de la somme des carrés des résidus.

$$Q = \sum_i \sum_j e_{i,j}^2 = \sum_i \sum_j (X_{i,j} - a(p(i-1) + j) - b_j)^2$$

où  $b_j \triangleq b + S_j$ . Nous admettrons que cette somme est minimale lorsque les dérivées partielles en  $b_j$  et  $a$  sont nulles. En ajoutant en plus le principe de conservation des aires ( $\sum S_j = 0$ ), on déduit pour les paramètres les valeurs suivantes

$$a = \frac{12}{np(n^2 - 1)} \left[ \sum_{i=1}^n (i \bar{X}_{i.}) - \frac{n(n+1)}{2} \bar{\bar{X}} \right]$$

$$b = \bar{\bar{X}} - \frac{a(np+1)}{2}$$

$$s_j = \bar{X}_{.j} - \bar{\bar{X}} - a\left(j - \frac{p+1}{2}\right)$$

On montre en fait que cet ajustement revient à déterminer la droite des moindres carrés sur la tendance définie par les moyennes annuelles affectées aux fins de périodes

$$(\bar{X}_{i\dots}, ip), \quad i = 1, \dots, n$$

et à estimer les coefficients saisonniers en écrivant que

$$\forall j = 1, \dots, p \quad s_j = \bar{X}_{.j} - aj - \text{cte}$$

La constante étant déterminée par le principe de conservation des aires:  $\sum s_j = 0$ .

### Filtre de Buys Ballot

La méthode Census est employée surtout en économie pour estimer simultanément la tendance et les variations saisonnières d'une chronique. Elle consiste à faire passer le filtre des moyennes mobiles, à estimer les coefficients saisonniers, à désaisonnaliser la série avec ces premières estimations, et à lisser à nouveau la série désaisonnalisée pour réestimer la tendance, ce qui permet enfin de réestimer définitivement les coefficients saisonniers.

La difficulté réside dans le choix des longueurs des moyennes mobiles. La première doit être évidemment de longueur égale à la période des variations saisonnières, et la longueur de la seconde est choisie après examen de la série désaisonnalisée.

### 2.5.3 Autocorrélation et variations saisonnières

On étudie l'influence de variations saisonnière sur les coefficients d'autocorrélation. Le but de ces études : mettre en évidence la période  $p$ .

**Théorème 2.5.3** Soit  $X_t$  une série chronique périodique de période  $p$  et de modèle  $X_t = S_r + \varepsilon_t$ , avec  $t = pq + r$ ,  $r > 0$ ,  $r \leq p$ . Alors les coefficients d'autocorrélations  $r_k$  sont approximativement périodique de période  $p$ .

#### différence de rang $p$

On appelle différence de rang  $p$  de la série  $X_t$  les différence de deux observations séparées par  $p$  terme  $X_{t+p} - X_t$ . On remarque que seul les différence premières sont égales à la différence de rang 1. Etudions l'effet de rang  $p$  sur une chronique sans tendance de période  $p$ :

$$X_t = S_r + \varepsilon_t \quad t = pq + r \quad 0 < r \leq p \quad \varepsilon_t \text{ bruit blanc}$$

Soit

$$Y_t = X_{t+p} - X_t = \varepsilon_{t+p} - \varepsilon_t$$

$$Y_{t+1} = X_{t+p+1} - X_{t+1} = \varepsilon_{t+p+1} - \varepsilon_{t+1}$$

La loi de  $X_{t+p} - X_t$  est donc indépendante de  $t$ ; Sous l'hypothèse que la série  $\varepsilon_t$  est un bruit blanc, il est évident que le coefficient d'autocorrélation de rang 1 est nul et qu'il en est de même pour tous ceux dont le rang est différent de  $p$  puisque les termes  $Y_t$  et  $Y_{t'}$  n'ont alors aucun point commun. Les coefficients d'autocorrélation de rang  $p$  sont par contre différents de 0<sup>¶</sup>:

$$Y_{t+p} = X_{t+2p} - X_{t+p} = \varepsilon_{t+2p} - \varepsilon_{t+p}$$

On en déduit facilement  $r_p = -\frac{1}{2}$

**Théorème 2.5.4** Soit  $X_t$  une chronique sans tendance et périodique de période  $p$ , alors la série des différence de rang  $p$  ne présente pas de variation saisonnière et tous les coefficients d'autocorrélation théorique sont nuls sauf le  $r_p$  est égale à  $-\frac{1}{2}$ .

Ce résultat, facile à montrer, est illustré dans la figure 2.17. La présence d'une tendance dans la chronique peut être envisagée : si elle est linéaire, le résultat est identique puisque cela revient à ajouter une constante aux différences  $Y_t$ . Si elle est de degré 2, elle n'est pas éliminée par le filtre des différences de rang  $p$ , qu'il faut éventuellement compléter par le filtre des différences premières.

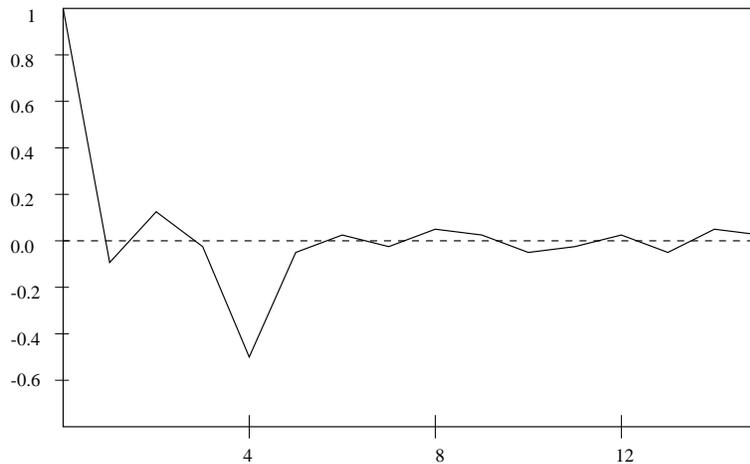


Figure 2.17: coefficients d'autocorrélation de  $Y_t = X_{t+4} - X_t$  pour une série de sans tendance de période  $p = 4$

<sup>¶</sup>  $Y_t \perp Y_{t'}$  pour  $t' > t$ , et  $t' \neq t + p$

# Chapter 3

## Modèles auto-régressifs et moyennes mobiles

### 3.1 Introduction aux processus stationnaires

On sait qu'une variable aléatoire réelle (v.a.r) est une application d'un espace probabilité  $(\Omega, \mathcal{A}, P)$  dans l'ensemble de nombre réels  $(R, \mathcal{B}_R)$

**Définition 3.1.1 Processus aléatoires discrets :** une famille de v.a.r.  $\{X_t\}_t \in Z$  définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $(R, \mathcal{B}_R)$

L'indice  $t$  qui désigne le temps  $Z$  est un ensemble discret, souvent  $Z$  est  $N$ , l'ensemble des entiers.

**Définition 3.1.2 Processus du 2 ordre :** Si  $E X_t$  et  $\text{Var } X_t$  existent pour  $\forall t \in Z$ , alors on appelle  $\{X_t\}$  un processus du 2 ordre

On ne considère que des processus du 2 ordre dans ce cours

**Remarque 3.1.3** On remarque

- Dans le cas d'une chronique, on ne connaît, pour un nombre fini de valeurs de  $t$ , qu'une seule réalisation des v.a.r.  $X_t$ , on ne peut pas utiliser les résultats des probabilité ni ceux de la statistique inférentielle qui sont établis à partir de la répétition de l'épreuve, on ne peut pas se payer le luxe de répéter ces expériences. La chronique est appelée trajectoire du processus.
- La connaissance d'une seule réalisation d'une v.a. ne permet pas d'en déduire des propriétés intéressantes. L'hypothèse de stationnarité, que nous définissons ci-dessous, permet de pallier cette difficulté, elle correspond à la modélisation suivante : chaque observation d'une série chronique est considérée comme la somme d'une valeur réelle déterminée par une fonction du temps et d'une valeur réelle résultant du hasard. Ces dernières sont supposées être des réalisations d'une même v.a.r.  $X$  dont les paramètres ne dépendent pas du temps. Ce qui permet d'appliquer les résultats de la statistique, puisque on dispose alors d'un échantillon  $\{X_t\}_{t=1, \dots}$  de la v.a.r. cet échantillon n'est pas nécessairement constitué des v.a.r. indépendantes

**Définition 3.1.4 Processus stationnaire** Un processus noté  $\{X_t\}$  est stationnaire si la loi du  $k$ -uple  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  est la même que celle du  $k$ -uple  $(X_{t_1+r}, X_{t_2+r}, \dots, X_{t_k+r})$  pour toutes valeurs de  $t_1, t_2, \dots, t_k, k, r$

C'est une propriété très stricte, qui implique en particulier qu'une même loi régit chacun des variables  $\{X_t\}, t \in I$ , et la loi de  $(X_1, X_2, \dots, X_n)$  est la même que celle de  $(X_{1+k}, X_{2+k}, \dots, X_{n+k})$ . Cette dernière implique qu'observer un échantillon de mesure  $(X_1, X_2, \dots, X_n)$  au temps  $1, 2, \dots, n$ , plutôt que faire plus tard cette observation de  $n$  mesures successives n'est pas important, que les lois mutuelles déterminées à partir d'observation du passé pourront être légitimement utilisées dans le future puisque qu'elles sont invariantes pour une translation du temps. En pratique, cette propriété de stationnarité sera supposée sans pouvoir être vérifiable ou testée.

**Définition 3.1.5 Processus stationnaire du second ordre** On appelle processus du 2 ordre un processus tel que la loi de la v.a.  $X_t$  est indépendante de  $t$ , et que la covariance de  $X_t$  et  $X_{t+k}$  ne dépend que de  $k$  pour  $\forall t$  et  $k$ .

d'une façon équivalente,  $X_t$  est stationnaire du 2 ordre si la loi de  $X_t$  est indépendante de  $t$  et  $\forall t, s, (s, t) \in I$ , la quantité  $\text{Cov}(X_t, X_s)$  ne dépend que de l'écart  $t - s$  et non de  $s$ .

**Remarque 3.1.6** En pratique

- Une série observée sera dite stationnaire quand un processus stationnaire modélise convenablement cette série.

- *Compte tenu des propriétés de la loi normale, un processus gaussien stationnaire du second ordre est stationnaire. Nous nous limitons dans la suite de ce chapitre aux processus stationnaires du 2 ordre que nous appellerons simplement processus stationnaires.*
- *On aura intérêt à avoir plutôt une série observée modélisable par un processus gaussien. Sinon, on envisage à le transformer en un processus qui s'avérerait modélisable par un processus gaussien.*

**Définition 3.1.7 processus centré :** Un processus  $\{X_t\}$  est centré si  $\mathbf{E} X_t = 0, \forall t$

Dans un premier temps, nous intéressons aux processus centrés et stationnaires, ils vérifient aussi la propriété dite d'homoscédasticité :

$$\mathbf{Var} X_t = \mathbf{E} X_t^2 = \sigma_X^2 \quad \forall t$$

de plus pour tout  $k$ ,

$$\mathbf{Cov}(X_t, X_{t-k}) = \mathbf{Cov}(X_t, X_{t+k}) = \mathbf{E} X_t X_{t-k} = c_k, \quad \forall t$$

et  $\rho_k = \frac{c_k}{\sigma_X^2}$

**Définition 3.1.8 Processus ergodique** *Un processus est dit ergodique quand les limites des moments empiriques convergent en moyenne quadratique vers les espérance correspondantes, lorsque la taille de la série s'accroît.*

Par exemple, la moyenne d'un processus stationnaire est définie par  $\mathbf{E} X_t$ , mais pour  $t$  fixé, nous avons aucune répétition d'échantillon pour estimer cette moyenne, l'ergodicité du processus stationnaire suppose que pendant un temps assez long, la moyenne empirique temporelle va être approximativement indépendante et qui permettra une estimation

$$\mathbf{E} X_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$$

Lorqu'on observe le processus suffisamment longtemps, les estimations des paramètres des modèles considérés faites à partir des observations sont alors valides et ces estimations ne seront pas influencées par des conditions initiales particulières

**Théorème 3.1.9** *Tout processus stationnaire gaussien  $X_t$  tel que*

$$\mathbf{Cov}(X_t, X_s) \longrightarrow 0 \quad \text{quand } |t - s| \rightarrow \infty$$

*est ergodique*

Pour les processus stationnaires considérés,  $\rho_k$  tend vers zéro quand  $k \rightarrow \infty$ , alors ils sont ergodique s'ils sont gaussien.

**Définition 3.1.10 Opérateur retard  $B$**  *Soit  $\{X_t\}$  un processus quelconque, on appelle opérateur retard l'opérateur  $B$  défini par*

$$BX_t = X_{t-1} \quad \forall t$$

Ce qui permet de définir un nouveau processus  $B\{X_t\}$ . On note  $\mathbb{1}$  l'opérateur identité et  $B^2$  l'opérateur  $B \circ B$ .

$$\mathbb{1}X_t = X_t, \quad B^2X_t = X_{t-2}, \quad \forall t$$

Les  $\mathbb{1}, B, B^2$  font parties de ce qu'on appelle opérateurs linéaires, qui définissent des nouveaux processus à partir de combinaisons linéaires de processus. Plus généralement, on note  $P(B)$  (comme on parle de polynôme) de degré  $n$ , en confondant la notation  $a_0$  avec celle de  $a_0\mathbb{1}$ :

$$P(B) = a_0 + a_1B + a_2B^2 + \dots + a_nB^n$$

et

$$P(B)X_t = (a_0 + a_1B + \dots + a_nB^n)X_t = a_0X_t + a_1X_{t-1} + \dots + a_nX_{t-n}$$

Les opérations habituelles sont valides, par exemple

$$(1 - B)^2X_t \triangleq (1 - 2B + B^2)X_t$$

On fera usage de fractions rationnelles, de décomposition en éléments simples de ces fractions, de développements polynomiaux. Par exemple, soit  $\lambda$  un nombre complexe de module inférieur à 1, on sait que

$$(1 - \lambda x)^{-1} = 1 + \lambda x + \lambda^2 x^2 + \dots + \lambda^n x^n + \dots$$

de sorte que avec  $|\lambda| < 1$ , on note

$$(1 - \lambda B)^{-1} = \sum_0^{\infty} \lambda^i B^i$$

ce qui permet de définir un nouveau processus par  $(1 - \lambda B)^{-1}X_t$ .

**Théorème 3.1.11** Des processus précédents tels que  $P(B)X_t$  et  $(1 - \lambda B)^{-1}X_t$  sont stationnaires si le processus  $\{X_t\}$  est stationnaire.

plus généralement, un opérateur  $f$  est un filtre linéaire quand c'est un opérateur invariant par changement de l'origine du temps. Le processus filtré  $\{Y_t\}$  défini par

$$Y_t = f(X_t)$$

vérifie ainsi

$$Y_{t-\tau} = f(X_{t-\tau}) \quad \forall t, \forall \tau$$

est alors stationnaire si  $\{X_t\}$  est stationnaire.

## 3.2 Fonction d'autocorrelation partielle

Dans le chapitre 2.4, les coefficients de corrélation  $\rho_k$  est estimé par les fonctions d'autocorrélation  $r_k$ . Sa représentation graphique est le corrélogramme.

**Remarque 3.2.1** On remarque que

1. Si  $\rho_k$  est proche de 1, c'est qu'une grande (resp. petite) valeur observée au temps  $t - k$  est suivie le plus souvent d'une grande (resp. petite) valeur au temps  $t$  et  $\forall t$ .
2. Si  $\rho_k$  est proche de  $-1$ , c'est qu'une grande (resp. petite) valeur observée au temps  $t - k$  est suivie le plus souvent d'une petite (resp. grande) valeur au temps  $t$ , et cela pour  $\forall t$
3. Si  $\rho_k$  est proche de 0, qu'une grande valeur ou une petite valeur observée au temps  $t - k$  est suivie le plus souvent d'une valeur quelconque au temps  $t$

Soit  $\{X_t\}$  un processus tel que\*

$$X_t = \phi_{11}X_{t-1} + R_{t,1}, \quad \forall t$$

où  $\phi_{11} \in \mathbf{R}$ , la variable résidu  $R_t$  est non corrélée à la variable explicative  $X_{t-1}$ . Soit en générale

$$X_t = \phi_{1,k}X_{t-1} + \phi_{2,k}X_{t-2} + \dots + \phi_{k,k}X_{t-k} + R_{t,k}$$

et  $R_{t,k}$  est non corrélée aux variables explicatives  $X_{t-1}, \dots, X_{t-k}$ . Cela implique que les variables du passé proche, associées aux premières variables explicatives, pourraient suffire pour expliquer totalement le présent. Dans le cas où  $k = 1$

$$X_t X_{t-1} = \phi_{11}X_{t-1}X_{t-1} + R_{t,1}X_{t-1} \quad \forall t$$

Donc pour  $\forall t \geq 1$

$$\mathbf{E} X_t X_{t-1} = \phi_{11} \mathbf{E} X_{t-1} X_{t-1} + \mathbf{E} R_{t,1} X_{t-1} = \phi_{11} \mathbf{Var} X_{t-1}$$

c'est à dire  $\phi_{11} = \rho_1$  de même, dans le cas où  $k = 2$  on a

$$X_t X_{t-1} = \phi_{12}X_{t-1}X_{t-1} + \phi_{22}X_{t-2}X_{t-1} + R_{t,2}X_{t-1}$$

$$X_t X_{t-2} = \phi_{12}X_{t-1}X_{t-2} + \phi_{22}X_{t-2}X_{t-2} + R_{t,2}X_{t-2}$$

l'espérance mathématique

$$\mathbf{E} X_t X_{t-1} = \phi_{12} \mathbf{E} X_{t-1}^2 + \phi_{22} \mathbf{E} X_{t-2} X_{t-1}$$

$$\mathbf{E} X_t X_{t-2} = \phi_{12} \mathbf{E} X_{t-1} X_{t-2} + \phi_{22} \mathbf{E} X_{t-2}^2$$

On divise tout par  $\mathbf{E} X_t = \sigma_x^2$  et en utilisant le fait que  $X_t$  est stationnaire centrée, on a

$$\begin{cases} \rho_1 &= \phi_{12} + \rho_1 \phi_{22} \\ \rho_2 &= \rho_1 \phi_{12} + \phi_{22} \end{cases} \implies \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

En générale la relation entre  $\phi_{kk}$  et  $\rho$  dans les régressions

$$X_t = \phi_{1,k}X_{t-1} + \phi_{2,k}X_{t-2} + \dots + \phi_{k,k}X_{t-k} + R_{t,k}$$

---

\*régression linéaire de  $X_t$  en  $X_{t-1}$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}_{k \times k}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}_{k \times k}} \quad (\text{Yules-Walker})$$

Par exemple si  $k = 2$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

**Définition 3.2.2** On appelle la fonction  $\Phi(k) \triangleq \phi_{kk}$  fonction autocorrélation partielle.

**Exercice 3.2.3** Montrer le corrélogramme d'un bruit blanc est nul pour  $\forall k \geq 1$ . son corrélogramme partiel  $\Phi$  l'est aussi.

nous dirons qu'un processus  $\{X_t\}$  est un processus centré engendré à partir d'un bruit blanc (noté  $\{\varepsilon_t\}$ ), si et seulement si on peut écrire

$$X_t = \sum_0^{\infty} \psi_j \varepsilon_j$$

avec  $\psi_0 = 1$ ,  $\psi_j \in \mathbf{R}$  et  $\sum_0^{\infty} |\psi_j| < \infty$  Tous les processus stationnaires considérés dans ce chapitre sont des processus qui peuvent être considérés comme engendrés à partir d'un bruit blanc.

**Définition 3.2.4** Le modèle auto-régressifs et moyenne mobile ARMA (autoregressive moving average), est

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

où  $\{\varepsilon_t\}$  est un bruit blanc.

1.  $\phi_1 = \phi_2 = \cdots = \phi_p = 0$  alors on appelle  $X_t$  moyenne mobile MA, si de plus  $\theta_q \neq 0$ , alors il s'agit d'un processus MA( $q$ ) d'ordre  $q$ .
2.  $\theta_1 = \theta_2 = \cdots = \theta_q = 0$  alors on appelle  $X_t$  modèle autoregressif, si de plus  $\phi_p \neq 0$ , alors il l'ordre de ce AR est  $p$ , noté AR( $p$ )

En générale, on note ce modèle par ARMA( $p, q$ ). On présente d'abord quelques modèles particuliers.

### 3.3 Modèle en moyenne mobile

#### 3.3.1 modèles en moyenne mobile d'ordre 1

On considère dans ce paragraphe le modèle MA(1).

$$X_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

où  $\theta$  est une constante et  $\{\varepsilon_t\}$  un bruit blanc. On peut alors calculer de manière explicite

$$\mathbf{E} X_t = \mu, \quad \mathbf{Var} X_t = (1 + \theta^2) \sigma^2$$

$$\mathbf{Cov}(X_t, X_{t-1}) = \theta \sigma^2, \quad \mathbf{Cov}(X_t, X_{t-j}) = 0, \quad j > 0$$

Comme  $\mathbf{E} X_t$  et  $\mathbf{Cov}(X_t, X_{t-j})$  ne dépende pas de  $t$ , MA(1) est donc un processus stationnaire. Les coefficients d'autocorrélation théorique sont (voir la figure 3.1)

$$\rho_1 = \frac{\theta \sigma^2}{(1 + \theta^2) \sigma^2} = \frac{\theta}{(1 + \theta^2)}, \quad \rho_2 = \rho_3 = \cdots = 0$$

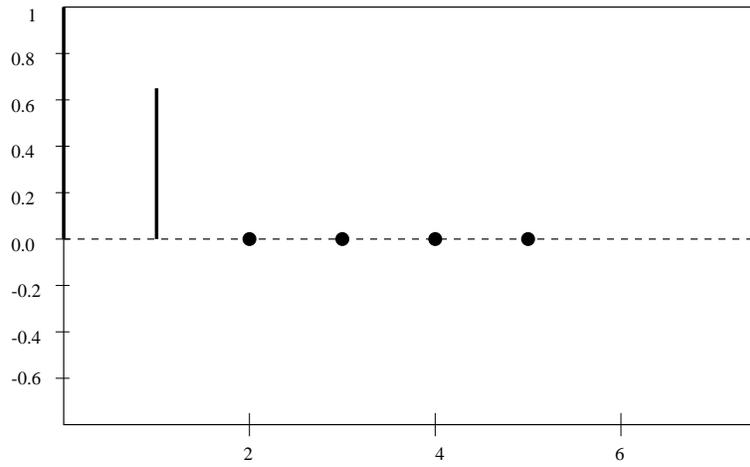


Figure 3.1: Coefficients d'autocorrélation de MA(1).

### 3.3.2 modèles en moyenne mobile d'ordre $q$

On considère le processus

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \quad \theta_q \neq 0$$

Après calculs on a

$$\begin{aligned} \mathbf{E} X_t &= \mu \\ \mathbf{Var} X_t &= \mathbf{E} (\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})^2 \\ &= (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \sigma^2 \\ \mathbf{Cov} (X_t, X_{t-j}) &= \mathbf{E} [(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}) \times (\varepsilon_{t-j} + \theta_1 \varepsilon_{t-j-1} + \cdots + \theta_q \varepsilon_{t-j-q})] \\ &= \begin{cases} (\theta_j + \theta_{j+1} \theta_1 + \theta_{j+2} \theta_2 + \cdots + \theta_q \theta_{q-j})^2 & j = 1, 2, \dots, q \\ 0, & j > q \end{cases} \end{aligned}$$

Par exemple, pour MA(2) nous avons

$$\begin{aligned} \gamma_j &= \mathbf{Cov} (X_t, X_{t-j}) \quad \text{Donc} \quad \gamma_0 = (1 + \theta_1^2 + \theta_2^2) \sigma^2 \\ \gamma_1 &= (\theta_1 + \theta_2 \theta_1) \sigma^2 \\ \gamma_2 &= \theta_2 \sigma^2 \\ \gamma_3 &= \gamma_4 = \cdots = 0 \end{aligned}$$

On pourra donc calculer facilement les coefficients de corrélation  $\rho_j = \frac{\gamma_j}{\gamma_0}$ ,  $j = 1, 2, \dots$ . le corrélogramme est alors  
Question : quelle est la condition stationnarité ?

## 3.4 Modèle autorégressif

Nous commençons par

### 3.4.1 Modèle autorégressif du premier ordre

Soit

$$X_t = c + \phi X_{t-1} + \varepsilon_t$$

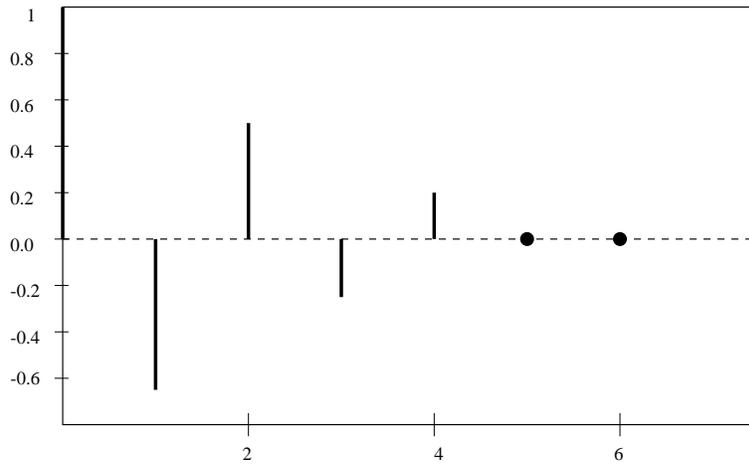


Figure 3.2: Corrélogramme MA(4),  $X_t = \varepsilon_t - 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} - 0.5\varepsilon_{t-3} + 0.5\varepsilon_{t-4}$

En fait si on prend  $W_t = c + \varepsilon$ , alors AR(1) est une équation différentielle  $X_t = \phi X_{t-1} + W_t$

$$X_1 = \phi X_0 + W_1 = \phi^2 X_{-1} + \phi W_0 + W_1$$

$$X_t = \phi^{t+1} X_{-1} + \phi^{t-1} W_0 + \dots + \phi W_{t-1} + W_t \text{ et}$$

$$X_{t+j} = \phi^{j+1} X_{t-1} + \phi^j W_t + \phi^{j-1} W_{t+1} + \dots + \phi W_{t+j-1} + W_{t+j}$$

Supposons  $X_{t-1}$  et  $W_t, W_{t+1}, \dots, W_{t+j}$  connus alors le multiplicateur dynamique est

$$\frac{\partial X_{t+j}}{\partial W_t} = \phi^j$$

**Remarque 3.4.1** Si  $W_t$  varie (et  $W_{t+1}, \dots, W_{t+j}$  ne changent pas), l'effet<sup>†</sup> sur  $X_{t+j}$  est décrit par  $\frac{\partial X_{t+j}}{\partial W_t}$ , d'où le nom multiplicateur dynamique.

On distingue alors plusieurs cas possibles

1.  $0 < \phi < 1$  alors  $\frac{\partial X_{t+j}}{\partial W_t}$  décroît géométriquement vers 0.
2.  $-1 < \phi < 0$  alors  $\frac{\partial X_{t+j}}{\partial W_t}$  change de signe de manière alternative et amplitude décroît géométriquement vers 0.
3.  $\phi > 1$  effet de  $W_t$  sur  $X_{t+j}$  est exponentiel par rapport à  $j$ , explosif.
4.  $\phi < -1$  oscillation explosive.

Donc si  $|\phi| < 1$  le système est stable, et si  $|\phi| > 1$  non stable. cela se traduit par

1. si  $|\phi| < 1$  alors  $X_t$  est stationnaire
2. si  $|\phi| > 1$  alors  $X_t$  est non stationnaire
3. si  $|\phi| = 1$  alors  $X_t$  est non stationnaire (à réfléchir)

Si  $|\phi| < 1$  on a

$$\begin{aligned} X_t &= (c + \varepsilon_t) + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3(c + \varepsilon_{t-3}) + \dots \\ &= \frac{c}{1-\phi} + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots \quad (\text{représentation MA}(\infty) \text{ de AR}) \end{aligned}$$

$$\mathbf{E} X_t = \frac{c}{1-\phi}$$

$$\begin{aligned} \mathbf{Var} X_t &= \mathbf{E} (X_t - \mathbf{E} X_t)^2 = \mathbf{E} (\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots)^2 \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \dots)\sigma^2 = \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

<sup>†</sup>il s'agit d'effet du présent sur la future

Attention, contrairement aux chapitres précédents, on a considéré la série comme

$$(\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots, X_t).$$

Par ailleurs  $\forall j \geq 0$

$$\begin{aligned} \gamma_j &= \mathbf{Cov}(X_t, X_{t-j}) \\ &= \mathbf{E} [(\varepsilon_t + \phi\varepsilon_{t-1} + \dots + \phi^j\varepsilon_{t-j} + \phi^{j+1}\varepsilon_{t-j-1} + \phi^{j+2}\varepsilon_{t-j-2} + \dots) \\ &\quad \times (\varepsilon_{t-j} + \phi\varepsilon_{t-j-1} + \phi^2\varepsilon_{t-j-2} + \dots)] \sigma^2 \\ &= [\phi^j + \phi^{j+2} + \phi^{j+4} + \dots] = \frac{\phi^j}{1 - \phi^2} \sigma^2 \end{aligned}$$

Les coefficients autocorrélations sont (voir la figure 3.3 par exemple)

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j$$

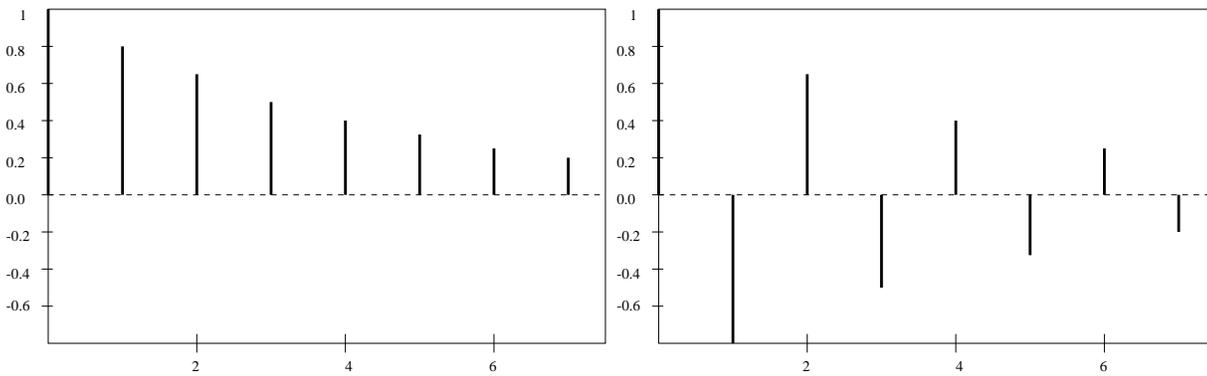


Figure 3.3: Corrélogramme AR(1) :  $X_t = 0.8X_{t-1} + \varepsilon_t$  et AR(1) :  $X_t = -0.8X_{t-1} + \varepsilon_t$

### 3.4.2 Modèle autorégressif du second ordre

On considère dans cette paragraphe le processus AR(2)

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

La condition pour que AR(2) soit stationnaire est que l'équation  $1 - \phi_1 z - \phi_2 z^2 = 0$  possède deux racines (réelles ou complexes) de module  $> 1$ .<sup>‡</sup> On calcule maintenant quelques statistiques pour AR(2).

$$\mu = \mathbf{E} X_t = c + \phi_1 \mu + \phi_2 \mu \implies \mu = \frac{c}{1 - \phi_1 - \phi_2}$$

Pour calculer  $\mathbf{E} X_t^2$  (ou  $\mathbf{Var} X_t$ ) on utilise

$$X_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \varepsilon_t$$

multipli  $X_{t-j} - \mu$  et calculer l'espérance on a

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, \quad j = 1, 2, \dots$$

diviser tout par  $\gamma_0$  on a aussi

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$$

en particulier,  $j = 1$  et  $j = 2$

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \implies \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

<sup>‡</sup>On admet ce résultat, il faut étudier la condition de stabilité d'équation différentielle  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$ .

On peut utiliser ces méthodes pour trouver

$$\gamma_0 = \frac{(1 - \phi_1)\sigma^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \quad \text{exercice}$$

En effet, il suffit de calculer

$$\begin{aligned} (X_t - \mu)^2 &= \phi_1^2(X_{t-1} - \mu)^2 + \phi_2^2(X_{t-2} - \mu)^2 + \varepsilon_t^2 \\ &\quad + 2\phi_1\phi_2(X_{t-1} - \mu)(X_{t-2} - \mu) + 2\phi_1(X_{t-1} - \mu)\varepsilon_t + 2\phi_2(X_{t-2} - \mu)\varepsilon_t \end{aligned}$$

$X_t - \mu$  étant stationnaire et moyenne nulle,  $\varepsilon_t$  un bruit blanc indépendant de  $X_{t-1}$  et  $X_{t-2}$ , prenons espérance mathématique nous avons

$$\gamma_0 = \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + \sigma^2 + 2\phi_1\phi_2\gamma_1 = \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + \sigma^2 + 2\phi_1\phi_2\rho_1\gamma_0$$

Donc

$$[1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2\rho_1] \gamma_0 = \sigma^2$$

Pour montrer le résultat, il suffit alors de remplacer  $\rho_1$  par  $\frac{\phi_1}{1 - \phi_2}$ . On peut aussi montrer la relation entre  $(\phi_1, \phi_2)$  et  $(\gamma_1, \gamma_2)$  (*exercice*)

$$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (\text{Yule-Walker}).$$

la condition stationnarité pour AR(2) est présentée par la figure 3.4.

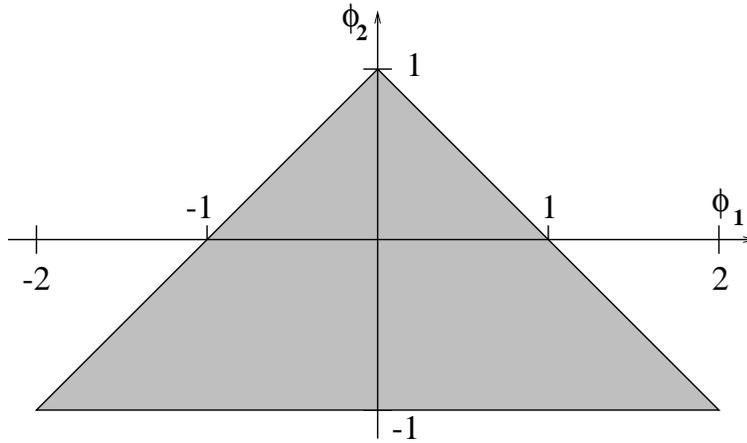


Figure 3.4: Condition de stationnarité pour un modèle de AR(2)

### 3.4.3 Modèle autorégressif d'ordre $p$

$$\text{AR}(p) : X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

Soit le polynôme caractéristique  $\Phi(z) = 1 - \sum_{k=1}^p \phi_k z^k$   $z \in \mathbb{C}$  ne possède pas de racines réelles ou complexes de module inférieur ou égale à 1 (condition stationnarité). On peut effectuer quelques calculs simples

$$\mathbf{E} X_t = \mu = c + \phi_1 \mu + \phi_1 \mu + \dots + \phi_p \mu$$

$$\mathbf{E} X_t = \mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

En écrivant AR( $p$ ) par

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t$$

Multipli par  $X_{t-j} - \mu$  et calculer l'espérance, on obtient

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p} & j = 1, 2, \dots \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p & j = 0 \end{cases}$$

diviser par  $\gamma_0$  on obtient l'équation de Yule-Walker:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \quad j = 1, 2, \dots$$

### 3.4.4 Inversibilité

On présente dans la suite la notion de inversibilité pour certain ARMA, elle est nécessaire pour identifier un ARMA. Pour comprendre cette notion, on introduit d'abord la notion *représentation*  $\text{AR}(\infty)$  d'un modèle ARMA. Soit  $\Phi(B)X_t = \Theta(B)\varepsilon_t$ .

#### Représentation $\text{MA}(\infty)$ et $\text{AR}(\infty)$ d'un ARMA

Pour un modèle  $\text{ARMA}(p, q)$ ,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

on peut obtenir une représentation  $\text{MA}(\infty)$

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{ou} \quad X_t = \Psi(B)\varepsilon_t$$

Par exemple, pour  $\text{ARMA}(1, 1)$  on peut

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} = \phi(\phi X_{t-2} + \varepsilon_{t-1} + \theta X_{t-2}) + \varepsilon_t + \theta \varepsilon_{t-1} \\ &= \phi^2 X_{t-2} + \varepsilon_t + (\theta + \phi)\varepsilon_{t-1} + \theta \varepsilon_{t-2} \\ &= \phi^2(\phi X_{t-3} + \varepsilon_{t-2} + \theta \varepsilon_{t-3}) + \varepsilon_t + (\theta + \phi)\varepsilon_{t-1} + \theta \varepsilon_{t-2} = \text{etc} = \text{etc} \end{aligned}$$

Il y a une deuxième méthode (plus efficace). Soit  $\Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$  on identifie les coefficients  $\psi_0, \psi_1, \psi_2, \dots$  par équation

$$\Phi(B)\Psi(B)\varepsilon_t = \Theta(B)\varepsilon_t$$

Dans l'exemple précédent, on obtient  $(1 - \phi B)\Psi(B) = 1 + \theta B$  on obtient donc

$$\begin{aligned} \psi_0 &= 1 = \text{coefficient devant } B^0 \\ \psi_1 &= \theta + \phi = \text{coefficient devant } B \\ \psi_2 &= \phi\psi_1 = \phi(\theta + \phi) = \text{coefficient devant } B^2 \\ \text{etc} & \quad \text{etc} \end{aligned}$$

donc

$$X_t = \varepsilon_t + (\theta + \phi)\varepsilon_{t-1} + \phi(\theta + \phi)\varepsilon_{t-2} + \dots$$

La troisième méthode consiste à inverser AR polynôme.

$$\begin{aligned} (1 - \phi B)X_t &= (1 + \theta B)\varepsilon_t \quad \text{alors} \quad X_t = (1 - \phi B)^{-1}(1 + \theta B)\varepsilon_t \\ X_t &= (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B) \end{aligned}$$

Si  $p \geq 2$ , on doit alors inverser un polynôme de degré  $\geq 2$  i.e.

$$X_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)^{-1}(1 + \theta B)\varepsilon_t$$

Or l'inversion d'un polynôme est souvent difficile, on suppose d'abord  $\exists$  un factorisation de  $\Phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  :

$$\Phi(B) = \left(1 - \frac{1}{r_1} B\right) \left(1 - \frac{1}{r_2} B\right) \dots \left(1 - \frac{1}{r_p} B\right)$$

où  $r_1, r_2, \dots, r_p$  sont les racines de  $\Phi(B)$ , alors

$$\Phi^{-1}(B) = \frac{a_1}{1 - \frac{1}{r_1} B} + \frac{a_2}{1 - \frac{1}{r_2} B} + \dots + \frac{a_p}{1 - \frac{1}{r_p} B}$$

et on écrit pour chaque terme de  $\Phi^{-1}(B)$  par  $\frac{1}{1 - \frac{1}{r_1}B} = 1 + \frac{1}{r_1}B + \left(\frac{1}{r_2}B\right)^2 + \dots$  pour obtenir la représentation infinie MA( $\infty$ ).

Le modèle ARMA a une représentation AR( $\infty$ ) également :

$$\varepsilon_t = \sum_{j=0}^{\infty} \xi_j X_{t-j}$$

Par exemple pour ARMA(1, 1)

$$\begin{aligned} \varepsilon_t &= X_t - \phi X_{t-1} - \theta \varepsilon_{t-1} = X_t - \phi X_{t-1} - \theta(X_{t-1} - \phi X_{t-2} - \theta \varepsilon_{t-2}) \\ &= X_t - (\phi + \theta)X_{t-1} + \theta(\phi + \theta)X_{t-2} - \theta^2(\phi + \theta)X_{t-2} + \dots (-\theta)^{k-1}(\phi + \theta)X_{t-k} + \dots \end{aligned}$$

On peut bien sûr utiliser l'identification du polynôme  $\Theta(B)$ . Soit  $\varepsilon_t = \Xi(B)X_t$  où  $\Xi(B) = \xi_0 + \xi_1 B + \xi_2 B^2 + \dots$  est à identifier. L'équation  $\Phi(B) = \Theta(B)\Xi(B)$  on obtient pour ARMA(1, 1)

$$\xi_0 = 1, \quad \xi_1 = -(\theta + \phi), \quad \xi_2 = \theta \xi_1, \quad \xi_3 = -\theta \xi_2, \quad \dots$$

La troisième méthode consiste aussi d'inverser un polynôme  $\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t$  pour ARMA( $p, 1$ ) on utilise  $(1 + \theta B)^{-1} = 1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \dots$ . Le cas où  $q \geq 2$  est difficile.

### Inversibilité d'un modèle ARMA

L'existence d'une représentation AR( $\infty$ ) :  $\varepsilon_t = \sum_{j=0}^{\infty} \xi_j X_{t-j}$  est importante pour l'estimation et la prédiction (c'est même essentielle si on veut faire de calculs numériques). Par exemple : pour un modèle MA(1) on a

$$\begin{aligned} X_t &= (1 + \theta B)\varepsilon_t = \varepsilon_t + \theta \varepsilon_{t-1} \\ \varepsilon_t &= (1 + \theta B)^{-1}X_t = (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \dots)X_t \end{aligned}$$

- Cas 1, si  $\theta > 1$ , le poids de  $X_{t-k}$  est  $(-\theta)^k$  croissant, on a besoin de tous les passés pour calculer  $\varepsilon_t$ .
- Cas 2, si  $\theta = 1$ , on a besoin de tous les passés  $X_{t-1}, X_{t-2}, \dots$  pour calculer  $\varepsilon_t$  et  $\sum_{j=0}^{\infty} |\xi_j|$  diverge
- Cas 3,  $|\theta| < 1$ , alors le poids de  $X_{t-k}$  décroît vers 0, on appelle ce cas *inversible*.

**Théorème 3.4.2** *La condition pour que un modèle ARMA( $p, q$ ) ( $\Phi(B)X_t = \Theta(B)\varepsilon_t$ ) soit stationnaire est que tous les racines du polynôme  $\Phi(z)$  (réelles ou complexes) ont des modules supérieurs à 1. La condition pour que un modèle ARMA soit inversible est que tous les racines du polynôme  $\Theta(B)$  (réelles ou complexes) ont des modules supérieurs à 1.*

*Preuve:* Admis. En effet, cela vient du fait que il faut

- Stationnarité : avoir une représentation convergente de MA( $\infty$ ).
- Inversibilité : avoir une représentation convergente de AR( $\infty$ ).

On remarque que le calcul des corrélogramme d'un modèle de ARMA mixte ne permet pas de identifier l'ordre ( $p, q$ ) d'un ARMA( $p, q$ ).

# Chapter 4

## Estimation, Identification et Prévision des modèles ARIMA

### 4.1 Introduction

L'identification de l'ordre  $(p, q)$  du modèle mixte est délicate dans la mesure où les méthodes à mettre en oeuvre sont tous imprécises ou deviennent si coûteuses quand la précision est recherchée, qu'elle sont pas toujours portés sur les logiciels. Il existe des méthodes heuristiques comme par exemple la méthode dite *du coin*. Une autre approche consiste à ne s'intéresser aux modèles mixtes qu'après avoir constaté qu'il ne peut s'agir ni d'un AR ni d'un MA :

- l'ordre  $p$  d'un  $AR(p)$  peut se lire sur corrélogramme  $\Phi$  des autocorrélations partielles, ce corrélogramme étant nulle à partir de  $k = p + 1$ .
- l'ordre  $q$  d'un  $MA(q)$  se lit sur corrélogramme, se comportant comme un corrélogramme nul à partir des décalages supérieurs ou égaux à  $q + 1$  et non nul en deça.

avec  $p, q$  données, il reste à estimer les paramètres de ce modèle en utilisant les observations  $(X_1, \dots, X_t)$ : la variation du bruit blanc, les coefficients  $\phi_1, \phi_2, \dots, \phi_p$  et  $\theta_1, \theta_2, \dots, \theta_q$ . En pratique, la valeur de  $p$  et  $q$  sont rarement  $\geq 3$ . Alors que signifie une bonne estimation ? Les critères sont

- sans biais, • petite erreur (au sens moyenne quadratique)

Un des objectifs principaux est de prédire l'avenir :

$$\mathbf{E}(X_{t+h} | X_1, X_2, \dots, X_t)$$

### 4.2 Estimation

On présente dans cette section deux méthodes d'estimation différentes, on considère tout d'abord des exemples simples.

#### 4.2.1 Méthode de moindre carré

Soit

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

Les paramètres à estimer sont  $\sigma^2$ ,  $\phi_1$  et  $\phi_2$ , on dispose des observations sous forme d'un vecteur de taille  $N$  :  $(X_1, \dots, X_N)$ . Il s'agit donc de chercher  $\phi_1$  et  $\phi_2$  t. q.

$$\sum_{t=1}^N (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2 \quad \text{soit minimale} \quad (4.1)$$

Les variables  $X_{-1}$  et  $X_0$  ne sont pas disponibles, on cherche

- $\hat{\phi}_1, \hat{\phi}_2$  t. q.

$$\sum_{t=3}^N (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2})^2 \quad \text{soit minimale}$$

On suppose en fait que les 2 premiers termes dans 4.1 sont nuls

- La valeur de  $\sigma^2$  sont estimée par

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{t=3}^N (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2})^2$$

On peut facilement appliquer le principe de ces méthodes pour des problèmes plus généraux, la méthode de moindre carré est facile à mettre en oeuvre, elle est valide quand la taille  $N$  est importante.

### 4.2.2 Méthode de maximum de vraisemblance

La méthode de maximum de vraisemblance est une des méthodes la plus utilisée en statistique, elle est très intéressante car sous des conditions assez générales, l'estimateur est asymptotiquement sans biais et de variance minimale. On suppose au départ que les  $\{X_t\}$ ,  $t = 1, \dots, N$  sont indépendantes de même distribution  $f(x, \Theta)$ , alors la fonction de vraisemblance est

$$f(X_1, X_2, \dots, X_N, \Theta) = \prod_{t=1}^N f(X_t, \Theta)$$

L'estimateur EMV est obtenu en maximisant cette fonction. Le problème dans les séries temporelles est que les  $\{X_t\}$  ne sont pas indépendantes, On peut néanmoins écrire la loi conjointe  $f(X_1, X_2, \dots, X_N, \Theta)$  à l'aide de densité conditionnelle.

$$\begin{aligned} f(X_1, \dots, X_N) &= f(X_N | \mathbb{X}_{N-1}) f(X_1, X_2, \dots, X_{N-1}) \\ &= f(X_N | \mathbb{X}_{N-1}) f(X_{N-1} | \mathbb{X}_{N-2}) f(X_1, X_2, \dots, X_{N-2}) \end{aligned}$$

où  $\mathbb{X} = (X_1, X_2, \dots, X_n)$ , on obtient  $f(X_1, \dots, X_N) = \prod_{t=2}^N f(X_t | \mathbb{X}_{t-1}) f(X_1)$ , produit des densités conditionnelles. la méthode de MV est applicable si le modèle considéré nous permet de calculer ces densités.

En effet, c'est le cas des modèles ARMA sous certaines conditions. Par exemple dans le cas où AR( $p$ ) est gaussien. En particulier, soit  $\varepsilon_t \in N(0, \sigma^2)$ , et que  $X_0$  est gaussienne, alors la densité conditionnelle d'une v. a. gaussienne sachant une autre variable gaussienne est une densité gaussienne\*.

### Exemples

On étudie maintenant quelques exemples.

**Exemple 4.2.1** Soit un modèle AR(1) gaussien

$$X_t = c + \phi X_{t-1} + \varepsilon_t \quad \varepsilon_t \text{ i.i.d. et } \varepsilon_t \in N(0, \sigma^2)$$

Les paramètres à estimer sont  $\Theta = (c, \phi, \sigma^2)$ . Calculer son EMV.

**Sol** La méthode de MV consiste donc à

- Evaluer la fonction de vraisemblance

$$f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N, \Theta)$$

- Trouver l'estimateur qui maximise cette fonction.

Pour commencer, nous considérons la distribution de  $X_1$ , d'après les résultats du chapitre 3 nous avons

$$\mu = \mathbf{E} X_t = \frac{c}{1-\phi}, \quad \mathbf{Var} X_t = \frac{\sigma^2}{1-\phi^2}$$

donc  $X_1$  est gaussienne,  $X_1 \sim N\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right)$ :

$$f_{X_1}(x_1, \Theta) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2/(1-\phi^2)}} \exp\left(-\frac{(x_1 - \frac{c}{1-\phi})^2}{2\sigma^2/(1-\phi^2)}\right)$$

Considérons maintenant la distribution de  $X_2$  sachant  $X_1 = x_1$ ,  $X_2 = c + \phi X_1 + \varepsilon_2$ . le conditionnement  $X_1 = x_1$  est traité comme si  $X_1$  est une constante déterministe donc la distribution conditionnelle est  $(X_2 | X_1 = x_1) \sim N(c + \phi x_1, \sigma^2)$ , c'est à dire

$$f_{X_2|X_1}(x_2|x_1, \Theta) = f(x_1, x_2, \Theta|x_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_2 - c - \phi x_1)^2}{2\sigma^2}\right)$$

---

\*C'est une propriété très importante pour les vecteurs gaussiens, c'est en partie grâce à cette propriété, on a trouvé le filtre de Kalman, voir le chapitre 7.

d'une manière récurrente, la distribution de  $X_N$  sachant  $X_1 = x_1, X_2 = x_2, \dots, X_{N-1} = x_{N-1}$  est

$$(X_N | X_1 = x_1, X_2 = x_2, \dots, X_{N-1} = x_{N-1}, \Theta) \sim N(c + \phi x_{N-1}, \sigma^2)$$

$$f_{X_N | X_1, \dots, X_{N-1}}(x_N | x_1, \dots, x_{N-1}, \Theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_N - c - \phi x_{N-1})^2}{2\sigma^2}\right)$$

car l'influence de  $(X_1 = x_1, X_2 = x_2, \dots, X_{N-1} = x_{N-1})$  sur  $X_N$  est faite seulement via  $X_{N-1} = x_{N-1}$ , et

$$X_N = c + \phi X_{N-1} + \varepsilon_N$$

Finalement, la formule  $f(X_1, X_2, \dots, X_N, \Theta) = \prod_{t=1}^N f(X_t | \mathbb{X}_{t-1}) \cdot f(X_1)$  se traduit par

$$\begin{aligned} f(x_1, x_2, \dots, x_N, \Theta) &= f_{X_1}(x_1, \Theta) \prod_{t=2}^N f_{X_t | X_{t-1}}(x_t | x_{t-1}, \Theta) \\ L(x_1, x_2, \dots, x_N, \Theta) &= f_{X_1}(x_1, \Theta) \prod_{t=2}^N f_{X_t | X_{t-1}}(x_t | x_{t-1}, \Theta) \\ \ln L(x_1, x_2, \dots, x_N, \Theta) &= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln(\sigma^2 / 1 - \phi^2) - \frac{[x_1 - (c/1 - \phi^2)]^2}{2\sigma^2 / 1 - \sigma^2} \\ &\quad - \frac{N-1}{2} \ln 2\pi - \frac{N-1}{2} \ln \sigma^2 - \sum_{t=2}^N \left( \frac{(x_t - c - \phi x_{t-1})^2}{2\sigma^2} \right) \end{aligned}$$

La distribution  $f(x_1, \Theta)$  prend une place importante dans l'expression  $L(\mathbf{x}_N, \Theta)$ . Souvent quand  $N$  est grand, on peut négliger  $X_1$  en fixant  $x_1$  : c'est à dire on considère  $X_1 = x_1$  comme connu et  $f_{X_1}(x_1, \Theta)$  est simplifiée par  $f_{X_1}(x_1, \Theta) = 1$ , alors la formule  $\ln L$  est beaucoup plus facile à évaluer. Maximiser la fonction  $\ln L(x_1, x_2, \dots, X_N, \Theta)$  par rapport à  $(c, \phi)$  en considérant  $\sigma^2$  comme connu est équivalent à minimiser la fonction

$$\sum_{t=2}^N (x_t - c - \phi x_{t-1})^2$$

Donc l'EMV pour AR(1) est (*exercice*)

$$\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} N-1 & \sum X_{t-1} \\ \sum X_{t-1} & \sum X_{t-1}^2 \end{pmatrix} \begin{pmatrix} \sum X_t \\ \sum X_{t-1} X_t \end{pmatrix} \quad \text{où } \sum = \sum_{t=2}^N$$

En faisant  $\frac{\partial}{\partial \sigma^2} \ln L(x_1, \dots, x_n, \sigma^2) = 0$  on obtient

$$-\frac{T-1}{2\sigma^2} + \sum_{t=2}^N \left( \frac{(x_t - c - \phi x_{t-1})^2}{2\sigma^4} \right) = 0$$

donc  $\hat{\sigma}^2 = \sum_{t=2}^N \left( \frac{(x_t - \hat{c} - \hat{\phi} x_{t-1})^2}{N-1} \right)$  (*exercice*)

**Exemple 4.2.2** On considère le modèle  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ , où  $\varepsilon_t$  est un bruit blanc gaussien, alors la densité conditionnelle de  $X_t$  sachant  $X_1, \dots, X_{t-1}$  est normale, la moyenne  $\mathbf{E}(X_t | \mathbb{X}_{t-1})$  est égale à  $\phi_1 X_{t-1} + \phi_2 X_{t-2}$  et la variance est

$$\mathbf{E}(X_t - (\phi_1 X_{t-1} + \phi_2 X_{t-2}))^2 = \mathbf{E}\varepsilon_t^2 = \sigma^2$$

c'est à dire

$$f(X_t | X_1, X_2, \dots, X_{t-1}, \Theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2}{2\sigma^2}\right\}$$

Montrer que l'EMV de  $\Theta = (\phi_1, \phi_2, \sigma^2)$  est

$$\begin{cases} (\hat{\phi}_1, \hat{\phi}_2) = \arg \min \sum_{t=3}^N (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2 \\ \hat{\sigma}^2 = \frac{1}{N-3} \sum_{t=3}^N (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2})^2 \end{cases}$$

**Sol** Etant donné que  $X_{-1}$ ,  $X_{-2}$  et  $X_0$  sont inconnus, et que les densités  $f(X_1)$ ,  $f(X_2|X_1)$  sont fixées à 1, on considère

$$L(\mathbb{X}_N, \Theta) = \prod_{t=3}^N f(X_t | \mathbb{X}_{t-1})$$

$$\ln L(\mathbb{X}_N, \phi_1, \phi_2) = -\frac{N-2}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=3}^N (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2$$

En utilisant la même méthode que l'exemple 4.2.1, on obtient le résultat.  $\square$

**Exemple 4.2.3** *Modèle AR(p) :  $X_t = c - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} + \varepsilon_t$ . Les paramètres sont  $\Theta = (c, \phi_1, \dots, \phi_p, \sigma^2)$ .*

**Sol** Maximiser la fonction de vraisemblance pour un AR(p) peut être réalisé par des méthodes numériques, on cherche alors à maximiser la fonction

$$\begin{aligned} & \ln f_{X_N, X_{N-1}, \dots, X_{p+1} | X_p, X_{p-1}, \dots, X_1} (x_N, x_{N-1}, \dots, x_{p+1} | x_p, \dots, x_1, \Theta) \\ &= -\frac{N-p}{2} \ln(2\pi\sigma^2) - \sum_{t=p+1}^N \frac{(x_t - c - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p})^2}{2\sigma^2} \end{aligned}$$

d'où

$$\left\{ \begin{array}{l} \hat{\sigma}^2 = \frac{1}{N-p} \sum_{t=p+1}^N (X_t - \hat{c} - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \dots - \hat{\phi}_p X_{t-p})^2 \\ (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p, \hat{c}) = \arg \min \sum_{t=p+1}^N (X_t - c - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2 \end{array} \right.$$

$\square$

**Exemple 4.2.4** *Calculer la fonction de vraisemblance pour un MA(1).*

**Sol** Soit

$$X_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \quad \text{où } \varepsilon_t \sim N(0, \sigma^2) \text{ i.i.d.}$$

Les paramètres à estimer sont  $\Theta = (\mu, \theta, \sigma^2)$ . Dans le cas gaussien

$$X_t | \varepsilon_{t-1} \sim N(\mu + \theta \varepsilon_{t-1}, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_t - \mu - \theta \varepsilon_{t-1})^2}{2\sigma^2} \right\}$$

On suppose que  $\varepsilon_0 = 0$  alors

$$X_1 | \varepsilon_0 = 0 \sim N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_1 - \mu)^2}{2\sigma^2} \right\} \text{ et } \varepsilon_1 = x_1 - \mu$$

$$X_2 | X_1, \varepsilon_0 = 0 \sim f_{X_2 | X_1, \varepsilon_0 = 0}(x_2 | x_1, \varepsilon_0, \Theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_2 - \mu - \theta \varepsilon_1)^2}{2\sigma^2} \right\}$$

$$\text{et } \varepsilon_2 = x_2 - \mu - \theta \varepsilon_1$$

... ..

$$X_t | X_{t-1}, \dots, X_1, \varepsilon_0 = 0 \sim f_{X_t | X_{t-1}, \dots, X_1, \varepsilon_0 = 0}(x_t | x_{t-1}, \dots, x_1, \varepsilon_0, \Theta) = f_{X_t | \varepsilon_{t-1}}(x_t | \varepsilon_{t-1}, \Theta)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_t - \mu - \theta \varepsilon_{t-1})^2}{2\sigma^2} \right\} = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{\varepsilon_t^2}{2\sigma^2} \right\}$$

Donc la fonction de vraisemblance est

$$f(X_N, X_{N-1}, \dots, X_1 | \varepsilon_0 = 0) = \left( \frac{1}{\sqrt{2\pi}} \right)^N \left( \frac{1}{\sqrt{\sigma^2}} \right)^N \exp \left\{ \sum_{t=1}^N \frac{\varepsilon_t^2}{2\sigma^2} \right\} \text{ et } \varepsilon_t = x_t - \mu - \theta \varepsilon_{t-1}$$

Les  $\varepsilon_t$  sont calculées de manière itérative.  $\square$

On remarque que contrairement aux problèmes 4.2.1, 4.2.2 et 4.2.3 où on peut calculer les EMV  $\hat{\Theta} = \arg \min \ln L(\mathbf{x}_N, \Theta)$  par des méthodes directes, dans l'exemple 4.2.4, on ne pourra pas le faire. Pour cela on présentera dans la section suivante une méthode implicite itérative.

**Exemple 4.2.5** *Processus* MA( $q$ ),  $\Theta = (\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)$ .

**Sol**  $\ln L(\Theta)$  est

$$\ln L(\mathbb{X}_N, \Theta) = \ln f(X_N, X_{N-1}, \dots, X_1 | \vec{\varepsilon}_0 = 0, \Theta) = -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{t=1}^N \frac{\varepsilon_t^2}{2\sigma^2}$$

Où le  $q$ -vecteur  $\vec{\varepsilon}_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1})$  et les  $\{\varepsilon_t\}$  sont obtenues par

$$\varepsilon_t = X_t - \mu - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

□

**Exemple 4.2.6** *Processus* ARMA( $p, q$ ). *Soit*

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$$\Theta = (c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)$$

**Sol** Soit

$$\left. \begin{aligned} \vec{\varepsilon}_0 &= (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1}) \\ \vec{x}_0 &= (x_0, x_{-1}, \dots, x_{-p+1}) \end{aligned} \right\} \text{condition initiale}$$

On peut calculer successivement pour  $t = 1, 2, \dots, N$

$$\varepsilon_t = x_t - c - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Les conditions initiales sont choisies :

$$\vec{\varepsilon}_0 = \overbrace{(0, 0, \dots, 0)}^p$$

$$\vec{x}_0 = \left( \overbrace{\left( \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}, \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}, \dots, \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \right)}^q \right)$$

Elles représentent espérances mathématiques de  $\vec{\varepsilon}_0$  et  $\vec{x}_0$ . On en déduit alors

$$\ln L(\mathbb{X}_N, \Theta) = f(X_N, X_{N-1}, \dots, X_1 | \vec{x}_0, \vec{\varepsilon}_0, \Theta) = -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{t=1}^N \frac{\varepsilon_t^2}{2\sigma^2}$$

□

Box et Jenkin suggèrent une formulation légèrement différente : les  $\varepsilon_t$  ne sont calculées qu'à partir de  $t = p+1, p+2, \dots, N$  avec  $\varepsilon_{p-q+1} = \varepsilon_{p-q} = \dots = \varepsilon_{p-1} = \varepsilon_p = 0$  et  $(x_1, x_2, \dots, x_p)$  ( $p$  valeurs observée) comme conditions initiales, pour  $t = p+1, \dots, N$

$$\varepsilon_t = x_t - c - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

La fonction  $\ln L(\Theta)$  devient

$$\begin{aligned} \ln L(\mathbb{X}_N, \Theta) &= f(X_N, X_{N-1}, \dots, X_1 | X_p, X_{p-1}, \dots, X_1, \varepsilon_p = \varepsilon_{p-1} = \dots = \varepsilon_{p-q+1} = 0, \Theta) \\ &= -\frac{N-p}{2} \ln(2\pi\sigma^2) - \sum_{t=p+1}^N \frac{\varepsilon_t^2}{2\sigma^2} \end{aligned}$$

□

### 4.3 Méthode numérique

On présente dans cette section une méthode itérative pour trouver l'EMV. Supposons qu'on dispose comme dans la figure 4.6 d'une boîte noire *subroutine*), capable de calculer à chaque valeur de  $\Theta$  et les observations  $(x_1, x_2, \dots, x_N)$  la fonction  $\ln L(\mathbb{X}_N, \Theta)$  correspondante. L'objectif est de trouver une suite  $\Theta^m$  qui converge vers  $\hat{\Theta}$  maximisant  $\mathcal{L}(\Theta) \triangleq \ln L(\Theta)$ .

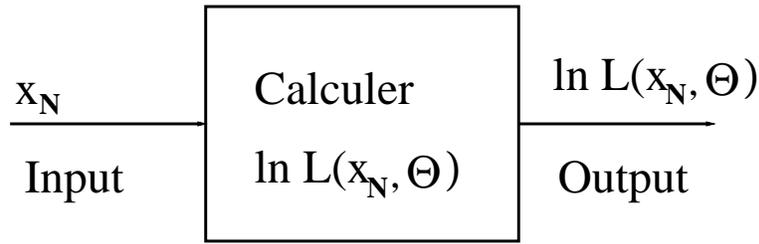


Figure 4.1: Boîte noire pour calculer  $\ln L(\mathbf{x}_N, \Theta)$

### 4.3.1 Méthode de recherche par grille

C'est la méthode la plus simple, on établit d'abord une grille (qui discrétise le domaine admissible de  $\Theta$ ), on teste ensuite point par point de ce domaine, cela permet d'identifier  $\arg \max \mathcal{L}(\Theta)$ . En pratique

- Il faut que la *grille* soit assez fine, pour capter le comportement de  $\mathcal{L}(\Theta)$  par rapport à  $\Theta$ .
- Raffiner sur la zone où  $\arg \max$  se trouve.
- Fixer un critère d'arrêt pour le raffinement.

**Exemple 4.3.1** En  $\dim = 1$ , AR(1) :  $X_t = c + \phi X_{t-1} + \varepsilon_t$ , on suppose que  $c = 0$ ,  $\sigma^2 = 1$  est connu. Alors selon l'exemple 4.2.1

$$\ln L(\phi) = -\frac{N}{2} \ln(1 - \phi^2) - \frac{1}{2}(1 - \phi^2)x_1^2 - \frac{1}{2} \sum_{t=2}^N (x_t - \phi x_{t-1})^2$$

Soit  $\mathbf{x}_N = (x_1 = 0.8, x_2 = 0.2, x_3 = -1.2, x_4 = -0.4, x_5 = 0.0)$ , La fonction  $\mathcal{L}(\phi)$  est illustrée par la figure 4.3

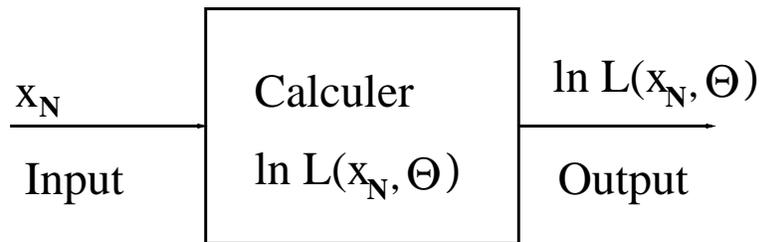


Figure 4.2: Méthode de recherche par grille

Cette méthode devient très coûteuse lorsque  $\dim \geq 2$ .

### 4.3.2 méthode de descente gradient

Pour maximiser la fonction  $\mathcal{L}(\Theta)$ , on définit

$$\begin{aligned} g(\Theta) &= (g_1(\Theta), g_2(\Theta), \dots, g_a(\Theta)) = \frac{\partial \mathcal{L}(\Theta)}{\partial \Theta} = D_{\Theta} \mathcal{L}(\Theta) \\ &= \text{grad}_{\Theta} \mathcal{L}(\Theta) \end{aligned}$$

La méthode consiste donc à calculer successivement la suite

$$\Theta^{m+1} = \Theta^m + s \cdot g(\Theta^m). \quad s : \text{paramètre de contrôle}$$

Si on ne dispose pas de  $g(\Theta)$  théorique, on se contentera d'avoir une approximation<sup>†</sup> :

$$\begin{aligned} g_i(\Theta^m) &= \frac{1}{\Delta} \{ \mathcal{L}(\Theta_1^m, \Theta_2^m, \dots, \Theta_{i-1}^m, \Theta_i^m + \Delta, \Theta_{i+1}^m, \dots, \Theta_a^m) \\ &\quad - \mathcal{L}(\Theta_1^m, \Theta_2^m, \dots, \Theta_{i-1}^m, \Theta_i^m, \Theta_{i+1}^m, \dots, \Theta_a^m) \} \end{aligned}$$

<sup>†</sup>gradient numérique

Pour  $i = 1, \dots, a$ . La valeur de  $s$  est choisie de la façon suivante : on test  $s = \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8$  etc (selon problème), on prendra la valeur pour laquelle  $\mathcal{L}(\Theta^m + sg(\Theta^m))$  est maximale, quand  $\Theta^m$  est proche de la solution finale, il faut alors raffiner encore plus. Enfin, pour éviter de tomber dans des maximums locaux, on effectue ces calculs avec plusieurs valeurs initiales distincts.

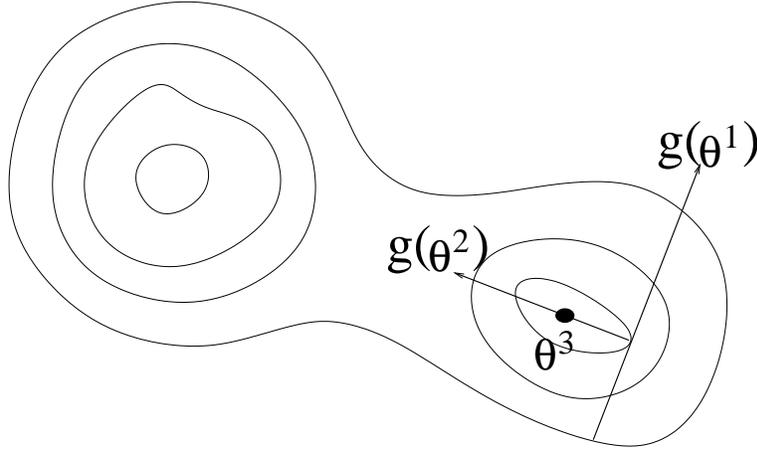


Figure 4.3: Méthode de descente gradient

### 4.3.3 Méthode de Newton

Soit  $H(\Theta^m)$  une matrice définie par

$$H(\Theta^m) = - \left. \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \Theta \partial \Theta} \right|_{\Theta = \Theta^m}$$

La méthode de Newton consiste à itérer

$$\Theta^{m+1} = \Theta^m + [H(\Theta^m)]^{-1} g(\Theta^m)$$

On remarque que si la fonction  $\mathcal{L}(\Theta)$  est une forme quadratique, alors une seule itération suffit :  $\Theta^1 = \text{EMV}$ . En générale, on utilise la formule

$$\Theta^{m+1} = \Theta^m + s[H(\Theta^m)]^{-1} g(\Theta^m) \quad s : \text{paramètre de contrôle}$$

## 4.4 Identification

### 4.4.1 Choix de $p$ et $q$ dans un processus AR( $p$ ) et MA( $q$ )

Comme signalé dans l'introduction, les indicateurs pour le choix de  $p$  et  $q$  sont essentiellement la forme des fonctions d'autocorrélation partielle  $\Phi(k)$  et les corrélogrammes  $r_k$ . On peut penser par exemple qu'une chute de  $\Phi(k)$  vers de valeurs proches de zéro pour  $k > p$  indique un AR( $p$ ), ou qu'une chute  $r_k$  vers des valeurs proches de zéro pour  $k > q$  indique un MA( $q$ ). Pour juger si les  $\Phi(k)$  ou  $r_k$  sont significativement différents de zéro, il est bon d'avoir une idée de leur écart-type. On peut utiliser deux résultats

1. Pour un MA( $q$ ), la variance de  $r_k$  vérifie

$$\forall k > q, \quad \text{Var } r_k \cong \frac{1}{T} \left[ 1 - 2 \sum_{i=1}^q r_k^2 \right]$$

2. Pour un AR( $p$ ), la variance de  $\Phi(k)$  vérifie

$$\forall k > p, \quad \text{Var } \Phi(k) \cong \frac{1}{T}$$

L'application de ces résultats sont assez simple, par exemple pour identifier l'ordre  $p$  d'un processus autorégressif, on peut calculer la suite des autocorrélations partielles estimées  $\Phi(k)$  et examiner à partir de quelle valeur elle reste dans la bande  $\left[ -\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right]$ .

Ainsi dans l'exemple ci-dessus, on peut considérer  $p = 3$ . Cette démarche consiste donc à tester la nullité des autocorrélations partielles  $\Phi(k)$  à partir d'un certain rang.

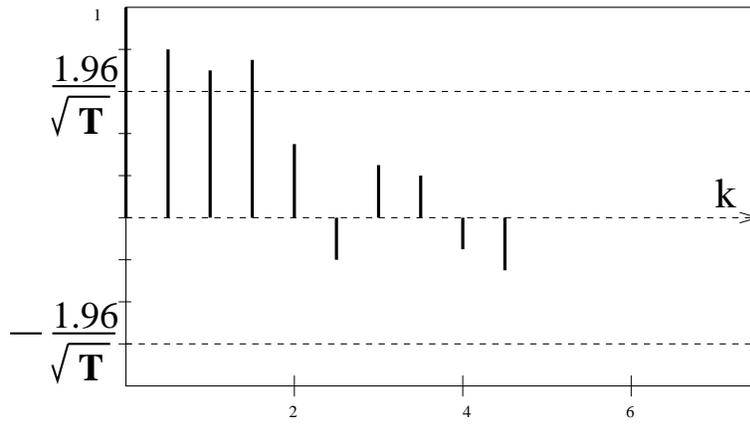


Figure 4.4: Corrélogramme des autocorrélations partielles

#### 4.4.2 Choix de $p$ et $q$ dans le cas d'un processus ARMA( $p, q$ )

Nous présentons la méthode de coin en admettant le résultat suivant

**Théorème 4.4.1** *Un processus ARMA( $p, q$ ) stationnaire satisfait la propriété suivante (propriété du coin).*

$$\begin{cases} \Delta(i, j) = 0 & i \geq p, j \geq q \\ \Delta(i, q-1) \neq 0 & i \geq p-1 \\ \Delta(p-1, j) \neq 0 & j \geq q-1 \end{cases}$$

avec

$$\Delta(i, j) = \det \begin{vmatrix} \gamma(j+1) & \gamma(j+2) & \cdots & \gamma(j+i+1) \\ \gamma(j) & \gamma(j+1) & \cdots & \gamma(j+i) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma(j+1-i) & \gamma(j-i) & \cdots & \gamma(j+1) \end{vmatrix}_{(i+1) \times (i+1)}$$

Les valeurs des  $\Delta(i, j)$  sont inconnues, elles peuvent être estimées en remplaçant dans le déterminant les corrélations  $\gamma(k)$  par leurs estimateurs  $r_k$ . Il est vraisemblable que dans le tableau dont les éléments sont de  $\Delta(i, j)$  on verra apparaître une rupture entre les lignes  $i = p-1$  et  $i = p$ , et entre les colonnes  $j = q-1$  et  $j = q$ .

En fait, ces nullités sont vérifiées par des tests statistiques (*test student* ou *test chi-2*). En pratique, à l'issue de ces tests, un ensemble de modèles correspondant aux différents choix de  $(p, q)$  seront acceptés. On peut par exemple accepter simultanément plusieurs modèles: MA(3), ARMA(1, 2), ARMA(3, 1) et AR(5). Après cet étape de identification (*dite Identification à priori*), on procède alors aux estimations des paramètres de chacun de ces modèles, une deuxième étape d'identification (*Identification a posteriori*) est alors effectuée afin de choisir le ou les meilleurs modèles. Le schéma de cette procédure peut être résumé par l'organigramme 4.5 ci-dessous

## 4.5 Prédiction

### 4.5.1 Principe de prédiction

On s'intéresse à la prédiction de la v.a.  $X_{t+1}$  basée sur le vecteur  $\mathbb{X}_t = (X_t, X_{t-1}, \dots, X_{t-m+1})$ . On désigne  $X_{t+1|t}^*$  la prédiction de  $X_{t+1}$  basée sur l'observation  $\mathbb{X}_t$ . Il faut d'abord mesurer la qualité d'une prédiction.

**Définition 4.5.1 Erreur quadratique moyenne (MSE)** *On appelle MSE (erreur quadratique moyenne) d'une prédiction  $X_{t+1|t}^*$  l'espérance*

$$\text{MSE } X_{t+1|t}^* \triangleq \mathbf{E} (X_{t+1} - X_{t+1|t}^*)^2$$

L'objectif est alors de minimiser cette erreur.

**Proposition 4.5.2** *La meilleure prédiction  $X_{t+1|t}^*$  au sens de MSE est l'espérance conditionnelle de  $X_{t+1}$  sachant  $\mathbb{X}_t$ .*

$$X_{t+1|t}^* = \mathbf{E} (X_{t+1} | \mathbb{X}_t)$$

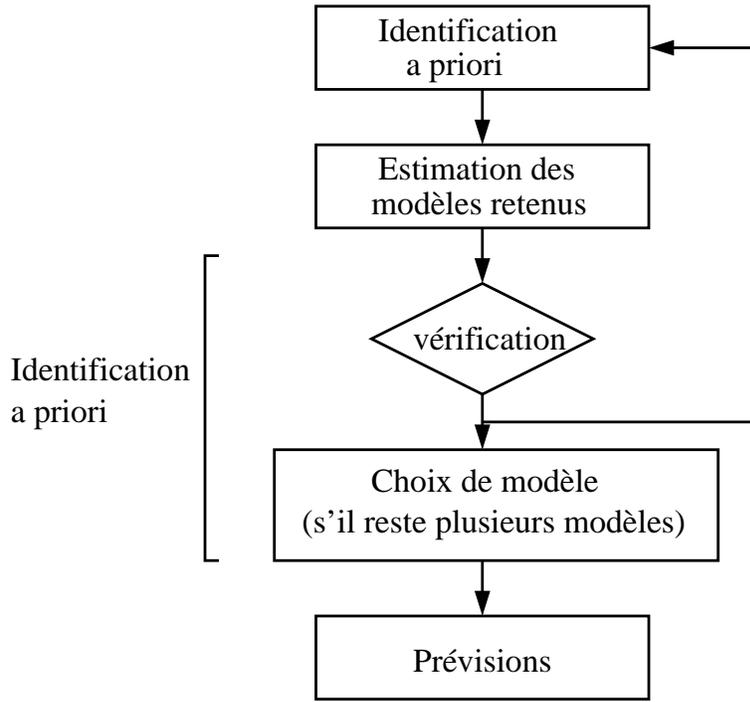


Figure 4.5: Schéma : Identification, estimation, prévision

**Preuve**<sup>‡</sup> Soit  $g(\mathbb{X}_t)$  une prévision quelconque  $X_{t+1|t}^* = g(\mathbb{X}_t)$  alors MSE correspondante est

$$\begin{aligned}
 \mathbf{E} (X_{t+1} - g(\mathbb{X}_t))^2 &= \mathbf{E} [X_{t+1} - \mathbf{E}(X_{t+1}|\mathbb{X}_t) + \mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)]^2 \\
 &= \mathbf{E} [X_{t+1} - \mathbf{E}(X_{t+1}|\mathbb{X}_t)]^2 \\
 &\quad + 2\mathbf{E} \{ [X_{t+1} - \mathbf{E}(X_{t+1}|\mathbb{X}_t)] [\mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)] \} \\
 &\quad + \mathbf{E} [\mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)]^2 \\
 &= \mathbf{E} I_1^2 + \mathbf{E} I_2 + \mathbf{E} I_3^2
 \end{aligned}$$

or  $\mathbf{E} I_2 = \mathbf{E} (\mathbf{E}(I_2|\mathbb{X}_t))$  et

$$\mathbf{E} (I_2|\mathbb{X}_t) = [\mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)] \mathbf{E} \{ [X_{t+1} - \mathbf{E}(X_{t+1}|\mathbb{X}_t)] | \mathbb{X}_t \}$$

car  $\mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)$  est  $\mathbb{X}_t$  mesurable<sup>§</sup>, donc  $\mathbf{E} I_2 = [\mathbf{E}(X_{t+1}|\mathbb{X}_t) - g(\mathbb{X}_t)] \times 0 = 0$ .

$\mathbf{E} I_1^2 \geq 0$  ne dépend pas du choix de  $g$ , et  $\mathbf{E} I_3^2 \geq 0$  atteint le minimum avec  $g(\mathbb{X}_t) = \mathbf{E}(X_{t+1}|\mathbb{X}_t)$  donc la meilleure prévision est  $\mathbf{E}(X_{t+1}|\mathbb{X}_t)$ .  $\square$

## 4.5.2 Espace $L^2$ , régression affine

**Définition 4.5.3**  $X \in L^2(\Omega, \mathcal{A}, P)$  (processus carrée intégrable) si  $\|X_t\|_2 = (\int |x|^2 dP)^{\frac{1}{2}} < \infty$ .

Il s'agit d'un espace Hilbert muni de produit scalaire  $\langle \cdot, \cdot \rangle$ , et la norme  $\|\cdot\|_2$  où  $\langle X, Y \rangle = \mathbf{E} XY$  et  $\|X\|^2 = \mathbf{E} X^2 = \mathbf{Var} X + (\mathbf{E} X)^2$ .

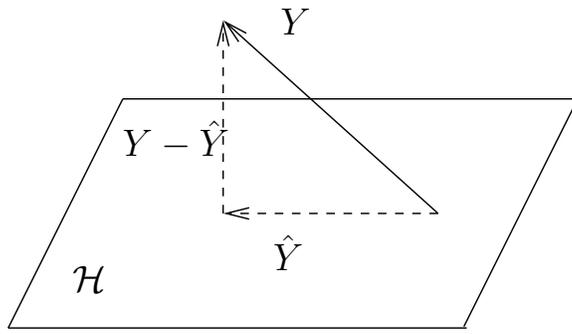
**Théorème 4.5.4 théorème de projection** Si  $\mathcal{H}$  est un sous espace fermé de  $L^2$  pour tout  $y \in L^2$ , il existe une unique v.a.  $\hat{y} \in \mathcal{H}$  t.q.

$$\|y - \hat{y}\| = \min_{H \in \mathcal{H}} \|y - H\|$$

caractérisé par  $\hat{Y} \in \mathcal{H}$  et  $y - \hat{y} \in \mathcal{H}^\perp$  (i.e.:  $\langle Y - \hat{Y}, H \rangle = 0, \forall H \in \mathcal{H}$ ).

<sup>‡</sup>On trouve un résultat similaire Proposition 7.3.3 dans le chapitre 7

<sup>§</sup> $\mathbf{E}(XY|\xi) = X\mathbf{E}(Y|\xi)$  si  $X$  est  $\xi$  mesurable



### Régression affine dans $L^2$

Soit  $X_1, X_2, \dots, X_n$   $n$  v.a.  $Y$  une v.a. dans  $L^2$ , alors la régression de  $Y$  sur  $X_1, \dots, X_n$  est la projection orthogonale dans  $L^2$  de  $Y$  sur  $\mathcal{H}^* = \text{Vect}(X_1, \dots, X_n)$  et la régression affine de  $Y$  sur  $X_1, \dots, X_n$  est la projection orthogonale dans  $L^2$  de  $Y$  sur  $\mathcal{H} = \text{Vect}(1, X_1, X_2, \dots, X_n)$  et on note  $\hat{Y} = \mathbf{E}L(Y|1, X_1, X_2, \dots)$ .

**Exercice 4.5.5** Soit  $X_1, \dots, X_n$  un vecteur,  $X_i \in L^2$  et  $\mathcal{H} = \text{Vect}(1, X_1, \dots, X_n)$  espace combinaison linéaire affine de  $X_1, \dots, X_n$ . Soit  $X = X_0 \in L^2$ ,

$$\gamma = \mathbf{Cov}(X, X_i)_{i=1, \dots, n}, \quad \sum = \text{matrice } \mathbf{Cov}(X_i, X_j)_{i, j=1, 2, \dots, n}$$

Alors

$$\hat{X} = \mathbf{E}L(X|1, X_1, \dots, X_n) = a_0 + a_1 + \dots + a_n$$

avec  $\mathbf{a} = (a_1, \dots, a_n)$  vérifiant  $\mathbf{a} = \sum^{-1} \gamma$

Solution : On suppose  $\hat{X} = a_0 + a_1 + \dots + a_n$ ,  $a_i$  à déterminer. On sait que  $\hat{X} \in \mathcal{H}$ , et  $X - \hat{X} \in \mathcal{H}^\perp$  cad

$$\langle X - \hat{X}, 1 \rangle = 0 \quad (4.2)$$

$$\langle X - \hat{X}, X_i \rangle = 0, \quad (4.3)$$

avec 4.2 on a  $\mathbf{E}X = \mathbf{E}\hat{X} = a_0 + a_1\mathbf{E}X_1 + \dots + a_n\mathbf{E}X_n$  avec 4.3 on a  $\mathbf{E}XX_i = \mathbf{E}\hat{X}X_i = a_0\mathbf{E}X_i + a_1\mathbf{E}X_1X_i + \dots + a_n\mathbf{E}X_nX_i$  donc

$$\mathbf{E}XX_i = \mathbf{E}X\mathbf{E}X_i - (a_1\mathbf{E}X_1 + \dots + a_n\mathbf{E}X_n)\mathbf{E}X_i \quad \begin{matrix} a_0 = \mathbf{E}X - a_1\mathbf{E}X_1 - \dots - a_n\mathbf{E}X_n \\ + a_1\mathbf{E}X_1X_i + \dots + a_n\mathbf{E}X_nX_i \end{matrix}$$

$\forall i = 1, \dots, n$  nous avons  $\mathbf{Cov}(X, X_i) = a_1\mathbf{Cov}(X_1, X_i) + a_2\mathbf{Cov}(X_2, X_i) + \dots + a_n\mathbf{Cov}(X_n, X_i)$

$$\left\{ \begin{array}{l} \mathbf{Cov}(X, X_1) = a_1\mathbf{Cov}(X_1, X_1) + a_2\mathbf{Cov}(X_2, X_1) + \dots + a_n\mathbf{Cov}(X_n, X_1) \\ \mathbf{Cov}(X, X_2) = a_1\mathbf{Cov}(X_1, X_2) + a_2\mathbf{Cov}(X_2, X_2) + \dots + a_n\mathbf{Cov}(X_n, X_2) \\ \dots \\ \mathbf{Cov}(X, X_n) = a_1\mathbf{Cov}(X_1, X_n) + a_2\mathbf{Cov}(X_2, X_n) + \dots + a_n\mathbf{Cov}(X_n, X_n) \end{array} \right.$$

$\gamma = \sum \mathbf{a}$  sous forme matricielle, d'où le résultat.

### Prévision linéaire

**Définition 4.5.6** Soit  $(X_t)_{t \in \mathbb{Z}}$  un processus  $L^2$ , on appelle meilleur prévision linéaire de  $X_t$  sachant son passé la régression linéaire de  $X_t$  sur son passé  $\mathcal{H} = \text{Vect}(1, X_{t-1}, X_{t-2}, \dots)$

**Définition 4.5.7** processus d'innovation est le processus

$$\eta_t = X_t - \mathbf{E}L(X_t|1, X_{t-1}, X_{t-2}, \dots)$$

Exemple : Soit  $Y_t \sim \text{BBG}(0, \sigma^2)$  le processus  $X_t$  est défini par

$$X_t = Y_t - \rho Y_{t-1}, \quad |\rho| < 1$$

Alors  $\mathbf{E}L(X_t|1, X_{t-1}, X_{t-2}, \dots) = -\sum_{i=1}^{\infty} \rho^i X_{t-i}$  est la régression linéaire de  $X_t$  sur son passé. et en plus, le processus d'innovation est  $Y_t$ .

Solution : Il est évident que  $X_t$  est stationnaire. Soit  $S_{t,n} = \sum_{i=1}^n \rho^i X_{t-i}$  alors  $X_t = Y_t - \rho Y_{t-1}$ , ajouter  $S_{t,n} = \sum_{i=1}^n$  à deux côtés et passer à la limite

$$\begin{aligned} X_t + S_{t,n} &= Y_t - \rho Y_{t-1} + \rho X_{t-1} + \rho^2(X_{t-2} + \dots + \rho^n X_{t-n}) \\ &= Y_t - \rho Y_{t-1} + \rho(Y_{t-1} - \rho Y_{t-2}) + \rho^2(Y_{t-2} - \rho Y_{t-3}) + \dots + \rho^n(Y_{t-n} - \rho Y_{t-n-1}) \\ &= Y_t + \rho^{n+1} Y_{t-n-1} \\ X_t + \sum_{i=1}^{\infty} \rho^i X_{t-i} &= Y_t \text{ car } |\rho| < 1 \end{aligned}$$

or la suite  $S_{t,n}$  est convergente :  $\|S_{t,n} - S_{t,m}\| = \|\sum_{i=m+1}^n \rho^i X_{t-i}\| \leq \|X_t\| \sum_{i=m+1}^n |\rho^i| \rightarrow 0$  quand  $n \rightarrow \infty$  et  $m \rightarrow \infty$ , c'est une suite de Cauchy, donc convergente. Soit  $S_{t,n} \rightarrow S_t$  on a  $X_t + S_t = Y_t$ . Maintenant  $\langle X_s, Y_t \rangle = 0, \forall s < t$  et  $\langle 1, Y_t \rangle = 0$  ( $Y_t$  est un BB!) et  $S_t = \mathbf{E}L(X_t | 1, X_{t-1}, \dots)$  donc tous les conditions sont réunies pour que  $Y_t$  soit un processus d'innovation de  $(X_1, X_2, \dots, X_n)$ .

## 4.6 Prédiction des modèles ARIMA, méthode de Box-Jenkins

Soit  $(X_1, X_2, \dots, X_T)$  une série stationnaire, on cherche à faire prédiction à horizon  $h$ . Dans un premier temps, les modèles AR, MA et ARMA sont supposés mis sous forme canonique, et n'avoir aucune racine unité, et tous les racines de  $\Phi(z)$  et  $\Theta(z)$  sont de modules strictement supérieur à 1. On va considérer successivement le cas AR(p), MA(q) et ARMA(p,q)

### 4.6.1 Prédiction de AR(p)

C'est pas le facile à faire.  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \Phi(B)X_t = \varepsilon_t$ . Donc la prédiction optimale à horizon 1, faite à la date  $T$  est

$$X_{T+1|T}^* = \mathbf{E}L(X_{T+1} | \mathbb{X}_T) = \phi_1 X_T + \dots + \phi_p X_{T-p}$$

car  $\varepsilon_t$  est l'innovation. De la même façon,  $X_{T+h} = \phi_1 X_{T+h-1} + \phi_2 X_{T+h-2} + \dots + \phi_n X_{T+h-n}$  alors  $X_{T+h|T}^* = \mathbf{E}L(X_{T+h} | \mathbb{X}_T)$  est donné de façon récursive par

$$X_{T+h|T}^* = \begin{cases} \phi_1 X_{T+h-1|T}^* + \dots + \phi_{h-1} X_{T+1|T}^* + \phi_h X_T + \dots + \phi_p X_{T+h-p} & \text{pour } h \leq p \\ \phi_1 X_{T+h-1|T}^* + \dots + \phi_p X_{T+h-p|T}^* & \text{pour } h > p \end{cases} \quad (4.4)$$

**Exercice 4.6.1 AR(1) moyenne non nulle** Calculer la prédiction à horizon  $h$  du modèle AR(1) :  $X_t = \mu + \phi X_{t-1} + \varepsilon_t$  ( $\mu$  considéré connu).

Solution :

$$\begin{aligned} X_{T+1|T}^* &= \mu + \phi X_T \\ X_{T+2|T}^* &= \mu + \phi X_{T+1|T}^* = \mu + \phi(\mu + \phi X_T) = \mu(1 + \phi) + \phi^2 X_T \end{aligned}$$

de façon itérative

$$X_{T+h|T}^* = \mu + \phi X_{T+h-1|T}^* = \mu(1 + \phi + \phi^2 + \dots + \phi^{h-1}) + \phi^h X_T.$$

On remarque que lorsque  $h \rightarrow \infty, X_{T+h|T}^*$  tend vers  $\frac{\mu}{1-\phi}$ , la vraie moyenne du processus  $X_t$ . L'erreur de prédiction à l'horizon  $h$  est

$$\begin{aligned} \varepsilon_h &= X_{T+h|T}^* - X_{T+h} = X_{T+h|T}^* - (\phi X_{T+h-1} + \mu + \varepsilon_{T+h}) = \dots \\ &= X_{T+h|T}^* - (\phi^h X_T + (\phi^{h-1} + \dots + \phi + 1)\mu + \varepsilon_{T+h} + \phi \varepsilon_{T+h-1} + \dots + \phi^{h-1} \varepsilon_{T+1}) \\ &= \varepsilon_{T+h} + \phi \varepsilon_{T+h-1} + \dots + \phi^{h-1} \varepsilon_{T+1} \end{aligned}$$

Son MSE est alors

$$\text{MSE } X_{T+1|T}^* = (1 + \phi^2 + \phi^4 + \dots + \phi^{2h-2})\sigma^2$$

où  $\sigma^2$  est la variance de  $\varepsilon_t$ .

**Exercice 4.6.2** On considère le processus suivant,

$$X_t = 4.0 + 0.8X_{t-1} + \varepsilon_t, \quad \varepsilon \sim N(0, 1)$$

dont les dernières observations ont été

11.287168, 19.002169, 18.664791, 17.550591, 21.724832, 21.8215, 22.59378, 23.025447, 22.023941,  
22.795715, 20.833277, 20.608492, 19.533135, 20.927484, 21.572549, 20.819312, 20.63309, 20.998086,  
22.245658, 20.709627, 21.132428, 20.935427, 20.701293, 20.16264, 18.862392, 20.378879, 21.843372,  
21.366562, 22.460745, 22.286604,

Calculer la prédiction à l'horizon 1, 2, ..., 7. Calculer intervalle de confiance de votre prédiction.

Solution.

La prévision  $X_{T+1|T}^*$  est

$$X_{T+1|T}^* = \mu + \phi X_T = 4.0 + 0.8 * 22.286604 = 21.829283$$

$$X_{T+2|T}^* = \mu + \phi X_{T+1|T}^* = 4.0 * 0.8 * 21.829283 = 21.463426$$

De meme on peut trouver

$$X_{T+3|T}^* = 21.170741, X_{T+4|T}^* = 20.936593, X_{T+5|T}^* = 20.749274$$

$$X_{T+6|T}^* = 20.599419 X_{T+7|T}^* = 20.479536$$

On peut biensur calculer  $MSE X_{T+h|T}^*$  facilement. On suppose ensuite que la variable aléatoire  $X_{T+h|T}^*$  suit une loi normale de moyenne  $X_{T+h|T}^*$ , et de variance  $MSE X_{T+h|T}^*$ , pour calculer ensuite intervalle de confiance de  $X_{T+h|T}^*$  :

$$\left[ X_{T+h|T}^* - \mu_{1-\alpha/2} \sigma \sqrt{1 + \phi^2 + \phi^4 + \dots + \phi^{2h-2}}, X_{T+h|T}^* + \mu_{1-\alpha/2} \sigma \sqrt{1 + \phi^2 + \phi^4 + \dots + \phi^{2h-2}} \right]$$

Par exemple,  $h = 2$  nous avons ici  $\sigma = 1$   $X_{T+2|T}^* = 21.463$ , avec intervalle de confiance de niveau 0.95

$$\left[ 21.829 - 1.96 \sigma \sqrt{1 + 0.8^2} 21.829 + 1.96 \sigma \sqrt{1 + 0.8^2} \right] = [21.829 - 2.51 21.829 + 2.51] = [19.319, 24.339]$$

où  $\alpha = 0.05$  et le niveau de signification est  $1 - \alpha = 0.95$ , donc la valeur critique est  $\mu_{0.975} = 1.96$  Voici le graphe explicatif

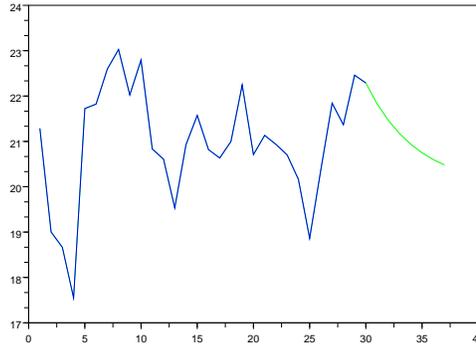


Figure 4.6:  $X_t$  et  $X_{T+h|T}^*$

**Exercice 4.6.3 suite** Soit  $Y_t = X_t - \frac{\mu}{1-\phi}$  calculer les prévisions  $Y_{T+h|T}^*$

Solution : Facile, après cette transformation,  $Y_t$  est un AR(1) de moyenne nulle :

$$\begin{aligned} Y_t &= X_t - \frac{\mu}{1-\phi} = \mu + \phi X_{t-1} + \varepsilon_t - \frac{\mu}{1-\phi} \\ &= \mu + \phi X_{t-1} - \frac{\phi\mu}{1-\phi} + \frac{\phi\mu}{1-\phi} + \varepsilon_t + \frac{\mu}{1-\phi} \\ &= \phi \left( X_{t-1} - \frac{\mu}{1-\phi} \right) + \varepsilon_t + \mu + \frac{\phi\mu}{1-\phi} - \frac{\mu}{1-\phi} = \phi \left( X_{t-1} - \frac{\mu}{1-\phi} \right) + \varepsilon_t \end{aligned}$$

D'après le résultat 4.4, nous avons  $Y_{T+h|T}^* = \phi Y_{T+h-1|T}^* = \dots = \phi^h Y_T$ . Revenons sur  $X_t$  nous avons

$$\begin{aligned} X_{T+h|T}^* &= Y_{T+h|T}^* + \frac{\mu}{1-\phi} = \phi^h Y_T + \frac{\mu}{1-\phi} \\ &= \phi^h \left( X_T - \frac{\mu}{1-\phi} \right) + \frac{\mu}{1-\phi} = \phi^h X_T + \mu \left( \frac{1-\phi^h}{1-\phi} \right) \end{aligned}$$

Comparer avec exercice précédent, on a le même résultat.

**Exercice 4.6.4 AR(2) moyenne non nulle** On considère le processus suivant :

$$X_t = 0.5 + 0.8\varepsilon_{t-1} - 0.2\varepsilon_{t-2}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

Est-ce qu'il s'agit d'un processus stationnaire? Calculer la moyenne  $E X_t$  de ce processus, et les prévisions à l'horizon 1, 2, ..., 5. Sachant les 9 premières observations sont

$$13.787168, 10.472434, 10.08357, 9.5911267, 14.340547, 14.995847, 15.265148, 15.163372, 13.681251,$$

Solution. Le calcul est presque identique que le cas, AR(1).

## 4.6.2 Prédiction de MA(q)

$X_t = \varepsilon_t + \theta_1 + \dots + \theta_1 \varepsilon_{t-q} = \Theta(B)\varepsilon_t$ . La prédiction optimale à horizon 1, faite à la date  $T$  est

$$X_{T+1|T}^* = \mathbf{E}(X_{T+1}|X_T, X_{T-1}, \dots) = \mathbf{E}L(X_{T+1}|\varepsilon_T \varepsilon_{T-1}, \dots)$$

car  $\sigma(X_T, X_{T-1}, \dots) = \sigma(\varepsilon_T, \varepsilon_{T-1}, \dots)$ , par conséquence

$$X_{T+1|T}^* = \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T+1-q} + 0$$

de même  $\forall h \geq 1$ , nous avons le résultat suivant

$$\begin{aligned} X_{T+h|T}^* &= \theta_h \varepsilon_T + \dots + \theta_q \varepsilon_{T+h-q} \quad \text{pour } h \leq q \\ &0 \quad \text{pour } h > q \end{aligned} \tag{4.5}$$

Inconvénient de cette méthode : Il faut connaître (estimer) les résidus passés, a priori non observables.  $\mathbf{E}(X_{T+1}|X_T, X_{T-1}, \dots) = \mathbf{E}L(X_{T+1}|\varepsilon_T \varepsilon_{T-1}, \dots)$ . Revenons sur  $X_t$  nous avons

### utilisation de AR( $\infty$ )

On peut écrire  $X_t = \Theta(B)\varepsilon_t$  par  $\Theta^{-1}(B)X_t = \varepsilon_t$  ce qui nous donne

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \quad \text{donc} \quad X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h}$$

On peut donc calculer  $X_{T+h|T}^*$  de façon itérative

$$X_{T+h|T}^* = \sum_{k=1}^{h-1} a_k X_{T+h-k|T}^* + \sum_{k=h}^{\infty} a_k X_{T+h-k|T}$$

cette formule n'est pas utilisable, car les  $(X_t)$  ne sont pas observés en pratique pour  $t < 0$ . On utilise une autre écriture

$$X_{T+h} = \sum_{k=1}^{\infty} a_k X_{T+h-k} + \varepsilon_{T+h} = \sum_{k=1}^{h-1} a_k X_{T+h-k} + \sum_{k=h}^{\infty} a_k X_{T+h-k} + \varepsilon_{T+h}$$

$\sum_{k=1}^{\infty} a_k X_{t+h-k}$  est une série convergente, le reste de cette série est négligeable. On peut considérer quand  $T$  suffisamment grand que

$$X_{T+h|T}^* = \sum_{k=1}^{h-1} a_k X_{T+h-k|T}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k} + \sum_{k=T+h+1}^{\infty} a_k X_{T+h-k}$$

le troisième terme étant négligeable, on approxime  $X_{T+h|T}^*$  par  $\hat{X}_{T+h|T}^*$

$$\hat{X}_{T+h|T}^* = \sum_{k=1}^{h-1} a_k \hat{X}_{T+h-k|T}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k}$$

Cette formule devient calculable

### Formule de mise à jour des résultats

Dans les calculs en temps réel, les observations  $X_t$  arrive successivement, idée est de comparer l'estimation  $X_{T+1|T}^*$  faite en  $T$  de  $X_{T+1}$  avec la vraie valeur,  $X_{T+1} - X_{T+1|T}^* = \varepsilon_{T+1}$ . L'équation 4.5 peut être modifiée. En se plaçant à la date  $T+1$ , on peut noter que  $X_{T+h|T+1}^* = X_{(T+1)+(h-1)|T+1}^*$  donc

$$X_{T+h|T+1}^* \begin{cases} \theta_{h-1} \varepsilon_{T+1} + X_{T+h|T}^* & \text{pour } h \leq q+1 \\ 0 & \text{pour } h > q+1 \end{cases}$$

donc pour  $h \leq q+1$

$$X_{T+h|T+1}^* = X_{T+h|T}^* + \theta_{h-1}(X_{T+1} - X_{T+1|T}^*) \tag{4.6}$$

Cette formule s'appelle *formule de mise à jour*.

### Exercice 4.6.5 MA(1) moyenne non nulle On considère le processus

$$X_t = 10.0 + \varepsilon_t - 0.5\varepsilon_{t-1}, \quad \text{où } \varepsilon \sim N(0, 1)$$

Les observations ont été 11.066351, 7.3288504, 10.476839, 8.8872297, 14.374981 Calculer la prédiction à l'horizon  $h = 1, 2, \dots$  de ce processus.

Solution. On définit d'abord un processus centré  $Y_t = X_t - 10 = (1.0663511, -2.6711496, 0.4768388, -1.1127703, 4.3749811)$   
 Pour calculer la prévision d'un modèle MA(q) on cherche d'abord sa forme AR( $\infty$ ).

$$\begin{aligned} \varepsilon_t &= Y_t + 0.5\varepsilon_{t-1} = Y_t + 0.5(Y_{t-1} + 0.5\varepsilon_{t-2}) = Y_t + 0.5(Y_{t-1} + 0.5(Y_{t-2} + 0.5\varepsilon_{t-3})) = \dots \\ &= \sum_{i=0}^{\infty} 0.5^i Y_{t-i} = \sum_{i=0}^{\infty} 0.5^i (X_{t-i} - 10.0) \end{aligned}$$

On obtient donc une forme AR( $\infty$ )

$$Y_t = \varepsilon_t + \sum_{k=1}^{\infty} a_k X_{t-k} = \varepsilon_t - 0.5Y_{t-1} - 0.5^2Y_{t-2} - 0.5^3Y_{t-3} - \dots$$

La formule qui permet de calculer la prévision à l'horizon h est

$$Y_{T+h|T}^* = \sum_{k=1}^{h-1} a_k Y_{T+h-k|T}^* + \sum_{k=h}^{T+h} a_k Y_{T+h-k}$$

Ici  $a_k = -0.5^k$ , et  $T = 4$ , lorsque  $h = 1$  nous avons

$$Y_{4+1|4}^* = \sum_{k=1}^{4+1} a_k Y_{4+1-k} = -0.5Y_4 - 0.5^2Y_3 - 0.5^3Y_2 - 0.5^4Y_1 - 0.5^5Y_0 = -1.8352794$$

Lorsque  $h=2$ , nous avons

$$\begin{aligned} Y_{4+2|4}^* &= \sum_{k=1}^{h-1} a_k Y_{T+h-k|T}^* + \sum_{k=h}^{T+h} a_k Y_{T+h-k} = a_1 Y_{5|4}^* + \sum_{k=2}^6 a_k Y_{6-k} \\ &= -0.5Y_{5|4}^* - 0.5^2Y_4 - \dots - 0.5^6Y_0 = 0.0 \end{aligned}$$

En fait, il est très facile de montrer que avec cette formule nous avons toujours

$$Y_{T+h|T} = 0.0, \quad \text{si } h \geq 2$$

Pour calculer  $X_{T+h|h}^*$ , il suffit alors de utiliser  $X_{T+h|T}^* = Y_{T+h|T}^* + 10$ , ce qui donne

$$X_{4+1|4}^* = 8.1642, \quad X_{4+2|4}^* = X_{4+3|4}^* = 10.0$$

Attention, ici le vecteur d'observation est  $\mathbf{X}_4 = (X_0, X_1, \dots, X_4)$ .

### 4.6.3 Prévision de ARMA(p,q)

Le processus  $\Phi(B)X_t = \Theta(B)\varepsilon_t$  :

$$\begin{aligned} X_t &= \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \\ X_{t+h} &= \sum_{i=1}^p \phi_i X_{t+h-i} + \varepsilon_{t+h} + \sum_{j=1}^q \theta_j \varepsilon_{t+h-j} \end{aligned}$$

On sait  $X_{T+h|T}^* = \mathbf{E}L(X_{T+h}|X_T, X_{T-1}, \dots) = \mathbf{E}L(X_{T+h}|\varepsilon_T, \varepsilon_{T-1}, \dots)$ . Quand  $h > q$  la prévision est similaire que le cas AR(p) (les termes qui contiennent  $\varepsilon$  sont tous indépendants de moyenne nulle par rapport à  $(X_T, X_{T-1}, \dots)$  quand  $h > q$ ).

$$X_{T+h|T}^* = \begin{cases} \phi_1 X_{T+h-1|T}^* + \dots + \phi_{h-1} X_{T+1|T}^* + \phi_h X_T + \dots + \phi_p X_{T+h-p} & \text{pour } h \leq p \\ \phi_1 X_{T+h-1|T}^* + \dots + \phi_p X_{T+h-p|T}^* & \text{pour } h > p \end{cases}$$

mais quand  $t \leq q$ , on ne peut pas enlever  $\varepsilon$ , et elle sont non observable....

#### utilisation de AR( $\infty$ ) pour ARMA(p,q)

Tout comme le cas de MA(q). on se sert de la forme AR( $\infty$ ) pour construire un algorithme calculable. L'équation  $\Phi(B)X_t = \Theta(B)\varepsilon_t$  peut toujours être écrite sous la forme (on suppose le processus est stationnaire et inversible)  $\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t$ . On obtient donc

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \implies X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h}$$

On utilise la même technique que dans le cas MA(q), en écrivant

$$X_{T+h} = \sum_{k=1}^{\infty} a_k X_{T+h-k} + \varepsilon_{t+h} = \sum_{k=1}^{h-1} a_k X_{T+h-k} + \sum_{k=h}^{T+h} a_k X_{T+h-k} + \sum_{k=T+h+1}^{\infty} a_k X_{T+h-k} + \varepsilon_{t+h}$$

le terme  $\sum_{k=T+h+1}^{\infty} a_k X_{T+h-k}$  est négligeable quand  $T$  est grande. D'où la formule itérative :

$$\hat{X}_{T+h|T}^* = \sum_{k=1}^{h-1} a_k \hat{X}_{T+h-k|T}^* + \sum_{k=h}^{T+h} a_k X_{T+h-k}$$

Ici,  $X_{T+h|T}^*$  est remplacée par  $\hat{X}_{T+h|T}^*$ , car c'est une approximation.