

# Séries temporelles : modèles et statistiques

Cet enseignement a fait l'objet d'un cours - td d'environ 40 heures au Master-2 "*Statistiques appliquées aux sciences du vivant*" de l'Université de Cotonou en novembre 2008. L'objectif était de présenter les principaux modèles de *séries temporelles* (ST), ou encore séries chronologiques (SC), leurs statistiques et les méthodes de prévision. L'étude d'une ST a un double objectif : modéliser pour expliquer un phénomène observé ; prédire la série dans le futur. Le plan de ce cours est le suivant :

1. *Modèles stationnaires au second ordre.*
2. *Modèles non-stationnaires.*
3. *Prédiction.*
4. *Modèles ARMA.*
5. *Un peu de théorie spectrale.*
6. *Estimation d'un ARMA.*
7. *Régression à résidus corrélés, ARIMA et SARIMA.*
8. *Prévision sans modèle.*
9. *SC bivariée ; modèle ARCH.*
10. *Annexes et références.*

A l'inverse d'un échantillon, les observations d'une SC sont dépendantes, cette dépendance temporelle étant modélisée de façon causale avec comme modèles centraux les *ARMA*. La littérature sur ce sujet est abondante. Deux livres "historiques" donnent la présentation moderne des SC : Anderson (1971), *The statistical analysis of time series* (J. Wiley) et Box et Jenkins (1976) : *Time Series Analysis : Forecasting and Control* (Holden-Day). Le livre de Brockwell et Davis (1991), *Time Series : Theory and Methods* (Springer) est une référence théorique incontournable mais difficile. Ces mêmes auteurs ont écrit un deuxième livre, *Introduction to Time Series and Forecasting* (2002), qui est une approche plus simple, claire, bien illustrée et à laquelle est jointe un tutoriel, *ITSM*, très pédagogique, ouvrage duquel nous avons retiré quelques idées ou exercices. Il existe également des polycopiés disponibles sous le label "Séries Temporelles"

## Le logiciel R

R est un outil indispensable à une bonne compréhension du cours et à sa mise en application. Nous conseillons vivement au lecteur d'avancer dans la lecture de ce cours en ouvrant une deuxième fenêtre "R" et en faisant les va-et-vient indispensables entre les notions et résultats présentés et les fonctions de R qui leurs sont dédiés. Quelques-unes de ces fonctions sont données dans le cours du texte. Pour l'installation de R et une première description, c.f. l'annexe du §10.5.

**Un exemple introductif : la série des niveaux du lac Huron.** Parcourant le package `DataSets` de R, en extraire la série temporelle `LakeHuron` : cette série annuelle donne les niveaux du lac Huron de 1875 à 1972 ( $n = 98$  données). Déclarant cette suite comme une série chronologique (`x=ts(LakeHuron)`) et utilisant la fonction `Plot`, représenter graphiquement cette série  $X_1, X_2, \dots, X_{98}$  (pour d'autres jeux de données, cf. § 10.5.1).

Qu'observe-t-on ? Une décroissance moyenne de ce niveau, approximativement affine, qu'on peut modéliser par :

$$X_t = a + bt + e_t, t = 1, 98.$$

Il est facile d'estimer cette tendance par moindres carrés ordinaires (fonction `lm` de R). On peut alors analyser les résidus estimés  $\hat{e}_t = X_t - (\hat{a} + \hat{b}t)$ ,  $t = 1, 98$ .

On commencera par représenter graphiquement cette série. En moyenne, les  $\hat{e}_t$  successifs gardent un même signe avec seulement 10 changements de signes sur 98 instants : ceci indique une corrélation positive entre les  $\hat{e}$  successifs. Ainsi, se pose une question centrale dans l'étude des SC :  $\hat{e}$  est-il un bruit blanc (BB), c-à-d une suite de variables centrées, de mêmes variances et non-corrélés ? Pour répondre à cette question, on va utiliser le programme `acf` (du package `stats`) qui fournit le graphique des premières auto-covariances empiriques de  $X$ , avec les limites de l'intervalle de confiance au niveau 95% sur cette acf empirique si  $\hat{e}$  était un BB : clairement,  $\hat{e}$  n'est pas un BB (on donnera des procédures de test statistique pour décider de cela).

Mais pourquoi est-il important de savoir si  $\hat{e}$  est un BB ou non ? Parce qu'il est intéressant de savoir s'il y a une tendance pour les résidus à garder localement un signe positif (période de recharge du lac) ou négative (période de baisse de niveau), cela une fois ôtée la tendance affine  $\hat{a} + \hat{b}t$ . Une autre raison est la suivante : si on veut prévoir la hauteur du lac en 1973 ou plus tard, il est important de savoir si  $\hat{e}$  est corrélé ou non : si  $\hat{e}$  est un BB, la bonne prévision naturelle en 1973 est  $\hat{X}_{1973} = \hat{a} + \hat{b} \times 99$  et dans ce cas, on ne peut pas faire mieux. Mais si  $\hat{e}$  n'est pas un BB, on peut faire mieux en utilisant la structure de corrélation de  $\hat{e}$ . Par exemple, s'il est raisonnable de modéliser  $\hat{e}$  par un  $AR(1)$ ,

$$\hat{e}_t = \rho \hat{e}_{t-1} + \eta_t \quad (\eta \text{ un BB}),$$

où on estime  $\rho$  par  $\hat{\rho}$ , alors la prédiction  $\tilde{X}_{1973} = \hat{a} + \hat{b} \times 98 + \hat{\rho} \times \hat{e}_{1972}$ , où  $\hat{e}_{1972} = X_{1972} - (\hat{a} + \hat{b} \times 1972)$ , est meilleure que la précédente.

Cet exemple simple soulève quelques-unes des questions qui seront étudiées ici :  $X$  est-elle de moyenne constante ? Si non, quel modèle proposer pour la moyenne de  $X$  ? Comment rendre  $X$  stationnaire en moyenne ? Le résidu de la régression envisagée est-il un BB ? Sinon, quel modèle stationnaire proposer pour ce résidu et comment l'estimer ? Comment utiliser cette structure au premier ordre (moyenne) et second ordre (covariance des résidus) pour prévoir au mieux la série dans le futur ? Comment valider le modèle retenu (premier et second ordre) ?

# 1 Modèle stationnaire au second ordre

## 1.1 Définitions

La notion fondamentale qui structure la modélisation d'un processus temporel est-celle de stationnarité. Commençons par donner quelques définitions

1. Une SC (un processus)  $X = (X_t, t \in \mathbb{Z})$  est dite *du second ordre* (on note  $X \in L^2$ ) si pour tout  $t$ ,  $E(X_t)^2 < \infty$ . Dans ce cas,  $E(|X_t|) < \infty$ , l'espérance existe et il est équivalent de dire que  $Var(X_t) < \infty$  pour tout  $t$ . On note  $m(t) = E(X_t)$  la fonction *moyenne de X* et  $C(s, t) = Cov(X_s, X_t)$  sa fonction de *covariance*.
2.  $X$  est *gaussienne* si toute combinaison linéaire finie des  $(X_t)$  est gaussienne : pour toute suite d'instant  $t_1, t_2, \dots, t_K$  et de poids réels  $a_1, a_2, \dots, a_K$ ,  $Y = \sum_{j=1}^K a_j X_{t_j}$  est gaussienne. Une série gaussienne est toujours du second ordre mais l'inverse n'est pas vrai. La loi d'un processus gaussien est entièrement caractérisée par ses fonctions moyenne  $m(\cdot)$  et covariance  $C(\cdot, \cdot)$ .
3. Un processus  $X$  est dit *stationnaire au second ordre* si pour tout  $t, s \in \mathbb{Z}$ , la moyenne est constante et la covariance  $C(s, t)$  ne dépend que de  $(t - s)$  :

$$\forall s, t : m(t) = E(X_t) = m \text{ et } Cov(X_s, X_t) = C(s, t) = \gamma(s - t).$$

$h \mapsto \gamma(h)$  est la *fonction d'auto-covariance* (acvf) de  $X$  et  $h \mapsto \rho(h) = \frac{\gamma(h)}{\gamma(0)}$  sa fonction d'*auto-corrélation* (acf).

4. Un processus  $X$  est *strictement stationnaire* si pour tout  $k > 0$ , toute suite  $t_1, t_2, \dots, t_k$  et tout  $h \in \mathbb{Z}$ , les lois de  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  et de  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$  sont identiques : la translation  $\mathcal{T}(h) : t \rightarrow t + h$  laisse invariante la loi de  $X$ . Un processus strictement stationnaire et du second ordre est stationnaire au second ordre mais l'inverse n'est pas vrai en général, sauf pour les processus gaussiens.

Dans ce cours, nous utiliserons principalement la notion de stationnarité au second ordre : sans mention explicite, "stationnaire" signifiera "stationnaire au second ordre".

5. Un processus  $e = (e_t, t \in \mathbb{Z})$  est un *bruit i.i.d.* si  $e$  est une suite i.i.d. centrée. Si la loi commune des  $e_t$  est  $\mathcal{L}$ , on notera  $e \sim i.i.d.(\mathcal{L})$  et  $e \sim i.i.d.(0, \sigma^2)$  si  $\mathcal{L}$  est centrée et de variance finie  $\sigma^2$ ,  $e \sim i.i.d.(\mathcal{N}(0, \sigma^2))$  si la loi commune est gaussienne centrée de variance  $\sigma^2$ .
6. Un processus  $e = (e_t, t \in \mathbb{Z})$  est un *bruit blanc faible* (noté BB) si  $e$  est stationnaire au second ordre, centré, avec  $Cov(e_t, e_s) = 0$  si  $s \neq t$ . On notera  $e \sim BB(0, \sigma^2)$  si  $Var(e_t) = E(e_t^2) = \sigma^2$ .

## 1.2 Exemples

**Marche aléatoire :** pour  $e \sim i.i.d.(0, \sigma^2)$ , on définit  $X_0 = e_0$  et pour  $t \geq 0$ ,  $X_{t+1} = X_t + e_{t+1}$ .  $X$  n'est pas stationnaire car  $Var(X_t) = t\sigma^2$  n'est pas constante. Plus généralement,  $Cov(X_t, X_s) = s\sigma^2$  si  $s \leq t$ .

**Processus MA(1) :** soit  $e \sim BB(0, \sigma^2)$ ,  $\theta \in \mathbb{R}$  et la série,

$$X_t = e_t + \theta e_{t-1}, t \geq 1 \quad (1)$$

$X$  est stationnaire centré de covariance,

$$\gamma(0) = \sigma^2(1 + \theta^2), \gamma(1) = \gamma(-1) = \theta\sigma^2 \text{ et } \gamma(h) = 0 \text{ si } |h| \geq 2.$$

L'acvf (l'acf) d'un MA(1) est de portée 1.

Plus généralement (cf. §4.2), pour  $q$  un entier  $> 0$ , un  $MA(q)$  est défini par,

$$X_t = e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}, t \geq 1.$$

Il est facile de voir que l'acvf est nulle si  $|h| \geq q + 1$ . Cette propriété est caractéristique d'un  $MA(q)$ .

**Processus AR(1) :** Soit  $|\phi| < 1$ . Supposons que la série,

$$X_t = \phi X_{t-1} + e_t \text{ où } e \sim BB(0, \sigma^2),$$

est stationnaire et que  $e_t$  est non-corrélé à  $X_s$  dès que  $s < t$ . D'une part,  $E(X_t) = \phi E(X_{t-1})$  et donc  $m = 0$  ( $X$  est centré); d'autre part, pour  $h \geq 1$  :

$$\begin{aligned} \gamma(h) &= Cov(X_t, X_{t-h}) = Cov(\phi X_{t-1} + e_t, X_{t-h}) \\ &= \phi \gamma(h-1) = \gamma(0) \phi^{|h|}. \end{aligned}$$

L'acvf d'un AR(1) décroît exponentiellement.

**Processus linéaire :** ces processus sont à la base de la construction des modèles ARMA (cf. §4.2). Soit  $e \sim BB(0, \sigma^2)$ ,  $(\theta_j, j \in \mathbb{Z})$  une suite réelle de  $l^2(\mathbb{Z})$  ( $\sum_{\mathbb{Z}} \theta_j^2 < \infty$ ). Alors le processus linéaire  $X$  associé aux  $(\theta_j)$  et à  $e$  comme défini ci-dessous est stationnaire :

$$X_t = \sum_{\mathbb{Z}} \theta_j e_{t-j} \text{ est d'acvf : } \gamma(h) = \sigma^2 \sum_{\mathbb{Z}} \theta_j \theta_{j+h}.$$

**Série périodique :** soit  $0 < \omega < \pi$  une fréquence donnée et  $X_t = A \cos(\omega t) + B \sin(\omega t)$ ,  $A$  et  $B$  étant décorréelées, centrées, de variances finies.  $(X_t)$  est centrée et  $Var(X_t)$  est constante si  $Var(A) = Var(B) = \sigma^2$ .  $X$  est alors stationnaire d'acvf :

$$\gamma(h) = \sigma^2 \cos(\omega t).$$

Plus généralement, soient  $0 < \omega_1 < \omega_2 < \dots < \omega_r < \pi$   $r$ -fréquences fixées,  $\{A_i, B_i, i = 1, r\}$  2*r* v.a. décorréelées, centrées telles que  $Var(A_i) = Var(B_i) = \sigma_i^2 < \infty, i = 1, r$ . Alors le mélange de  $r$ -processus périodiques associés aux  $r$  fréquences est stationnaire :

$$X_t = \sum_{i=1,r} \{A_i \cos(\omega_i t) + B_i \sin(\omega_i t)\} \text{ d'acvf } \gamma(h) = \sum_{i=1,r} \sigma_i^2 \cos(\omega_i t). \quad (2)$$

Un mélange "continu" de telles composantes infinitésimales est de fait la forme générale d'une SC du second ordre (cf. (30)).

### 1.3 Moyenne, acvf et acf empiriques

Supposons que  $X(n) = (x_1, x_2, \dots, x_n)$  est l'observation de  $X$ , une série stationnaire. La moyenne empirique est :

$$\bar{x} = \hat{m} = \frac{1}{n} \sum_1^n x_i.$$

L'acvf empirique à distance  $h$  est :

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (x_{i+|h|} - \bar{x})(x_i - \bar{x}), \quad -n < h < n.$$

Diviser par  $n$ , et non par  $(n-|h|)$ , le nombre de termes de la somme, garantit que  $\hat{\Gamma}_n = [\hat{\gamma}(i-j)]_{i,j=1}^n$  est s.d.p., propriété importante pour l'estimation et la prédiction.

Enfin, l'acf empirique est  $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$  pour  $-n < h < n$ .

#### 1.3.1 Exercices

1. Simuler  $n = 200$  valeurs *i.i.d.*  $\mathcal{N}(0, 1)$  et calculer moyenne et auto-corrélations empiriques (fonctions `rnorm` et `acf` de `R`).
2. Soit  $\{Z_t\} \sim i.i.d. \mathcal{N}(0, \sigma^2)$ ,  $a, b$  et  $c$  trois constantes. Parmi lesquels des processus suivants, lesquels sont stationnaires et préciser la moyenne et l'acvf de ceux qui le sont :
  - (a)  $X_t = a + bZ_t + cZ_{t-2}$  ;
  - (b)  $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$  ;
  - (c)  $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$  ;
  - (d)  $X_t = Z_{2t}$  ;
  - (e)  $X_t = Z_t Z_{t-1}$  ;
  - (f)  $X_t = f(Z_t, Z_{t-1})$  où  $f$  est continue et bornée ;
  - (g)  $X_t = Z_t Z_{2t}$  ;
  - (h)  $X_t = Z_t^2 Z_{t+1}$ .
3. *BB i.i.d. ou Bruit Blanc faible* : soit  $\{Z_t\} \sim i.i.d. \mathcal{N}(0, 1)$  et  $X$ ,

$$X_t = \begin{cases} Z_t & \text{si } t \text{ est pair et} \\ \frac{1}{\sqrt{2}}(Z_{t-1}^2 - 1) & \text{sinon.} \end{cases} \quad (3)$$

- (a) - Vérifier que  $X \sim BB(0, 1)$  mais que  $X$  n'est pas une suite *i.i.d.*. Que vaut  $E(X_{n+1} | X_n)$  suivant la parité de  $n$  ?
- (b) - Simuler une série  $X$  de longueur 200. Graphique et acvf empirique  $X$ . Cela confirme-t-il la non-corrélation de  $X$  ?
- (c) - Soit  $Y_t = X_t^2, t = 1, 200$  ; acvf empirique de  $Y$  ;  $Y$  est-elle un *BB* ? En déduire que la suite  $X$  n'est pas *i.i.d.*. Et si  $Y_t = |X_t|$  ?

4. On note  $\hat{\rho}(h)$  l'acf des observations  $x_1, x_2, \dots, x_n$ .
- a - Montrer que  $\hat{\rho}(h) \rightarrow 1$  si  $n \rightarrow \infty$  pour  $x_t = a + bt, b \neq 0$ .
- b - Montrer que  $\hat{\rho}(h) \rightarrow \cos(ah)$  si  $n \rightarrow \infty$  pour  $x_t = \cos(at)$ .

## 1.4 Propriétés

**Auto-covariance :**

- $\gamma(0) \geq 0, |\gamma(h)| \leq \gamma(0)$  et  $\gamma(h) = \gamma(-h)$  pour tout  $h$ .
- Pour toute suite  $a_1, a_2, \dots, a_K$  réelle :

$$\text{Var}\left(\sum_{t=1}^K a_t X_t\right) = \sum_{i,j=1}^K a_i a_j \gamma(i-j) \geq 0. \quad (4)$$

Cette propriété traduit que l'acvf  $\gamma(\cdot)$  est *semi-définie positive* (s.d.p.). Elle est *définie positive* (d.p.) si on a la stricte positivité dès que  $a \neq 0$ . Cette définition est analogue à la s.d.p. des matrices  $[\gamma(i-j)]_{i,j=1}^K$  pour tout  $K \geq 1$ . On a la propriété réciproque et caractéristique suivante :

- Une fonction réelle sur  $\mathbb{N}$  est une acvf si et seulement si (ssi) elle est paire et s.d.p..
- S'assurer qu'une fonction  $\kappa(\cdot)$  est une acvf? Le plus simple est de "proposer" une série dont l'acvf est  $\gamma(\cdot)$  (cf. aussi la caractérisation d'une acvf via la densité spectrale de  $X$ , §5.1).

Par exemple, montrons que la fonction  $\kappa(\cdot)$  définie ci-dessous est une covariance ssi  $|\rho| \leq \frac{1}{2}$  :

$$\kappa(0) = 1, \kappa(h) = \rho \text{ si } |h| = 1 \text{ et } \kappa(h) = 0 \text{ sinon.} \quad (5)$$

- *La condition est suffisante* : le MA(1) (1) admet pour acvf  $\gamma(h) = \frac{\theta}{1+\theta^2} \rho^{|h|}$ , équation en  $\theta$  qui admet bien une racine réelle si  $|\rho| \leq \frac{1}{2}$ ; pour avoir  $\gamma(0) = 1$ , on imposera  $\sigma^2(1+\theta^2) = 1$ .
- *La condition est nécessaire* : ou  $\kappa(\cdot)$  n'est pas une acvf sinon. Pour cela, on propose  $a$  t.q. (4) n'est pas vérifiée : si  ${}^t a = (1, -1, 1, -1, \dots) \in \mathbb{R}^n$ , alors  ${}^t a K_n a = n - 2\rho(n-1)$  si  $K = [\kappa(i-j)]_{i,j=1}^n$  et donc que  ${}^t a K_n a < 0$  si  $|\rho| > \frac{1}{2}$  et  $n$  grand :  $\kappa(\cdot)$  n'est pas une acvf.

### 1.4.1 Exercices

- $X$  et  $Y$  sont deux SC stationnaires au second ordre et indépendantes. Démontrer que  $X+Y$  est stationnaire.
- En identifiant une série associée appropriée, vérifier que  $C_1$  et  $C_2$  sont des covariances :

$$\begin{aligned} C_1(h) &= 4 \times (0.5)^{|h|} + \cos(ah); \\ C_2(h) &= 1 \text{ si } h = 0, C_2(h) = 0 \text{ si } |h| = 1 \text{ ou } |h| \geq 3 \text{ et } C_2(h) = 0.4 \text{ si } |h| = 2. \end{aligned}$$

## 1.5 Propriétés des estimations empiriques

Soit  $X(n) = (X_1, X_2, \dots, X_n)$  l'observation d'une série  $X$  stationnaire au second ordre.

### 1.5.1 Moyenne empirique.

- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  estime sans biais la moyenne  $m = E(X_1)$ .
- Si  $\gamma(h) \xrightarrow{h \rightarrow \infty} 0$ , alors  $\text{Var}(\bar{X}_n - \mu) = E(\bar{X}_n - \mu)^2 \xrightarrow{n \rightarrow \infty} 0$  (on note  $\bar{X}_n \xrightarrow{L^2} \mu$  cette convergence en moyenne quadratique).
- Si de plus  $v = \sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$ , alors :  $\text{Var}(\bar{X}_n) \sim \frac{v}{n}$ .

Anticipant la définition des modèles linéaires (cf. §4.2), on a la convergence en loi suivante :

- Si de plus  $X$  est un processus linéaire (i.e. un ARMA) associé à un bruit  $i.i.d.(0, \sigma^2)$  :

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{v}{n}\right).$$

L'intervalle de confiance (IC) approché de  $m$  à 95% est  $[\bar{X}_n - 1.96\sqrt{\frac{v}{n}}, \bar{X}_n + 1.96\sqrt{\frac{v}{n}}]$ .

## 1.5.2 Lois des covariances et corrélations empiriques.

Soient  $\hat{\gamma}(h)$  et  $\hat{\rho}(h)$  pour  $-n < h < n$  les *acvf* et *acf empiriques* de  $X$  calculées à partir de l'observation  $X(n)$ . On a :

- Ces estimations sont approximativement sans biais si  $n$  est “grand”.
- Une règle est d'avoir au moins  $n > 40$  observations et de se limiter aux  $h < \frac{n}{4}$ .
- Les matrices de covariances (de corrélations) associées aux  $\hat{\gamma}(\cdot)$  sont s.d.p.. Elles sont d.p. dès que la loi de  $X_1$  est à densité.
- Ces estimations sont asymptotiquement normales si  $X_t = \sum_{\mathbb{Z}} \psi_i \varepsilon_{t-i}$  pour  $\varepsilon \sim BB(0, \sigma^2)$  à condition que  $\sum_{\mathbb{Z}} |\psi_i| < \infty$  et  $\sum_{\mathbb{Z}} |i| \psi_i^2 < \infty$  (BD).
- Ces deux dernières conditions sont satisfaites si  $X$  est un ARMA.
- **Formule de Bartlett** : elle précise la matrice de covariance d'un vecteur de corrélations empiriques; soit  $\boldsymbol{\rho} = (\rho(1), \rho(2), \dots, \rho(k))$  et  $\hat{\boldsymbol{\rho}}$  son estimation empirique. Alors,  $\hat{\boldsymbol{\rho}} \sim N_k(\boldsymbol{\rho}, \frac{W}{n})$  où  $W$  est donnée par la *formule de Bartlett* :

$$w_{ij} = \sum_{l \in \mathbb{Z}} \{\rho(l+i) + \rho(l-i) - 2\rho(l)\rho(i)\} \{\rho(l+j) + \rho(l-j) - 2\rho(l)\rho(j)\}. \quad (6)$$

La formule de Bartlett est la clé de questions de base concernant un processus stationnaire. Un cas particulier important est à la base des tests de :

“( $H_0$ ) : les  $(X_i)$  sont *i.i.d.* de variance finie”,

Sous ( $H_0$ ),  $w_{ii} = 1$  et  $w_{ij} = 0$  pour tout  $i \neq j$ , et donc pour si  $n$  grand :

$$\boxed{\text{Sous } (H_0) : \hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(k) \text{ sont } i.i.d. \mathcal{N}(0, \frac{1}{n}).}$$

## 1.6 Tests de bruit blanc

### 1.6.1 Tests paramétriques

On veut tester l'hypothèse,

$$(H_0) : X \sim i.i.d.(m, \sigma^2),$$

de loi commune centrée ( $m = 0$ ) ou non. Il y a 3 versions test du *Chi 2* résultant de (6).

*Test du portemanteau* :

$$\text{Sous } (H_0) : Q = n \sum_{k=1}^K \hat{\rho}^2(k) \sim \chi_K^2 \text{ si } n \text{ est grand.}$$

$\mathcal{R}_\alpha = \{n \sum_{k=1}^K \hat{\rho}^2(k) \geq q(\alpha; K)\}$  est une région de rejet approximative de ( $H_0$ ) au niveau  $\alpha$  pour le  $\alpha$ -quantile  $P(\chi_K^2 \geq q(\alpha; K)) = \alpha$ .

*Test de Ljung-Box* : est issu d'un raffinement de l'approximation par un  $\chi^2$  :

$$\text{Sous } (H_0) : Q_{LB} = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}^2(k)}{n-k} \sim \chi_K^2.$$

*Test de McLeod-Li* : sous  $(H_0)$ , la suite  $Y_t = X_t^2$  est *i.i.d.* de variance finie si  $E(X_1^4) < \infty$  ; notant  $\rho_{XX}$  la corrélation des  $(X_i^2, i = 1, n)$ , on a :

$$\text{Si } E(X_1^4) < \infty \text{ et sous } (H_0) : Q_{ML} = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_{XX}^2(k)}{n-k} \sim \chi_K^2.$$

Ce test permet de détecter qu'une suite est non *i.i.d.* même si son acvf empirique est proche de celle d'un BB. D'autres transformations  $Y_t = f(X_t)$  peuvent être utilisées, par exemple  $Y_t = |X_t|$ . Vous pouvez à ce sujet reprendre l'exemple de la série  $X$  définie à l'exercice 1.3.1-2).

## 1.6.2 Tests non-paramétriques

*Le test du point tournant.*

$2 \leq i \leq n-1$  est un point tournant d'une suite numérique  $x = (x_1, x_2, \dots, x_n)$  si soit  $x_{i-1} < x_i$  et  $x_i > x_{i+1}$  (pic local de la suite  $x$  en  $i$ ), soit  $x_{i-1} > x_i$  et  $x_i < x_{i+1}$  (vallée locale de la suite  $x$  en  $i$ ). Notons :

$$(\tilde{H}_0) : \text{les } (X_i) \text{ sont } i.i.d. \text{ de loi commune à densité,}$$

et  $T$  le nombre de points tournants de  $x$ . Il est intuitif que  $T$  est faible si les  $X_i$  successifs sont corrélés positivement et que  $T$  est grand si les  $X_i$  successifs sont corrélés négativement. On propose donc comme région de rejet de  $(\tilde{H}_0)$  :

$$\mathcal{R}_\alpha = \{|T - E_{\tilde{H}_0}(T)| \geq s(\alpha)\},$$

les valeurs de  $T$  autour de son espérance sous  $\tilde{H}_0$  correspondant à l'acceptation de  $(\tilde{H}_0)$ . Il faut évaluer la loi de  $T$  sous  $(\tilde{H}_0)$ . Sous l'hypothèse que la loi marginale commune de  $X_t$  est à densité, on a (cf. exercice 1.6.4-1) :

$$E(T) = \frac{2}{3}(n-2) \text{ et } Var(T) = \frac{16n-29}{90}.$$

D'autre part, pour  $n$  grand, on peut montrer que  $T$  est approximativement normal. La région de rejet approximative au niveau  $\alpha$  pour  $n$  grand est,  $t(\beta)$  étant le  $\beta$ -quantile de la normale réduite :

$$\mathcal{R}_\alpha = \left\{ \left| T - \frac{2}{3}(n-2) \right| \geq t\left(\frac{\alpha}{2}\right) \sqrt{\frac{16n-29}{90}} \right\}.$$

*Le test de signe sur la différence.*

On compte  $S$  le nombre de fois que  $x_i > x_{i-1}$  pour  $i = 2, n$  : sous  $(\tilde{H}_0)$ ,  $E(S) = \frac{1}{2}(n-1)$  et on peut aussi montrer que  $Var(S) = \frac{n+1}{12}$ ,  $S$  étant approximativement normale pour  $n$  grand. Les petites ou les grandes valeurs de  $S$  mettant en doute  $(\tilde{H}_0)$ , on propose comme région de rejet,

$$\mathcal{R}_\alpha = \left\{ \left| S - \frac{n-1}{2} \right| \geq t\left(\frac{\alpha}{2}\right) \sqrt{\frac{n+1}{12}} \right\}.$$

### Puissance des tests.

Comme on devrait le faire pour chaque test en cas de maintien de l'hypothèse nulle, il faut évaluer la puissance du test en spécifiant la famille d'alternatives envisagées. A ce sujet, le test de signe sur la différence présente un risque car il conduit à accepter  $(\tilde{H}_0)$  si  $(x_i)$  est proche de la périodicité ( $S$  est alors proche de  $\frac{n-1}{2}$ ) : le test de signe est de faible puissance pour la famille d'alternatives où  $x$  est proche de la périodicité.

### 1.6.3 Test du caractère gaussien

On suppose observé un échantillon *i.i.d.*  $\{e_i, i = 1, n\}$  et on veut tester  $(H_0) : e_1 \sim \mathcal{N}(m, \sigma^2)$ .

**Approche graphique : droite de Henry ou qq-plot** Classons par ordre croissant les  $e : e_{(1)} < e_{(2)} < \dots < e_{(n)}$  et notons  $u_{(1)} < u_{(2)} < \dots < u_{(n)}$  un échantillon gaussien réduit réordonné,  $m_j = E(u_{(j)})$  l'espérance de ces statistiques d'ordre. Le qq-plot des  $\{e_i\}$  est le graphique croisant les *quantiles théoriques* et les *quantiles empiriques*  $qq(e) = \{(m_j, e_{(j)}), j = 1, n\}$  : sous  $(H_0)$ , et pour  $n$  grand, les points de  $qq(e)$  sont approximativement alignés sur une droite de pente  $\sigma$  et d'ordonnée à l'origine  $m$ . Une forte déviation à cette propriété fait douter de  $(H_0)$ , c-à-d du caractère gaussien des observations. La fonction `qqplot` du package `Stats` de `R` permet de juger de cet alignement. Une règle statistique de rejet est construite sur la base de  $r^2$ , le carré du coefficient de corrélation empirique du graphique  $qq(e)$ ,

$$r^2 = \frac{\{\sum (e_i - \bar{e})m_i\}^2}{\{\sum (e_i - \bar{e})^2\}\{\sum m_i^2\}}$$

On rejettera  $(H_0)$  au niveau  $\alpha$  si  $\mathcal{R}_\alpha = \{r^2 < s(\alpha)\}$ . Pour  $n < 100$ , il existe des tables donnant  $s(\alpha)$ ; pour  $n = 200$ ,  $s(5\%) = .987$  et  $s(10\%) = .989$ .

**Approche paramétrique : test de Jarque-Berra (JB)** La statistique de JB est construite sur la constatation suivante : une loi gaussienne est de *coefficient d'asymétrie*  $\kappa_3 = E(X - \mu)^3 / \sigma^3$  nul (*skewness*,  $\kappa_3 = 0$  si  $X$  symétrique) et de *coefficient d'applatissage*  $\kappa_4 = E(X - \mu)^4 / \sigma^4 = 3$  (*kurtosis*).

Jarque et Bera proposent un test de l'hypothèse plus faible :

$$(\tilde{H}_0) : \kappa_3 = 0 \text{ et } \kappa_4 = 3.$$

Soient  $S$  et  $K$  les estimateurs empiriques de  $\kappa_3$  et  $\kappa_4$  calculés à partir de  $(e_i)$ ; alors la statistique :

$$JB = \frac{n-p}{6} \left\{ S^2 + \frac{1}{4}(K-3)^2 \right\}$$

prend une valeur faible sous l'hypothèse de normalité. Jarque et Bera ont montré que la loi asymptotique de  $JB$  est, sous  $(\tilde{H}_0)$ , un  $\chi^2$  à 2 ddl. On rejette l'hypothèse gaussienne si on tombe dans la région de rejet  $\mathcal{R}_\alpha = \{JB \geq q(2; \alpha)\}$ .

### 1.6.4 Exercices et fonctions R utiles

1. *Test du point tournant.* Soient  $X_1, X_2, X_3$  trois v.a.r. *i.i.d.* de densité  $f$ .
  - a - Démontrer que  $p = P(X_1 < X_2 \text{ et } X_3 < X_2) = \int_{\mathbb{R}} P(Y_1 < Y_2 \text{ et } Y_3 < Y_2 \mid Y_2 = y)f(y)dy = \frac{1}{3}$ . Que vaut  $q = P(Y_1 > Y_2 \text{ et } Y_3 > Y_2)$  ?
  - b - En déduire que l'espérance du nombre de points tournants d'une série  $\{X_i, i = 1, n\}$  de v.a.r. *i.i.d.* à densité est  $\frac{2}{3}(n-2)$ .
2. *Loi asymptotique pour la covariance empirique d'un bruit i.i.d.  $(0, \sigma^2)$ .*  
Soient  $(e_t, t = 1, 2, \dots)$  un bruit *i.i.d.*  $(0, \sigma^2)$  et sa covariance empirique à distance 1 calculée pour un échantillon  $e_1, e_2, \dots, e_n$  :

$$\hat{\gamma}_n(1) = \frac{1}{n} \sum_{i=1, n-1} U_i \text{ où } U_i = e_i e_{i+1}, i \geq 1.$$

- a - Vérifier que  $(U_i, i \geq 1)$  est une suite strictement stationnaire du second ordre. Vérifier de plus que  $U_i$  et  $U_j$  sont indépendantes si  $|i - j| > 1$ . Quelle est l'acvf de  $U$  ?

b - Utilisant le TCL pour les suites  $m$ -indépendantes (cf. Annexe §10.4), démontrer que :

$$\sqrt{n}\widehat{\gamma}_n(1) \rightsquigarrow \mathcal{N}(0, \sigma^4) \text{ et } \sqrt{n}\widehat{\rho}_n(1) \rightsquigarrow \mathcal{N}(0, 1).$$

c - Généraliser ce résultat à la corrélation empirique à distance  $k$  et à la loi jointe de deux corrélations empiriques.

3. Simuler un 200-échantillon d'une loi  $\mathcal{L}$ , dresser son qq-plot et tester la gaussianité de  $\mathcal{L}$  par :
  - (i) le test sur  $r^2$  (`cor(x, y)` de `stats`);
  - (ii) la statistique  $JB$ .

On envisagera les lois  $\mathcal{L}$  suivantes : une gaussienne (`rnorm(n, mean, sd)` est un générateur de  $n$  v.a. *i.i.d.*  $\mathcal{N}(m, \sigma^2)$ ); une Student (`rt(n, df, ncp)`, `ncp` = paramètre de non centralité); une loi exponentielle (`rexp(n, rate)`); une loi uniforme (`runif(n, min, max)`); une loi  $> 0$  log-normale (`rlnorm(n, meanlog, sdlog)`).

*Fonctions R utiles :*

Package `stats` : `rxxx(n, paramètres)` (simulation d'un  $n$ -échantillon de la loi `xxx`), `acf`, `qqplot`, `Box.test` (tests de Box-Pierce et Ljung-Box)

Package `tseries` : `jarque.bera.test`, `run.test` (test des runs du caractère *i.i.d.* d'une série binaire)

## 2 Modèles non-stationnaires

Ce sujet sera repris au chapitre 7 traitant des SARIMA et des régressions à résidu processus. Présentons quelques situations de non-stationnarité.

### 2.1 Non-stationnarité en moyenne

C'est le cas où  $t \mapsto m(t) = E(X_t)$  n'est pas constante.

#### 2.1.1 Modèle de régression

On dispose d'un modèle paramétrique  $m(t) = g(t, \beta)$  où  $\beta \in \mathbb{R}^p$  et  $g$  est une fonction connue. Le modèle est linéaire en  $\beta$ , c-à-d  $g(t, \beta) = {}^t g(t)\beta$ , ou non. Dans un modèle linéaire,

$$X_t = {}^t g(t)\beta + e(t), t = 1, 2, \dots \quad (7)$$

où  $e$  est un résidu centré;  $e$  pourra être stationnaire de modèle paramétrique  $e(\varphi)$ ,  $\varphi \in \mathbb{R}^l$ ; dans ce cas  $X$  est paramétré au premier ordre ( $\beta$ ) et au second ordre ( $\varphi$ ). Si  $e$  n'est pas stationnaire, on devra opérer des transformations pour le stationnariser (transformation de Box-Cox, cf. §2.2; différenciation, cf. §2.1.4).

Donnons quelques exemples classiques de modèles (7) :

- $g(t)$  est un vecteur de variables exogènes observables.
- $g(t)$  est le vecteur des monomes en  $t$  de degré  $\leq r$ .
- $g(t)$  est associé à une composante saisonnière périodique.
- ou  $g(t)$  est une combinaison de ces modèles.

Une modélisation classique *tendance + saisonnalité + résidu* est :

$$X_t = m(t) + s_t + e_t, t \geq 1, \quad (8)$$

$m(t)$  est une *tendance* et  $(s_t)$  est une *saisonnalité de période  $d$* , c-à-d que  $\forall t \geq 0, s_{t+d} \equiv s_t$ . Par exemple,  $m(t) = a + bt$  est une tendance affine et  $s$  est une *saisonnalité centrée* et (8) est de

dimension  $d + 1$  ; le centrage  $s_1 + s_2 + \dots + s_d = 0$  est indispensable afin de rendre les paramètres  $(a, s)$  identifiables.

*Exemple :* La série **AirPassengers** (cf. **datasets** de **R**) est un classique de l'étude des SC (cf. Box & Jenkins). Cette série *mensuelle* donne le nombre total de passagers avion de 1949 à 1960. Ouvrir **R** et charger cette série. En donner la représentation graphique. Constaté l'existence d'une saisonnalité mensuelle ( $d = 12$ ) et d'une croissance moyenne approximativement affine (modèle (8)). On constate qu'il subsiste une non-stationnarité en variance qui nécessitera une transformation de Box-Cox des données afin de les stationnariser en variance (cf. §2.2).

Si la régression  $t \mapsto m(t, \beta)$  n'est pas connue et si l'observation est  $X_1, X_2, \dots, X_n$ , on estimera  $\beta$  par MCO. Cette estimation pourra être améliorée dans un second temps par MCG une fois identifiée et estimée la structure du résidu  $(e_t)$  (cf. § 7.1 et annexe §10.3).

La prévision de  $X_{n+h}$ , reposera sur : (i) l'estimation de  $\beta$  ; (ii) la prévision du résidu  $e_{n+h}$  :

$$\widehat{X}_{n,h} = m(n + h, \widehat{\beta}) + \widehat{e}_{n,h}.$$

Deux facteurs d'imprécision apparaissent mais si  $n$  est grand et si  $\widehat{\beta}$  est une estimation consistante, c'est l'incertitude sur la prévision  $\widehat{e}_{n,h}$  à pas  $h$  qui est dominante.

### 2.1.2 Filtrage d'une tendance et/ou d'une saisonnalité par moyenne mobile

**Tendance sans saisonnalité** Supposons que  $X_t = m(t) + e_t$  pour un résidu  $(e_t)$  centré. Sans modèle sur la tendance  $t \mapsto m(t)$ , le filtrage par *moyenne mobile* (MA pour Moving Average) est une méthode locale non-paramétrique permettant de filtrer la tendance  $m(t) = E(X_t)$ .

Supposons que l'on observe  $X(n) = (X_1, X_2, \dots, X_n)$  et que  $m(t)$  soit à "variation régulière et lente" autour de  $t$ , approximativement affine sur  $[t - q, t + q]$ . Alors le filtre en  $(2q + 1)$ -MA :

$$\widehat{m}(t) = \frac{1}{2q + 1} \sum_{j=-q}^q X_{t+j} \text{ pour } q + 1 \leq t \leq n - q,$$

donne une "bonne approximation" de  $m(t)$  pour laquelle  $\widehat{e}_t = X_t - \widehat{m}(t)$  est approximativement centrée. Par exemple, si  $m(s) = a + bs$ ,  $\widehat{m}(t) = m(t) + \widetilde{e}_t$  où  $\widetilde{e}_t = \frac{1}{2q+1} \sum_{j=-q}^q e_{t+j}$  est "petit", parce que centré et de variance  $\frac{\sigma^2}{2q+1}$  si  $e \sim BB(0, \sigma^2)$ . L'étude portera alors sur la nouvelle série des résidus  $\widehat{e}_t = X_t - \widehat{m}(t)$ . D'autres filtres plus efficaces peuvent être utilisés.

**Tendance et saisonnalité** Supposons que  $X_t = m(t) + s_t + e_t$  où  $s$  est une  $d$ -saisonnalité centrée :

- on commence par estimer la tendance avec un filtre éliminant la saisonnalité ; par exemple si  $d$  est pair,

$$\widehat{m}_1(t) = \frac{1}{2d} (X_{t-d} + X_{t-d+1} + \dots + X_{t+d-1} + X_{t+d}) \text{ pour } \frac{d}{2} < t \leq n - \frac{d}{2}.$$

- on estime ensuite la saisonnalité en prenant la moyenne arithmétique  $w_k$  des déviations  $\{X_{k+jd} - \widehat{m}_1(k + jd), \frac{d}{2} < k + jd \leq n - \frac{d}{2}\}$  ;

- on recentre cette estimation en  $\widehat{s}_k = w_k - \bar{w}$ . La *série désaisonnalisée* est  $d_t = X_t - \widehat{s}_t$ .

- on termine en réestimant la tendance par filtrage en moyenne mobile à partir de la série désaisonnalisée  $(d_t)$ , soit  $\widehat{m}_2(t)$ .

Il reste alors à étudier la série recentrée  $\widehat{e}_t = X_t - (\widehat{m}_2(t) + \widehat{s}_t)$ ,  $t = 1, n$ .

### 2.1.3 Exercice et fonctions R utiles

Soit  $B$  l'opérateur de retard :  $BZ_t = Z_{t-1}$ .

1 - Déterminer les constantes  $a, b$  et  $c$  t.q. le filtre  $F(B) = 1 + aB + bB^2 + cB^3$  laisse invariant les tendances affines mais absorbe les 2-saisonnalités.

2 - Vérifier que le filtre  $F(B) = \sum_{i=-2}^2 a_i B^i$  de coefficients  $(a_{-2}, a_{-1}, a_0, a_1, a_2) = \frac{1}{9}(-1, 4, 3, 4, -1)$  laisse invariant les tendances polynomiales de degré  $\leq 3$  mais absorbe les 3-saisonnalités.

Fonctions R utiles :

Package `stats` : `decompose` (en  $T(t) + S(t) + e(t)$ ), `monthplot` (ajustement saisonnier)

Package `forecast` : `seasadj` (désaisonnalise une série)

Package `tseries` : `kpss.test` (test de stationnarité)

### 2.1.4 Stationnarisation par différenciation

**Stationnarisation de la moyenne.** La constatation de base est la suivante : notons

$$\nabla Z_t = Z_t - Z_{t-1}$$

l'opérateur de différence d'ordre 1 : si  $X_t = a + bt + e_t$ ,  $\nabla X_t = b + \varepsilon_t$  où  $\varepsilon_t = \nabla e_t = e_t - e_{t-1}$  est stationnaire centré si  $e$  l'est. Ainsi, la différenciation  $\nabla$  stationnarise en moyenne une série dont la tendance est affine en préservant la stationnarité du résidu. Il est facile d'étendre ce résultat aux tendances polynomiales de degré  $k$  qui sont absorbées par la  $k$ -itérée  $\nabla^k$  de  $\nabla$  :

$$\nabla^k P(t) = c \text{ si } P \text{ est un polynome de degré } k.$$

Ainsi, si  $X_t = P(t) + e_t$  où  $P$  est polynomiale de degré  $k$ , alors  $Y_t = \nabla^k X_t = c + \varepsilon_t$  où  $\varepsilon$  est stationnaire si  $e$  l'est. Cette constatation justifie l'utilisation de la différenciation  $\nabla$  ou de ses premières itérées  $\nabla^k$  lorsque la tendance d'une série est régulière et varie lentement. Un inconvénient de cette différenciation est de "compliquer" éventuellement la structure du bruit résiduel  $\varepsilon$  (i.e. si  $e$  est un  $BB$ ,  $\varepsilon$  devient un  $MA(1)$ ).

**Elimination d'une  $d$ -saisonnalité.** Soit  $\nabla_d$  la différenciation d'ordre  $d$  (ne pas confondre avec  $\nabla^d$ ) :

$$\nabla_d Z_t = Z_t - Z_{t-d}.$$

Si  $s = (s_t)$  est une  $d$ -saisonnalité,  $\nabla_d s_t \equiv 0$  : la  $d$ -différenciation absorbe les  $d$ -saisonnalités. Par exemple, si  $X_t = s_t + e_t$ ,  $\nabla_d X_t = \varepsilon_t$  où  $\varepsilon_t = \nabla_d e_t$  est stationnaire si  $e$  l'est. Remarquons que  $\nabla_d$  stationnarise également une SC à tendance affine : si  $X_t = a + bt + s_t + e_t$ ,  $\nabla_d Z_t \equiv d \times b + \varepsilon_t$  est stationnaire. Cependant on optera souvent pour une composition des deux différenciations  $\nabla \nabla_d = \nabla_d \nabla$  dans une telle situation.

**Retour à la série d'origine** La technique de différenciation permet de passer d'une série  $X$  non stationnaire en moyenne à une série  $Y$  stationnaire en moyenne. Pour la prédiction de  $X$ , on passera par la prédiction sur  $Y$  puis on remontera par filtrage inverse à la prédiction recherchée sur  $X$ . Illustrons cela à travers un exemple : on observe  $X(n) = (X_1, X_2, \dots, X_n)$  et on veut prédire  $X_{n+2}$  ; on pense être en présence d'une tendance régulière pour la série  $X$  qui est trimestrielle. On propose donc la stationnarisation suivante :

$$Y_t = \nabla_4 \nabla X_t = (X_t - X_{t-1}) - (X_{t-4} - X_{t-5}) = X_t - X_{t-1} - X_{t-4} + X_{t-5}, \quad t = 6, n$$

On constate un *rétrécissement* du domaine d'observation pour la série différenciée  $Y$  ( $n - 5$  observations  $Y_t$  au lieu de  $n$  pour  $X$ ). On prévoit  $Y_{n+2}$  par  $\hat{Y}_{n,2}$  puis on calcule la prédiction associée  $\hat{X}_{n,2} = \hat{Y}_{n,2} + \hat{X}_{n,1} + X_{n-2} - X_{n-3}$ , et à nouveau  $\hat{X}_{n,1} = \hat{Y}_{n,1} + X_n + X_{n-3} - X_{n-4}$ . On trouve :

$$\hat{X}_{n,2} = \hat{Y}_{n,2} + \hat{Y}_{n,1} + X_n + X_{n-2} - X_{n-4}.$$

### 2.1.5 Exercices

1. Soit  $\{e_t\}$  un bruit stationnaire centré,  $a$  et  $b$  deux constantes,  $X_t = (a + bt)s_t + \varepsilon_t$  et  $Y_t = (\nabla_{12})^2 X_t$ . Vérifier que  $Y$  est stationnaire. Déterminer sa covariance en fonction de celle de  $e$ .
2. (a) - Simuler la série  $X_t = a + bt + \sigma e_t$  pour  $t = 1, 200$ ,  $a = 2, b = 0.1, \sigma = 2$ ,  $e$  étant *i.i.d.*( $\mathcal{N}(0, 1)$ ). Graphe, acvf et pacf empiriques de  $(X_t, t = 1, 200)$ . Calculer l'estimation des MCO de  $(a, b)$  et l'estimation déduite de  $\sigma^2$  (cf. Annexe §10.3). Tester  $b = 0$  contre  $b > 0$ .  
 (b) - Simuler la série  $X_t = a + bt + \eta_t$  pour  $t = 1, 200$ ,  $a = 2, b = 0.1$ ,  $\eta$  étant le modèle  $AR(1)$  d'équation :

$$\eta_t = \rho \eta_{t-1} + e_t, t = 0, 200 \text{ pour } \rho = \frac{\sqrt{3}}{2}.$$

Vérifier que  $Var(\eta_t) = 4$ . Tracer la série  $(\eta_t, t = 1, 200)$ , son acvf et sa pacf empiriques. Estimer  $b$  par MCO. Comment tester  $b = 0$  ?

3. (a) - Soit  $(e_t, t = 1, 200) \sim i.i.d.(\mathcal{N}(0, 1))$ ,  $X_t = t + e_t$ ,  $Y_t = \sqrt{t}e_t$  et  $Z_t = \cos \frac{\pi t}{4} + e_t$  pour  $t = 1, 200$ . Quelles sont les acvf empiriques de  $X, Y$  et  $Z$ . Quelles conclusions en tirez-vous ?  
 (b) - Soit  $X_{1,t} = X_t - X_{t-1}$  pour  $t = 2, 200$  et  $Z_{1,t} = Z_t - Z_{t-8}$  pour  $t = 9, 200$ . Mêmes questions sur  $X_1$  et  $Z_1$ .

## 2.2 Non-stationnarité en variance

C'est la situation où  $Var(X_t) = \sigma^2(t)$  dépend de  $t$ . Réexaminons la série **AirPassengers** : la variance  $\sigma^2(t)$  est une fonction croissante de la moyenne  $m(t)$  :

$$\sigma^2(t) = f(m(t)). \tag{9}$$

Soit  $cv(X) = \frac{\sigma(X)}{E(X)}$  le *coefficient de variation* d'une v.a.  $X > 0$ . De nombreux phénomènes  $X$  sont à  $cv(X) = \kappa$  constant : l'ordre de grandeur de la variabilité de  $X$  (son écart type) est du même ordre que la moyenne de  $X$  (nombre de malades d'un canton, trafic de voyageurs, chiffre d'affaires d'une entreprise). Dans ce cas, le modèle pour (9) est  $\sigma(t) = \kappa m(t)$ .

### 2.2.1 Transformation stabilisant la variance

Une solution consiste à trouver une transformation  $g(X_t)$  telle que  $Var(g(X_t))$  est approximativement constante. L'approche est la suivante : supposons que  $X \sim \mathcal{N}(m, \kappa^2 f(m))$  et que  $\kappa$  est petit. Le résultat (cf. (45) de l'annexe, §10.4) précise la loi approchée de la déformation d'une gaussienne de petite variance si  $g$  est différentiable autour de  $m$ ,

$$g(X) \sim \mathcal{N}(g(m), \kappa^2 f(m)g'(m)^2).$$

La variance est stabilisée si  $f(m)g'(m)^2 = c$ . D'où l'équation différentielle sur  $g$  :

$$g'(m) = \frac{c}{\sqrt{f(m)}},$$

### 2.2.2 Trois exemples

- $X$  est à  $cv$  constant  $\kappa$  petit,  $\sigma^2 = f(m) = \kappa^2 m^2$  et  $g'(m) = \frac{c}{m}$  : on obtient  $g(m) = \kappa \log(m)$  : le log stabilise la variance d'un modèle à  $cv$  constant et petit.
- $\sigma^2 = \kappa^2 m$ ,  $\kappa$  petit :  $\sigma^2 = f(m) = \kappa^2 m$  et  $g(m) = \kappa m^{-\frac{1}{2}}$  stabilise la variance.

– Plus généralement, la transformation de *Box-Cox* :

$$f(m) = \frac{m^a}{a}, a > 0 \text{ et } f(m) = \log(m) \text{ pour } a = 0,$$

stabilise la variance de  $X$  si  $\sigma^2 = \kappa^2 m^{2-2a}$ ,  $\kappa$  petit, la situation frontière  $a = 0$  correspondant au cas d'un *cv* constant.

### 2.2.3 Exercices et fonctions R utiles

On considère la série  $X_t = (a + bt)Z_t$ ,  $t \geq 0$ ,  $a$  et  $b > 0$  où  $(Z_t)$  est une suite stationnaire de v.a. sur  $\mathbb{R}^+$ ,  $E(Z_1) = \mu > 0$ ,  $Var(Z_1) = \sigma^2$ .

a - Calculer l'espérance et la variance de  $X_t$ . Que vaut le coefficient de variation de  $X_t$  ?

b - Soit  $Y_t = \nabla X_t = X_t - X_{t-1}$  pour  $t \geq 1$ .  $(Y_t)$  est-elle stationnaire en espérance ? en variance ?

c - Soit  $W_t = \log(\nabla X_t)$ . Montrer que  $(W_t)$  est approximativement stationnaire si  $t$  est grand et/ou si  $a \gg b$ .

d - On observe  $X(n) = (X_1, X_2, \dots, X_n)$  et on dispose d'une prévision  $\widehat{W}_{n,1}$  de  $W_{n+1}$  en fonction des  $W_t$  passés. Comment prévoir  $X_{n+1}$  ?

*Fonctions R utiles :*

Dans le package `forecast` : `BoxCox`, `InvBoxCox`.

Dans le package `stats` : `diff.ts`, `diffinv`

## 3 Prédiction $L^2$ et auto-corrélation partielle

### 3.1 Prédiction au second ordre : le cas général

Soient  $Y, W = (W_1, W_2, \dots, W_n)$ ,  $(n + 1)$  v.a.r. du second ordre. Supposons connues les moyennes et les variances-covariances suivantes :

$$\begin{aligned} \mu_Y &= E(Y), \mu_W = {}^t(\mu_i = E(W_i), i = 1, n) \in \mathbb{R}^n; \\ \gamma &= Cov(W, Y) \in \mathbb{R}^n; Var(Y) \in \mathbb{R}; \\ \Gamma &= Cov(W), \text{ matrice } n \times n. \end{aligned}$$

L'objectif est de prédire au mieux, pour le critère de l'*écart quadratique moyen* (EQM),  $Y$  à partir d'une combinaison affine des  $W_i, i = 1, n$ . Il s'agit donc de trouver la combinaison affine  $\widehat{Y} = a_0 + a_1 W_1 + a_2 W_2 + \dots + a_n W_n$  qui minimise l'EQM :

$$S(a) = E(Y - \widehat{Y})^2 = E\{Y - (a_0 + a_1 W_1 + a_2 W_2 + \dots + a_n W_n)\}^2.$$

On parle aussi de la *régression affine* de  $Y$  sur  $W$  :  $\widehat{Y}$  est la projection orthogonale (pour le produit scalaire de la covariance de  $Y$ ) sur l'espace affine engendré par les  $W$ .

**Proposition :** La prédiction optimale s'écrit  $\widehat{Y} = \mu_Y + {}^t a(W - \mu_W)$  où  $a = {}^t(a_1, a_2, \dots, a_n)$  vérifie :

$${}^t a \Gamma = \gamma \text{ et } EQM(\widehat{Y}) = Var(Y) - {}^t a \gamma. \quad (10)$$

*Preuve :* Soit  $L^2$  l'espace de Hilbert des variables centrées de variances finies muni du produit scalaire de la covariance,  $\langle U, V \rangle = Cov(U, V)$ . Recentrons les  $(n + 1)$  variables  $Y$  et  $W$ . Minimiser le critère de l'EQM revient à projeter orthogonalement  $Y$  sur l'espace linéaire des  $W$ . Cette projection existe toujours et l'orthogonalité de  $(Y - \widehat{Y})$  aux  $W_i, i = 1, n$ , s'écrit, notant  $\Gamma_i$  la  $i$ -ième colonne de  $\Gamma$  :

$$\begin{aligned} Cov(Y - \widehat{Y}, W_i) &= 0, i = 1, n, \text{ soit : } {}^t a \Gamma_i = \gamma_i \text{ pour } i = 1, n, \\ \text{c-à-d : } &{}^t a \Gamma = \gamma. \end{aligned} \quad (11)$$

Un tel  $a$  existe toujours et est unique si  $\Gamma$  est régulière :  $a = \Gamma^{-1} \gamma$ . Un calcul direct donne l'erreur de prédiction. .... ■

## 3.2 Le cas d'un processus stationnaire

### 3.2.1 Prédiction à l'horizon $h$

Un objectif central dans l'étude des SC est la prédiction à l'horizon  $h > 0$  de  $X_{n+h}$  si on observe  $X(n) = {}^t(X_1, X_2, \dots, X_n)$ . Si la série est centrée et si  $X$  est stationnaire d'acvf  $\gamma(\cdot)$ , la propriété précédente nous dit que cette prévision vaut,

$$\begin{aligned}\widehat{X}_{n,h} &= {}^t a X(n) \text{ où } {}^t a \Gamma = \gamma, \text{ avec} \\ \gamma &= {}^t Cov(X_{n+h}, X(n)) = {}^t(\gamma(n+h-1), \dots, \gamma(h+1), \gamma(h)) \text{ et} \\ \Gamma &= Cov(X(n)) = (\gamma(i-j))_{i,j=1}^n.\end{aligned}$$

L'erreur de prévision est donnée par (10). Si  $\gamma(\cdot)$  n'est pas connue, on la remplacera par son estimation ; dans ce cas, l'erreur de prévision donnée par (10) n'est qu'approchée puisqu'un facteur d'imprécision sur les paramètres  $\widehat{\gamma}(k)$  s'ajoute à l'incertitude sur le processus.

### 3.2.2 Algorithme récursif de prédiction ; prévision de valeurs manquantes

Une difficulté dans l'obtention de  $\widehat{X}_{n,h}$  réside dans l'inversion de la matrice  $\Gamma$  de taille  $n \times n$  lorsque  $n$  est grand. Il existe des *algorithmes récursifs* qui lèvent cette difficulté (i.e. algorithme de Durbin-Levinson, cf. BD). Leur principe est le suivant : une fois calculée  $\widehat{X}_{n,1}$ , on cherche à obtenir par une récursion simple (sans inversion de matrice) la prédiction  $\widehat{X}_{n+1,1}$  sur la base des observations et de  $\widehat{X}_{n,1}$ .

La prédiction  $L^2$  permet également d'estimer une *valeur manquante*  $X_k$ ,  $1 < k < n$  de  $X^k(n) = (X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$  :

$$\widehat{X}_k(n) = {}^t a X^k(n) \text{ où } {}^t a \Gamma = \gamma \text{ pour } \Gamma = Cov(X^k(n)) \text{ et } \gamma = {}^t Cov(X_k, X^k(n)).$$

### 3.2.3 Exercices

#### 1. Prévision optimale sur la dernière observation

a - Préciser la loi conditionnelle  $\mathcal{L}(Y | X = x)$  si  ${}^t(X, Y) \sim \mathcal{N}_2({}^t(a, b), \Sigma)$  de covariance  $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  inversible.

b - Soit  $X_1, X_2, \dots$  un processus stationnaire au second ordre de moyenne  $m$  et d'ACF  $\rho(\cdot)$ . Démontrer que le prédicteur optimal de  $X_{n+h}$  en fonction de la seule observation  $X_n$  et de la constante 1 est  $\widehat{X}_{n,h} = \rho(h)X_n + m(1 - \rho(h))$ . Quelle est la loi de cette prédiction si  $X$  est gaussienne ?

#### 2. Prédiction de valeurs manquantes.

a -  $X_1, X_2, X_4$  et  $X_5$  sont les observations d'un MA(1) :  $X_t = Z_t + \theta Z_{t-1}$  où  $Z \sim BB(0, \sigma^2)$ . Déterminer la meilleure prédiction linéaire de la *valeur manquante*  $X_3$  en précisant chaque fois les erreurs de prédictions : (a1) en fonction de  $X_2, X_4$  ; (a2) en fonction de  $X_1, X_2$  ; (a3) en fonction de  $X_1, X_2, X_4$  et  $X_5$ .

b - Mêmes questions mais  $X$  est l'AR(1) stationnaire :  $X_t = \phi X_{t-1} + Z_t$ .

## 3.3 Auto-corrélation partielle

### 3.3.1 Le cas général

Soient  $Y, Z$  et  $W = (W_1, W_2, \dots, W_n)$ ,  $(n+2)$  v.a. réelles de variances finies, de moyennes et de variances-covariances connues. Soit  $\widehat{Y}$  (resp.  $\widehat{Z}$ ) la régression affine de  $Y$  (resp.  $Z$ ) sur  $W$  et  $\widehat{\varepsilon}(Y | W)$  (resp.  $\widehat{\varepsilon}(Z | W)$ ) le résidu de cette régression :

$$Y = \widehat{Y} + \widehat{\varepsilon}(Y | W) \text{ et } Z = \widehat{Z} + \widehat{\varepsilon}(Z | W).$$

**Définition :** Le coefficient de corrélation partielle  $\rho_P(Y, Z | W)$  entre  $Y$  et  $Z$  à  $W$  connu est le coefficient de corrélation entre les résidus  $\widehat{\varepsilon}(Y | W)$  et  $\widehat{\varepsilon}(Z | W)$  des régressions de  $Y$  sur  $W$  et de  $Z$  sur  $W$  :

$$\rho_P(Y, Z | W) = \rho(\widehat{\varepsilon}(Y | W), \widehat{\varepsilon}(Z | W)). \quad (12)$$

Ce coefficient mesure le degré de liaison affine entre  $Y$  et  $Z$  une fois retirée l'information affine qu'apporte  $W$  à l'une et à l'autre des variables  $Y$  et  $Z$ . Par exemple, si on observe un  $n$ -échantillon de la taille  $Y$ , du poids  $Z$  et de l'âge  $W$  d'une population, la corrélation  $\rho_P(Y, Z | W)$  est le bon indice de liaison affine entre taille et poids à "âge fixé" ; la corrélation habituelle  $\rho(Y, Z)$ , plus grande que la corrélation partielle, intègre le fait que taille et poids sont deux fonctions croissantes de l'âge et en fait masque la liaison intrinsèque entre la taille et le poids.

*Exemple :* Soit  $(Y, Z, W) \in \mathbb{R}^3$  de moyenne  $m$ , de variances  $\sigma^2 \in \mathbb{R}^3$  et de matrice de corrélations  $\rho$ . On peut alors expliciter la forme analytique de  $\rho_P(Y, Z | W)$ . Les deux régressions sont :

$$\begin{aligned} Y - m_Y &= \rho_{YW} \frac{\sigma_Y}{\sigma_W} (W - m_W) + \widehat{\varepsilon}(Y | W) \text{ et} \\ Z - m_Z &= \rho_{ZW} \frac{\sigma_Z}{\sigma_W} (W - m_W) + \widehat{\varepsilon}(Z | W). \end{aligned}$$

Reste à calculer la corrélation simple  $\rho(\widehat{\varepsilon}(Y | W), \widehat{\varepsilon}(Z | W))$ . Un calcul direct conduit à :

$$\rho_P(Y, Z | W) = \frac{\rho_{YZ} - \rho_{YW}\rho_{ZW}}{\sqrt{1 - \rho_{YW}^2}\sqrt{1 - \rho_{ZW}^2}}.$$

### 3.3.2 Auto-corrélation partielle pour un processus stationnaire

Soit  $X$  un processus stationnaire au second ordre.

**Définition :** La fonction d'*auto-corrélation partielle* (notée *pacf*) de  $X$  à distance  $h > 0$  est définie par :

$$\alpha(h) = \rho_P(X_0, X_h | X_1, X_2, \dots, X_{h-1}),$$

$\alpha(h)$  est la corrélation entre  $X_0$  et  $X_h$  aux valeurs intermédiaires  $(X_1, X_2, \dots, X_{h-1})$  fixées. Par convention, on pose  $\alpha(0) = 1$ .

*Lien avec la prédiction.*

La prédiction de  $X_h$  sur la base de  $X_0^{h-1} = (X_0, X_1, X_2, \dots, X_{h-1})$  est donnée par  $\widehat{X}_{h,1} = {}^t a_h X_0^{h-1}$  où  $a_h = \Gamma_h^{-1} \gamma_h$ ,  $\Gamma_h = (\gamma(i-j))_{i,j=0}^{h-1}$  et  $\gamma_h = {}^t(\gamma(h), \gamma(h-1), \dots, \gamma(1))$ . On a la propriété suivante (cf. BD) :

**Propriété :** Soit  $a_{0,h}$  le coefficient de  $X_0$  dans  $\widehat{X}_{h,1}$ . Alors la *pacf* de  $X$  à distance  $h$  vaut  $a_{0,h}$  :

$$\alpha(h) = \rho(X_0, X_h | X_1, X_2, \dots, X_{h-1}) = a_{0h}.$$

Anticipons sur les modèles ARMA : une propriété caractéristique des  $AR(p)$  est que la *pacf* d'un  $AR(p)$  est nulle au delà de  $p$  :  $\alpha(h) = 0$  si  $h > p$ . Cette propriété est la propriété duale de celle d'un  $MA(q)$  pour lequel l'acf vérifie :  $\rho(h) = 0$  si  $h > q$ .

### 3.3.3 Exercices

1. Soient  $X, Y$  et  $Z$  trois variables *i.i.d.*  $(0, 1)$ ,  $X_1 = X + Z$  et  $Y_1 = Y + Z$ . Calculer  $\rho(X_1, Y_1 | Z)$ .
2. Utilisant la définition de la *pacf*, démontrer que la *pacf* d'un  $AR(1)$  (cf. §4.2) est nulle à partir de la distance 2. Généraliser au cas d'un  $AR(p)$ .

## 4 Modèles ARMA

Les modèles ARMA (Auto-Regressive with Moving Average) sont des modèles stationnaires de petite dimension dont la covariance approche une large famille de covariances. En particulier, pour toute covariance  $\gamma(\cdot)$  et tout entier  $K > 0$ , il existe un modèle ARMA (et même un AR)  $X$  tel que  $\gamma(k) = \gamma_X(k)$  pour  $0 \leq k \leq K$ . Ces modèles vont être définis comme des *modèles linéaires* associés à un BB. Commençons par définir de tels modèles linéaires.

### 4.1 Processus linéaire

Soit  $\psi$  une suite de  $l^1$  ( $\sum_{\mathbb{Z}} |\psi_i| < \infty$ ) et  $\varepsilon \sim BB(0, \sigma^2)$ . Un processus linéaire est :

$$X_t = \sum_{i \in \mathbb{Z}} \psi_i \varepsilon_{t-i}, \quad t \in \mathbb{Z}. \quad (13)$$

$X_t$  est donc une moyenne glissante infinie ( $MA(\infty)$ ) et bilatérale du bruit  $\varepsilon$  pour des poids  $\psi$ . La série définissant  $X_t$  converge en probabilité et dans  $L^2$  sous la condition  $\psi \in l^1$  (cf. §10.2). L'acvf de  $X$  vaut :

$$\gamma_X(h) = \sigma^2 \sum_{i \in \mathbb{Z}} \psi_i \psi_{i+h}. \quad (14)$$

En particulier,  $\gamma(h)$  tend vers 0 si  $h \rightarrow \infty$ .

### 4.2 Modèle ARMA( $p, q$ )

Un ARMA( $p, q$ ) associé à  $\varepsilon \sim BB(0, \sigma^2)$  et de paramètres  $\phi \in \mathbb{R}^p$ ,  $\theta \in \mathbb{R}^q$ , est un processus stationnaire  $X$  vérifiant pour chaque  $t \in \mathbb{Z}$  :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (15)$$

Notons  $P(z)$  (resp.  $Q(z)$ ) le polynôme associé à la composante AR( $p$ ) (resp. MA( $q$ )) :

$$\begin{aligned} P(z) &= 1 - \phi_1 z - \dots - \phi_p z^p \quad (\phi_p \neq 0); \\ Q(z) &= 1 + \theta_1 z + \dots + \theta_q z^q \quad (\theta_q \neq 0). \end{aligned}$$

On supposera toujours que  $P$  et  $Q$  sont sans facteur commun. Si  $B$  est l'opérateur de retard défini pour tout  $t \in \mathbb{Z}$  par  $BX_t = X_{t-1}$ , l'équation génératrice de l'ARMA s'écrit :

$$P(B)X_t = Q(B)\varepsilon_t. \quad (16)$$

**Propriétés :** (15) admet une solution stationnaire ssi  $P(z) \neq 0$  pour  $|z| = 1$ .

*Preuve :* Si  $P(z) \neq 0$  pour  $|z| = 1$ , alors  $\exists \delta > 0$  t.q.  $P^{-1}$  admette un développement de Laurent uniformément convergent sur la couronne  $1 - \delta \leq |z| \leq 1 + \delta$ ,

$$P^{-1}(z) = \sum_{i \in \mathbb{Z}} \chi_i z^i \quad \text{avec } \chi_0 = 1, \quad (17)$$

et la suite des  $(\chi_i)$  est à décroissance exponentielle. On en déduit que  $\frac{Q(z)}{P(z)} = \sum_{i \in \mathbb{Z}} \psi_i z^i$ , avec  $\psi_0 = 1$ , où  $(\psi)$ , étant à décroissance exponentielle, est dans  $l^1$ . Définissons alors :

$$X_t = \sum_{i \in \mathbb{Z}} \psi_i \varepsilon_{t-i}. \quad (18)$$

$X$  est solution de (16) et son acvf est donnée par (14). ..... ■

Les ARMA sont donc des processus linéaires particuliers dont les poids  $\psi$  décroissent exponentiellement ;  $X$  est un AR( $p$ ) si  $q = 0$  ;  $X$  est un MA( $q$ ) si  $p = 0$ .

### 4.2.1 Causalité, inversibilité

**Causalité et innovation de  $X$**  Une représentation (15) est *causale* si la représentation (18) de  $X_t$  ne fait intervenir que les bruits  $\varepsilon_s, s \leq t$  : notons  $Y_u^v$  le sous-espace linéaire de  $L^2$  engendré par les v.a.  $\{Y_t, u \leq t \leq v\}$ . Si la représentation de  $X$  est causale,  $X_{-\infty}^t \subseteq \varepsilon_{-\infty}^t$ .  $\varepsilon_t$  est l'*innovation* de  $X$  à l'instant  $t$ , caractérisée par :

$$X_{-\infty}^t = X_{-\infty}^{t-1} \oplus (\varepsilon_t) \text{ et } Cov(\varepsilon_t, X_s) = 0 \text{ si } s < t. \quad (19)$$

**Propriété :** (15) est causale ssi  $P(z) \neq 0$  pour  $|z| \leq 1$ , c-à-d si  $P$  n'a pas de racine sur le disque unité.

*Preuve :* En effet, cette condition assure que le développement de Laurent (17) de  $P(z)^{-1}$  est le développement ordinaire en série entière  $P^{-1}(z) = \sum_{i \geq 0} \chi_i z^i$ , convergent si  $|z| \leq 1$ . . . . . ■

Si  $P(z) = \prod_{k=1}^p (1 - \frac{z}{z_k})$ , le développement de  $P(z)^{-1}$  s'obtient à partir des développements des fractions rationnelles :

$$\frac{1}{1 - \frac{z}{z_k}} = \sum_{i \geq 0} (z_k)^{-i} z^i \text{ pour } |z| \leq 1.$$

**Inversibilité de l'ARMA** La représentation (15) est *inversible* si pour tout  $t$  :

$$\varepsilon_t = \sum_{i \geq 0} \varphi_i X_{t-i}. \quad (20)$$

La représentation est inversible si  $Q(z) \neq 0$  pour  $|z| \leq 1$ , c-à-d si le polynôme MA n'a pas de racine sur le disque unité. Pour un ARMA inversible,  $\varepsilon_{-\infty}^t \subseteq X_{-\infty}^t$ .

Ainsi si la représentation ARMA est causale et inversible, on a l'égalité entre les espaces engendrés par  $X$  et par  $\varepsilon$  :

$$\text{Si } X \text{ est causal et inversible : } \forall t : \varepsilon_{-\infty}^t = X_{-\infty}^t.$$

### 4.2.2 Exercice

Parmi les modèles suivants, lesquels sont causaux? inversibles? ( $e$  est un  $BB(0, \sigma^2)$ ) et pour ceux qui sont causaux, calculer les 6 premiers coefficients  $\psi_i$  de la représentation causale  $X_t = \sum_{i \in \mathbb{Z}} \psi_i e_{t-i}, t \in \mathbb{Z}$  :

- $X_t = 0.8X_{t-1} + e_t + 1.2e_{t-1}$
- $X_t + 1.9X_{t-1} + 0.88X_{t-2} = e_t + 0.2e_{t-1} + 0.7e_{t-2}$ .
- $X_t - 0.9X_{t-2} = e_t - 0.16e_{t-2}$ .

### 4.2.3 Exemples d'ARMA

**Les modèles en moyenne glissante  $MA(q)$**  Considérons :

$$MA(q) : X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (21)$$

et posons  $\theta_0 = 1$ .  $X$  a pour acvf :

$$\gamma(h) = \sigma^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h} \text{ pour } |h| \leq q \text{ et } \gamma(h) = 0 \text{ sinon.}$$

La portée de l'acvf d'un  $MA(q)$  est  $q$ . Cette propriété est caractéristique des MA (BD) : un processus stationnaire dont l'acvf est de portée  $q$  peut s'écrire comme un  $MA(q)$ . La représentation est inversible si  $Q(z) = 1 + \sum_{j=1}^q \theta_j z^j$  ne s'annule pas sur  $|z| \leq 1$  : dans ce cas,  $\forall t : \varepsilon_{-\infty}^t = X_{-\infty}^t$ .

**Les modèles d'auto-régression  $AR(p)$**  Le modèle stationnaire :

$$AR(1) : X_t = \phi X_{t-1} + \varepsilon_t \text{ où } \varepsilon \sim BB(0, \sigma^2), \quad (22)$$

existe ssi  $|\phi| \neq 1$ . La représentation est causale si  $|\phi| < 1$  puisque la racine de  $P(z) = 1 - \phi z$  est  $z_1 = \phi^{-1}$ . Dans ce cas, l'écriture linéaire de  $X$  et son acvf sont :

$$X_t = \sum_{i \geq 0} \phi^i \varepsilon_{t-i} \text{ et } \gamma(h) = \sigma^2 \sum_{i \geq 0} \phi_i \phi_{i+h} = \gamma(0) \phi^{|h|} \text{ où } \gamma(0) = \frac{\sigma^2}{1 - \phi^2}.$$

La propriété fondamentale (19) de l'innovation  $\varepsilon_t$  est d'être non-corrélée aux valeurs passées  $\{X_s, s < t\}$  et d'être le bon complément à  $X_{t-1}$  pour expliquer  $X_t$ .

**Exemple :** soit l'équation  $AR(2)$ ,  $X_t = 0.7X_{t-1} - 0.1X_{t-2} + \varepsilon_t$ . Le polynôme caractéristique associé est  $P(z) = 1 - 0.7z + 0.1z^2$  de racines  $z_1 = 2$  et  $z_2 = 5$ . Il existe donc une solution stationnaire et la représentation est causale. L'écriture linéaire associée utilise la décomposition en éléments simples,

$$\begin{aligned} \frac{1}{P(z)} &= a_1 \frac{1}{1 - \frac{z}{z_1}} + a_2 \frac{1}{1 - \frac{z}{z_2}} \text{ avec } a_1 = \frac{z_2}{z_2 - z_1} \text{ et } a_2 = \frac{z_1}{z_1 - z_2}, \\ X_t &= \sum_{i \geq 0} \psi_i \varepsilon_{t-i} \text{ où } \psi_i = a_1 z_1^{-i} + a_2 z_2^{-i}, \\ \gamma(h) &= \frac{\sigma^2 z_1^2 z_2^2}{(z_1 z_2 - 1)(z_2 - z_1)} \left\{ \frac{z_1^{1-h}}{z_1^2 - 1} - \frac{z_2^{1-h}}{z_2^2 - 1} \right\}. \end{aligned}$$

$\gamma(h)$  est une somme pondérée des exponentielles  $z_1^{-h}$  et  $z_2^{-h}$  associées aux racines de  $P$ .

Si  $P$  était à racines complexes conjuguées,  $P(z) = (1 - \frac{z}{z_1})(1 - \frac{z}{\bar{z}_1})$  avec  $z_1 = r \exp i\omega$ , la condition de causalité est  $r > 1$ , l'écriture causale,

$$X_t = \sum_{j \geq 0} \psi_j \varepsilon_{t-j} \text{ avec } \psi_j = r^{-j} \times \frac{\sin(j+1)\omega}{\sin \omega}.$$

où  $j \mapsto \psi_j$  est une sinusoïde amortie ;  $|\gamma(h)|$  décroît en  $r^{-|h|}$ .

**Le modèle  $ARMA(1, 1)$**  Considérons l' $ARMA(1, 1)$  définit, pour  $\theta + \phi \neq 0$ , par :

$$ARMA(1, 1) : X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \text{ où } \varepsilon \sim BB(0, \sigma^2),$$

La représentation est *causale* si  $\phi < 1$ , de représentation linéaire,

$$\begin{aligned} X_t &= P(B)^{-1} Q(B) \varepsilon_t = \left\{ \sum_{i \geq 0} \phi^i B^i \right\} \{1 + \theta B\} \varepsilon_t \\ &= \sum_{i \geq 0} \psi_i \varepsilon_{t-i} \text{ où } \psi_0 = 1 \text{ et } \psi_i = \phi^{i-1} (\theta + \phi) \text{ si } i \geq 1. \end{aligned} \quad (23)$$

La représentation est *inversible* si  $|\theta| < 1$ . Intervertissant les rôles de  $\theta$  et  $\phi$ , on obtient,

$$\varepsilon_t = \sum_{i \geq 0} \varphi_i X_{t-i} \text{ avec } \varphi_0 = 0 \text{ et } \varphi_i = \theta^{i-1} (\theta + \phi) \text{ si } i \geq 1.$$

Comme on va le voir de façon générale pour les  $ARMA$ , l'acvf décroît comme un mélange des exponentielles associées aux  $p$  racines du polynôme AR, ici  $\gamma(h) \propto \phi^{|h|}$ . Plus précisément, utilisant la représentation causale (23), on obtient :

$$\begin{aligned} \gamma(0) &= \sigma^2 \left\{ 1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right\}, \\ \gamma(1) &= \sigma^2 (\theta + \phi) \left\{ 1 + \frac{\phi(\theta + \phi)}{1 - \phi^2} \right\} \text{ et} \\ \gamma(h) &= \phi^{h-1} \gamma(1) \text{ pour } h \geq 2. \end{aligned}$$

**Réexamen de la formule de Bartlett pour un AR ou un MA** Rappelons que la formule de Bartlett (6) donne la loi asymptotique du vecteur des corrélations empiriques,  $\widehat{\rho} \sim N_k(\rho, \frac{W}{n})$ , pour un processus linéaire  $X_t = \sum_{\mathbb{Z}} \psi_i \varepsilon_{t-i}$  associé à un BB générateur *i.i.d.*(0,  $\sigma^2$ ) de coefficients ( $\psi_i$ ) vérifiant :  $\sum_{\mathbb{Z}} |\psi_i| < \infty$  et  $\sum_{\mathbb{Z}} |i| \psi_i^2 < \infty$ . Ces deux conditions de sommabilité sont toujours satisfaites pour un ARMA. Spécifions les poids  $w_{ii}$  pour un MA(1) et un AR(1).

• *X est un MA(1) :  $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$  où  $\varepsilon \sim i.i.d.(0, \sigma^2)$ . Alors, notant  $\rho(1) = \frac{\theta}{1+\theta^2}$  la corrélation de  $X$  à distance 1, on a :*

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1) & \text{si } i = 1, \\ 1 + 2\rho^2(1) & \text{si } i > 1. \end{cases}$$

• *X est un MA(q) : si X est un MA(q),  $w_{ii} = 1 + 2(\rho^2(1) + \dots + \rho(q)^2)$  si  $i > q$ . Ce résultat est utile pour tester la nullité des corrélations de  $X$  à partir de la distance  $q + 1$  et pour juger ainsi du caractère MA de  $X$ .*

•• *X est un AR(1) :  $X_t = \phi X_{t-1} + \varepsilon_t$ ,  $|\phi| < 1$  et  $\varepsilon \sim i.i.d.(0, \sigma^2)$ . Alors :*

$$\text{Si } i \geq 1 : w_{ii} = (1 - \phi^{2i}) \frac{1 + \phi^2}{1 - \phi^2} - 2i\phi^{2i}.$$

•• *pacf d'un AR(p) : si X est un AR(p), sa pacf à distance  $h$  est nulle si  $h \geq p + 1$ , d'estimation  $\widehat{\alpha}(h) \sim \mathcal{N}(0, \frac{1}{n})$ . Ce résultat sera utile pour tester la nullité des pacf de  $X$  à partir de la distance  $p + 1$  et ainsi de juger du caractère AR de  $X$ .*

#### 4.2.4 Exercices

- (a) - *IC sur la moyenne* : un échantillon de taille  $n = 100$  d'un AR(1)  $X$  de moyenne inconnue  $m$ , de paramètres  $\phi = 0.6$  et de variance d'innovation  $\sigma_\varepsilon^2 = 2$ , a pour moyenne empirique  $\widehat{m}_{100} = \bar{x} = 0.271$ . Construire un intervalle de confiance au niveau 95% pour  $m$ . Ces données sont-elles compatibles avec l'hypothèse que  $X$  est centré ?  
 (b) - *IC sur la covariance* : on suppose que  $X$  est un AR(1) centré de paramètre inconnu  $\phi$ . Un 100-échantillon de  $X$  donne les estimations  $\widehat{\rho}_1 = 0.438$  et  $\widehat{\rho}_2 = 0.145$ . Construire un intervalle de confiance au niveau 95% pour  $\rho(1)$  et  $\rho(2)$ . Sur la base de ces deux IC, peut-on dire que les données observées sont compatibles avec l'hypothèse que  $X$  est un AR(1) de paramètre  $\phi = 0.8$  ?  
 (c) - Même données qu'en (b) mais pour  $X$  un MA(1) : IC au niveau 95% pour  $\rho(1)$  et  $\rho(2)$ . Sur la base de ces deux IC, peut-on dire que les données sont compatibles avec l'hypothèse que  $X$  est un MA(1) de paramètre  $\theta = 0.6$  ?
- Soit l'AR(2) :  $X_t = X_{t-1} - 0.5X_{t-2} + e_t$ . Que valent  $\gamma(0)$  et  $\gamma(1)$ . Montrer que :

$$\gamma(h) = 2^{-\frac{h}{2}} \left\{ \cos \frac{\pi h}{4} + \frac{1}{3} \sin \frac{\pi h}{4} \right\} \text{ si } h \geq 0.$$

### 4.3 ACVF d'un ARMA.

Si  $X$  est l'ARMA( $p, q$ ) de représentation (15), son acvf s'obtient à partir de sa représentation linéaire (18). Si la représentation est causale,  $X_t = \sum_{i \geq 0} \psi_i \varepsilon_{t-i}$ , on obtient :

$$\gamma(h) = \sigma^2 \sum_{i \geq 0} \psi_i \psi_{i+h}.$$

Comme  $\psi$ , l'acvf de  $X$  décroît exponentiellement. Donnons une autre approche de cette propriété de décroissance.

### 4.3.1 Equations de Yule-Walker.

Une équation de Yule-Walker sur la covariance  $\gamma$  de  $X$  s'obtient à partir de l'équation (15) à l'instant  $t$ , en la multipliant par  $X_{t-k}$ , puis en prenant l'espérance :

$$E(X_{t-k}\{X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}\}) = E(X_{t-k}\{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}\}).$$

Pour  $k \leq p$ , cela donne :

$$\gamma(k) - \sum_{i=1,p} \phi_i \gamma(k-1) = \begin{cases} \sigma^2 \sum_{i \geq 0} \theta_{k+i} \psi_i & \text{si } 0 \leq k \leq q, \\ 0 & \text{si } k \geq q+1, \end{cases} \quad (24)$$

et :

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0 \text{ si } k \geq k_0 = \max\{p, q+1\}.$$

A partir du rang  $k_0 = \max\{p, q+1\}$ ,  $\gamma$  suit la récurrence linéaire de polynôme caractéristique  $P$  : notant  $\{z_l\}$  les racines de  $P$ , la théorie des équations récurrentes linéaires nous dit que  $\gamma(h) = \sum_{l=1,p} a_l z_l^{-|h|}$  :  $\gamma(h)$  décroît exponentiellement vers 0 si  $h \rightarrow \infty$ .

### 4.3.2 Exercices

1. *Deux MA de même acvf* : vérifier que les deux modèles ci-dessous ont même acvf,

$$\begin{aligned} X_t &= Z_t + \theta Z_{t-1}, Z \sim BB(0, \sigma^2), 0 < |\theta| < 1 \text{ et} \\ Y_t &= V_t + \theta^{-1} V_{t-1}, V \sim BB(0, \theta^2 \sigma^2). \end{aligned}$$

Ce résultat montre qu'un  $MA(1)$  non-inversible admet une représentation équivalente inversible. Ceci reste vrai pour un  $MA(q)$  général.

2. *AR(1) non-causal et écriture causale associée* : soit  $\{X_t\}$  vérifiant les équations d'auto-régression :

$$X_t = \Phi X_{t-1} + e_t \text{ où } |\Phi| > 1 \text{ où } e \sim i.i.d.(0, \sigma^2). \quad (25)$$

(a) - Montrer que  $X_t = -\sum_{i=1}^{\infty} \Phi^{-i} e_{t+j}$  est (l'unique) solution stationnaire de (25) et que  $\gamma_X(h) = \gamma_X(0) \phi^{-|h|}$ .

(b) - Pour cette solution stationnaire, on définit la nouvelle suite :

$$\varepsilon_t = X_t - \frac{1}{\Phi} X_{t-1}. \quad (26)$$

Montrer que  $\{\varepsilon_t\} \sim BB(0, \tau^2)$  et exprimer  $\tau^2$  à partir de  $\sigma^2$  et  $\Phi$ .

Ce résultat montre que la solution stationnaire de l'équation *AR(1) non-causale* (25) est aussi l'*AR(1) causale* solution de (26). Ce résultat reste vrai pour un  $AR(p)$  général. De même, tout  $ARMA(p, q)$  admet une représentation équivalente qui soit à la fois causale (partie  $AR$ ) et inversible (partie  $MA$ ; cf. (5.4)).

3. *Génération d'un modèle ARMA(1, 1)* : soient  $e$  et  $\varepsilon$  deux  $BB$  centrés non-corrélés de variances respectives  $\sigma_e^2$  et  $\sigma_\varepsilon^2$  et la série  $Y = (Y_t)$  :

$$\begin{aligned} Y_t &= X_t + \varepsilon_t, t \geq 1 \text{ où } X_t \text{ est l'AR(1) stationnaire,} \\ X_t &= aX_{t-1} + e_t, t \geq 1, |a| < 1. \end{aligned}$$

(a) - Montrer que  $Y$  est stationnaire au second ordre et donner son acvf.

(b) - Vérifier que  $U_t := Y_t - aY_{t-1}$  étant non-corrélé à distance  $> 1$  est un  $MA(1)$ .

(c) - En conclure que  $Y$  est un  $ARMA(1, 1)$  dont on précisera les 3 paramètres en fonction de  $a$ ,  $\sigma_e^2$  et  $\sigma_\varepsilon^2$ .

4. Déterminer l'acvf de  $X_t = e_t + \theta_1 e_{t-1} + \theta_{12} e_{t-12}$  ( $e$  est un  $BB(0, \sigma^2)$ ).
5. Soit  $X$  le  $MA(2)$  :  $X_t = e_t + \theta e_{t-2}$  où  $e \sim BB(0, \sigma^2)$ . Déterminer l'acvf de  $X$ . Que vaut  $Var(X_1 + X_2 + \dots + X_n)$ ? Que vaut la pacf de  $X$  à distance 1?
6. Montrer que la pacf à distance 2 du  $MA(1)$  :  $X_t = e_t + \theta e_{t-1}$  où  $e \sim BB(0, \sigma^2)$  est  $\alpha(2) = -\frac{\theta^2}{1+\theta^2+\theta^4}$ .
7. On considère l' $AR(2)$  à trou stationnaire de résidu  $e$  un  $BB(0, \sigma^2)$  :

$$X_t = 0.8X_{t-2} + e_t. \quad (27)$$

- (a) - Vérifier que cette représentation est causale. Multipliant (27) par  $X_{t-1}$  et prenant l'espérance, vérifier que  $\gamma(1) = 0$  et qu'il en est de même des covariances à distances impaires. Que vaut  $\gamma(0)$ ? Calculer les covariances aux distances paires.
  - (b) Donner l'écriture linéaire de  $X$  et retrouver les résultats précédents.
  - (c) - Représenter les fonctions d'acvf et la pacf du modèle ainsi que leurs estimations empiriques pour un 200-échantillon de ce modèle.
8. Soit  $e = (e_t) \sim BB(0, 1)$  et la série  $X_t = e_t + e_{t-2}$ ,  $t \in \mathbb{Z}$ . Déterminer la prédiction optimale de  $X_3$  sur  $\{X_1, X_2, X_4, X_5\}$ . Quelle est l'erreur de prédiction? Calculer la pacf  $\alpha(4)$  de  $X$  à distance 4.

#### 4.4 Prédiction pour les ARMA

Notons  $\widehat{X}_{n,h}$  la prédiction de  $X_{n+h}$  sur les observations  $\{X_1, X_2, \dots, X_n\}$ .

Pour un  $AR(p)$  causal,  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$ , les résidus vérifient,  $Cov(\varepsilon_t, X_s) = 0$  si  $s < t$ . Supposons que  $n \geq p$ . Alors,

$$\begin{aligned} \widehat{X}_{n,1} &= \phi_1 X_n + \dots + \phi_p X_{n+1-p}, \\ X_{n+1} - \widehat{X}_{n,1} &= \varepsilon_{n+1} \text{ et } Var(X_{n+1} - \widehat{X}_{n,1}) = \sigma_\varepsilon^2. \end{aligned}$$

On obtient de façon récursive les prévisions et erreurs de prédiction à distance  $h = 2, 3, \dots$ . Par exemple :

$$\begin{aligned} \widehat{X}_{n,2} &= \phi_1 \widehat{X}_{n,1} + \phi_2 X_n + \dots + \phi_p X_{n+2-p} \\ &= (\phi_1^2 + \phi_2) X_n + (\phi_3 + \phi_1 \phi_2) X_{n-1} + \dots + (\phi_1 \phi_{p-1} + \phi_p) X_{n+2-p} + \phi_1 \phi_p X_{n+1-p}, \\ X_{n+2} - \widehat{X}_{n,2} &= \phi_1 (X_{n+1} - \widehat{X}_{n,1}) + \varepsilon_{n+2} = \phi_1 \varepsilon_{n+1} + \varepsilon_{n+2}, \quad Var(X_{n+2} - \widehat{X}_{n,2}) = \sigma_\varepsilon^2 (1 + \phi_1^2). \end{aligned}$$

Pour un  $ARMA(p, q)$  causal et inversible,  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \varepsilon_{t-q}$ , on a :  $Cov(\varepsilon_t, X_s) = 0$  si  $s < t$ . De plus :

$$X_{-\infty}^t = \varepsilon_{-\infty}^t, \quad X_t = \sum_{i \geq 0} \psi_i \varepsilon_{t-i} \text{ et } \varepsilon_t = \sum_{i \geq 0} \pi_i X_{t-i}.$$

Supposons dans un premier temps qu'on observe tous les  $X_s, s \leq n$  et que l'on veuille prévoir  $X_{n+h}$  à partir de  $X_{-\infty}^n = \varepsilon_{-\infty}^n$ . On a :

$$\begin{aligned} \widehat{X}_{n,h} &= P(X_{n+h} | X_{-\infty}^n) = P(X_{n+h} | \varepsilon_{-\infty}^n) = \sum_{i \geq h} \pi_i \varepsilon_{n+h-i}, \quad (28) \\ X_{n+h} - \widehat{X}_{n,h} &= \sum_{i=0}^{h-1} \pi_i \varepsilon_{n+h-i}, \quad Var(X_{n+h} - \widehat{X}_{n,h}) = \sigma_\varepsilon^2 \left( \sum_{i=0}^{h-1} \pi_i^2 \right). \end{aligned}$$

Deux problèmes se posent (28) :

- (i)  $\varepsilon$  n'étant pas observé, il faut calculer  $\varepsilon_t = \sum_{i \geq 0} \pi_i X_{t-i}$  pour le calcul effectif de  $\widehat{X}_{n,h}$ .
- (ii) Ne disposant que des observations  $\{X_1, X_2, \dots, X_n\}$ , on obtiendra une prévision approchée  $\widehat{X}_{n,h}^*$  en faisant  $X_s = 0$  si  $s \leq 0$  dans (28); pour  $n \rightarrow \infty$ ,  $\widehat{X}_{n,h}^* \rightarrow \widehat{X}_{n,h}$  et la prévision  $\widehat{X}_{n,h}^*$  est donc correcte pour  $n$  grand.

Si la représentation *ARMA* n'est pas causale et inversible, on commencera par la rendre causale et inversible par la technique de pivotage des racines expliquée au §5.4. Et si on veut obtenir la prévision exacte, on utilisera les formules exactes de prédiction linéaire données au §3.2.

Pour un *MA*( $q$ ) inversible,  $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$ , la prévision pour les *ARMA* s'applique encore :

$$\widehat{X}_{n,h} = \theta_h \varepsilon_n + \dots + \theta_q \varepsilon_{n-q} \text{ si } 0 \leq h \leq q \text{ et } \widehat{X}_{n,h} = 0 \text{ si } h \geq q + 1.$$

#### 4.4.1 Exercices et fonctions R utiles

1. On observe la réalisation  $X_1, X_2, \dots, X_n$  d'un processus stationnaire. Déterminer les prévisions à distance 1 et 2 pour un modèle : *AR*(1), *AR*(2), *MA*(1), *MA*(2) et *ARMA*(1, 1).
2. On considère le modèle  $X_t = \phi(X_{t-1} + X_{t-2}) + e_t$ ,  $\phi > 0$ , où  $e \sim BB(0, \sigma^2)$ . A quelle condition cette représentation est-elle causale? Déterminer la prédiction  $\widehat{X}_{n,3}$  de  $X_{n+3}$  sur  $\{X_1, X_2, \dots, X_n\}$ .

Fonctions R utiles :

Package `fArma` : `armaSim`, `armaFit`, `armaRoots`, `armaTrueacf`

Package `forecast` : `forecast.Arima`

Package `stats` : `arima.sim`, `ARMAacf`, `ARMAtoMA` (écriture *MA*( $\infty$ ) d'un *ARMA*), `predict.Arima`

## 5 Un peu de théorie spectrale

### 5.1 Covariance sommable et densité spectrale

Soit  $X$  un processus stationnaire de covariance  $\gamma$  sommable,  $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$ . La densité spectrale de  $X$  est la fonction  $f(\cdot)$  définie par,

$$f(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-ih\lambda} \gamma(h) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \cos(h\lambda) \gamma(h), \lambda \in [-\pi, \pi[.$$

La sommabilité de  $\gamma$  implique que la série définissant  $f$  soit absolument convergente.

**Propriété** :  $f$  est paire, positive et intégrable sur  $[-\pi, +\pi[$ . Réciproquement, toute fonction paire,  $\geq 0$  et intégrable sur  $[0, 2\pi[$  définit une acvf  $\gamma(\cdot)$  :

$$\gamma(k) = \int_{-\pi}^{+\pi} e^{ik\lambda} f(\lambda) d\lambda = \int_{-\pi}^{+\pi} \cos(k\lambda) f(\lambda) d\lambda. \quad (29)$$

*Preuve* :

• La parité de  $f$  découle de celle de  $\cos$  et de  $\gamma$ . La formule donnant  $\gamma(k)$  est la formule d'inversion de Fourier. Montrons que  $f \geq 0$ ; pour  $N > 0$ , définissons,

$$\begin{aligned} f_N(\lambda) &= \frac{1}{2\pi N} E \left| \sum_{l=1, N} X_l e^{-il\lambda} \right|^2 = \frac{1}{2\pi N} E \left( \sum_{l=1, N} X_l e^{-il\lambda} \sum_{m=1, N} X_m e^{im\lambda} \right) \\ &= \frac{1}{2\pi} \sum_{|h| < N} \left(1 - \frac{|h|}{N}\right) e^{ih\lambda} \gamma(h). \end{aligned}$$

$\gamma$  étant sommable,  $f_N(\lambda) \rightarrow f(\lambda)$  et  $f(\lambda) \geq 0$  comme limite d'une suite  $\geq 0$ . La covariance  $\gamma$  est la transformée de Fourier de  $f$  et la positivité de  $f$  correspond à la s.d.p. de  $\gamma$ .

•• Réciproquement, vérifions que  $\gamma$  définie par (29) est une acvf.  $\gamma(0) = \int_{-\pi}^{+\pi} f(\lambda)d\lambda \geq 0$  et  $\gamma$  est paire. Reste à voir que  $\gamma$  est s.d.p.. Pour  $n$  réels  $a_1, a_2, \dots, a_n$ , on a :

$$\begin{aligned} \sum_{u,v=1,n} a_u a_v \gamma(u-v) &= \int_{-\pi}^{+\pi} \sum_{u,v=1,n} a_u a_v e^{i(u-v)\lambda} f(\lambda) d\lambda \\ &= \int_{-\pi}^{+\pi} \left| \sum_{j=1,n} a_j e^{ij\lambda} \right|^2 f(\lambda) d\lambda \geq 0. \end{aligned}$$

..... ■

Un processus centré stationnaire de covariance sommable peut donc être *caractérisé au second ordre* soit par sa covariance  $\gamma$ , soit par sa densité spectrale  $f$ . Cette caractérisation est utile pour vérifier que  $\gamma$  est bien une covariance ; il suffit en effet de vérifier que la fonction  $f$  qui lui est associée est  $\geq 0$ . Par exemple, la fonction  $\kappa(\cdot)$  définie en (5) est une acvf ssi  $f(\lambda) = \frac{1}{2\pi}(1 - 2\rho \cos \lambda)$  est  $\geq 0$ , c-à-d ssi  $|\rho| \leq \frac{1}{2}$ .

### 5.1.1 Exercice

Les fonctions  $\gamma$  définies ci-dessous sont-elles des acvf ?

- (i) :  $\gamma(0) = 1, \gamma(h) = 0.5$  si  $|h| = 1, \gamma(h) = 0.2$  si  $|h| = 2$  et  $\gamma(h) = 0$  si  $|h| \geq 3$ .
- (ii) :  $\gamma(0) = 1, \gamma(h) = -0.5$  si  $|h| = \pm 2, \gamma(h) = -0.25$  si  $|h| = \pm 3$  et  $\gamma(h) = 0$  sinon.

## 5.2 Exemples de densité spectrale

- La densité spectrale d'un *BB* faible  $e \sim (0, \sigma^2)$  est constante :  $f(\lambda) \equiv \frac{\sigma^2}{2\pi}$ .
- La densité spectrale du *MA(1)*  $X_t = e_t + \theta e_{t-1}$  où  $e \sim BB(0, \sigma^2)$  est :

$$f(\lambda) = \frac{\sigma^2}{2\pi} \{ (1 + \theta^2) + 2\theta \cos \lambda \} = \frac{\sigma^2}{2\pi} |Q(e^{i\lambda})|^2.$$

- La densité spectrale de l'*AR(1)*  $X_t = \phi X_{t-1} + e_t, |\phi| < 1$ , vaut :

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi(1-\phi^2)} \left\{ 1 + \sum_{h \geq 1} \phi^h (e^{i\lambda h} + e^{-i\lambda h}) \right\} \\ &= \frac{\sigma^2}{2\pi} (1 - \phi \cos \lambda + \phi^2)^{-1} = \frac{\sigma^2}{2\pi |1 - \phi e^{i\lambda}|^2} = \frac{\sigma^2}{2\pi |P(e^{i\lambda})|^2}. \end{aligned}$$

## 5.3 Représentation spectrale d'un processus stationnaire

Le processus périodique (2) n'admet pas de densité spectrale car sa covariance n'est pas sommable. Il existe une notion plus générale pour de tels processus stationnaires, celle de *mesure spectrale*  $F$  sur  $] -\pi, \pi ]$  : pour (2),  $F(\lambda) = \sum_1^r \sigma_i^2 F_i(\lambda)$  où

$$F_i(\lambda) = \begin{cases} 0 & \text{si } \lambda < -\omega_i, \\ 0.5 & \text{si } -\omega_i \leq \lambda < \omega_i \text{ et} \\ 1 & \text{si } \omega_i \leq \lambda. \end{cases}$$

$F$  est symétrique, concentrée sur les fréquences  $\{-\omega_i, \omega_i, i = 1, r\}$  avec des sauts  $\sigma_i^2$  en ces  $r$ -fréquences. Formellement, un processus stationnaire au second ordre  $X$  admet la représentation spectrale :

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} Z(d\lambda), \quad (30)$$

où  $Z$  est un processus à accroissements décorrélés et  $Var(Z(d\lambda)) = F(d\lambda)$ . La densité spectrale  $f$  existe si  $F$  est absolument continue,  $F(d\lambda) = f(\lambda)d\lambda$ . Si  $X$  admet une densité spectrale  $f$ , un pic de  $f$  en  $\omega$  détecte une fréquence dominante  $\omega$  pour  $X$ .

### 5.3.1 Exercice et fonctions de R utiles

La série mensuelle  $X$  `sunspots` donne le nombre de taches solaires de 1749 à 1983. Représenter  $X$  et son acvf empirique pour  $0 \leq h \leq 50$ . En déduire une estimation  $\hat{f}$  de la densité spectrale de  $X$ . Vérifier que la fréquence  $\omega_{10} = \frac{2\pi}{10} \approx 0.63$  est un pic pour  $\hat{f}$ , suggérant une périodicité de 10 à 11 ans. Retrouver cette périodicité sur le graphe de  $X$ .

*Fonctions R utiles :*

Package `stats` : `spec` (estimation de  $f$ ), `spec.ar` (estimation de  $f$  et ajustement AR), `spec.pgram` (estimation de  $f$  par régularisation du périodogramme), `spec.taper` (estimation de  $f$  sur données rabotées).

## 5.4 Filtré linéaire d'un processus stationnaire

Soit  $X$  un processus stationnaire de covariance sommable et de densité spectrale  $f_X$ . Un filtré linéaire de  $X$  est un processus  $Y$  associé à  $X$  et à une suite sommable  $\psi$ ,

$$Y_t = \sum \psi_i X_{t-i} = \Psi(B)X_t. \quad (31)$$

**Propriété :**  $Y$  est stationnaire de densité spectrale  $f_Y(\lambda) = |\Psi(e^{i\lambda})|^2 f_X(\lambda)$ . En particulier, l'ARMA  $P(B)X_t = Q(B)e_t$  associé à  $e$ , un  $BB(0, \sigma^2)$ , admet pour densité spectrale :

$$f_X(\lambda) = \frac{\sigma^2 |Q(e^{i\lambda})|^2}{2\pi |P(e^{i\lambda})|^2}. \quad (32)$$

*Preuve :* L'acvf de  $Y$  est  $\gamma_Y(h) = \sum_{j,k \in \mathbb{Z}} \psi_j \psi_k \gamma_X(h+j-k)$ . La substitution des valeurs  $\gamma_X(h+j-k) = \int_{-\pi}^{+\pi} e^{i(h+k-j)\lambda} f_X(\lambda) d\lambda$  et quelques manipulations conduisent au résultat annoncé. ■

**Corollaire :** Tout ARMA( $p, q$ ) admet une représentation ARMA( $p, q$ ) causale et inversible de même acvf.

*Preuve :* La densité spectrale de l'ARMA s'écrit :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{Q}{P}(e^{i\lambda}) \right|^2$$

La nouvelle représentation doit faire disparaître les racines de  $P$  et de  $Q$  de module  $< 1$ . On fait la constatation suivante : si  $|\chi| \neq 1$  et  $|z| = 1$ , alors :

$$\begin{aligned} z - \chi &= z\chi(\chi^{-1} - z^{-1}), \quad |z - \chi| = |\chi| |z - \bar{\chi}^{-1}| \\ \text{c-à-d} &: \left| \chi \frac{z - \bar{\chi}^{-1}}{z - \chi} \right| = 1. \end{aligned}$$

Notons alors  $\{\xi_i, i \in I\}$  (resp.  $\{\zeta_j, j \in J\}$ ) les racines de  $Q$  (resp. de  $P$ ) de module  $< 1$  et considérons la fraction rationnelle associée :

$$R(z) = \prod_{i \in I} \xi_i \frac{z - \bar{\xi}_i^{-1}}{z - \xi_i} / \prod_{j \in J} \zeta_j \frac{z - \bar{\zeta}_j^{-1}}{z - \zeta_j}.$$

Alors  $|R(e^{i\lambda})| \equiv 1$  et  $X_t = R(B) \frac{Q}{P}(B) e_t$  est une représentation  $ARMA(p, q)$  causale et inversible de même densité spectrale, donc de même acvf, que l'ARMA initial. .... ■

## 6 Estimation d'un ARMA

Soit  $Y$  un modèle stationnaire de moyenne  $\mu = E(Y_1)$ ,  $X = Y - \mu$  étant un  $ARMA(p, q)$  à  $(p + q + 1)$  paramètres  $\phi, \theta$  et  $\sigma^2$ ,

$$X_t - (\phi_1 X_{t-1} + \dots + \phi_p X_{t-p}) = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \text{ où } \varepsilon \sim BB(0, \sigma^2). \quad (33)$$

Si on observe  $y_1, y_2, \dots, y_n$ , on commence par estimer  $\mu$  par  $\hat{\mu} = \bar{y}$  et on travaille ensuite sur les observations recentrées  $x_i = y_i - \bar{y}$ ,  $i = 1, n$ .

### 6.1 Estimation de Yule-Walker d'un $AR(p)$ .

Soit  $X$  un  $AR(p)$  causal :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t, \varepsilon \sim BB(0, \sigma^2)$$

$X$  étant causal,  $X_t = \sum_{i \geq 0} \psi_i \varepsilon_{t-i}$  avec  $\psi_0 = 1$  et  $Cov(X_s, \varepsilon_t) = 0$  si  $s < t$ . En particulier,  $Cov(X_t, \varepsilon_t) = \psi_0 \sigma^2 = \sigma^2$ . Multipliant l'équation génératrice de  $X_t$  par  $X_{t-k}$ ,  $0 \leq k \leq p$ , on obtient les  $(p + 1)$  équations de Yule-Walker :

$$\begin{aligned} \text{pour } k = 0 : \gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma^2 \text{ et} \\ \text{pour } 0 < k \leq p : \gamma(k) &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \dots + \phi_p \gamma(k-p). \end{aligned}$$

Notant  $\gamma_p = {}^t(\gamma(1), \gamma(2), \dots, \gamma(p))$  et  $\Gamma_p = (\gamma(i-j))_{i,j=1,p}$ , ces équations s'écrivent :

$$\Gamma_p \phi = \gamma_p \text{ et } \sigma^2 = \gamma(0) - {}^t \phi \gamma_p.$$

Les estimations de YW de  $\phi, \theta$  et  $\sigma^2$  sont des estimateurs de moments obtenus en identifiant les  $\gamma(k)$  aux distances  $0 \leq k \leq p$  à leurs estimations empiriques  $\hat{\gamma}(k)$ ,

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \text{ et } \hat{\sigma}^2 = \hat{\gamma}(0) - {}^t \hat{\phi} \hat{\gamma}_p.$$

En comparaison d'autres méthodes plus efficaces (i.e. le MV), l'estimation de YW est une estimation préliminaire qui pourra être utilisée comme valeur initiale démarrant un algorithme itératif d'optimisation (i.e. nécessaire à la recherche de l'estimateur du MV).

#### 6.1.1 Propriété de l'estimation de YW

**Propriétés (BD).**

1 - L'estimation  $\hat{\phi}$  de YW fournit toujours une représentation  $AR(p)$  causale :

$$\hat{P}(z) = 1 - \hat{\phi}_1 z - \dots - \hat{\phi}_p z^p \neq 0 \text{ si } |z| \leq 1.$$

2 - Les acvf empiriques  $\{\widehat{\gamma}(h), h = 0, p\}$  coïncident avec celles de l' $AR(p)$  au paramètre estimé  $(\widehat{\phi}, \widehat{\sigma}^2)$  :

$$\widehat{\gamma}(h) = \gamma((\widehat{\phi}, \widehat{\sigma}^2)) \text{ si } 0 \leq h \leq p.$$

3 - Si  $\varepsilon \sim i.i.d.(0, \sigma^2)$ , on a pour  $n \rightarrow \infty$  :

$$\widehat{\sigma}^2 \xrightarrow{P} \sigma^2 \text{ et } \sqrt{n}(\widehat{\phi} - \phi) \sim \mathcal{N}_p(0, \sigma^2 \Gamma_p^{-1}).$$

*Commentaires :*

- Une conséquence de la propriété (2-) est que pour toute covariance régulière  $\Gamma_{p+1} = (\gamma(i-j))_{i,j=1,p+1}$ , il existe un  $AR(p)$  dont les covariances aux distances  $k = 0, 1, \dots, p$  sont  $\gamma(0), \gamma(1), \dots, \gamma(p)$ .

- Cette propriété ne se transfère pas aux  $MA$ . Par exemple, pour  $\gamma(0) = 1$  et  $\gamma(1) = \gamma$  où  $|\gamma| < 1$ ,  $\Gamma = (\gamma(i-j))_{i,j=0,1}$  est bien d.p. mais il n'existe pas un  $MA(1)$  admettant ces covariances si  $|\gamma| > \frac{1}{2}$  (cf. l'équation (5)). Alors qu'il existe un  $AR(1)$  avec ces covariances à distance 0 et 1.

- La propriété (3-) permet de construire la région de confiance approchée au niveau  $(1 - \alpha)$  pour  $\phi$  :

$$\mathcal{R}_{1-\alpha} = \{\phi : {}^t(\widehat{\phi} - \phi)\widehat{\Gamma}_p^{-1}(\widehat{\phi} - \phi) \leq n^{-1}\widehat{\sigma}^2\chi^2(\alpha; p)\}, \text{ où } P(\chi_p^2 \leq \chi^2(1 - \alpha; p)) = 1 - \alpha.$$

- Si  $\widehat{v}_{jj} = (\widehat{\sigma}^2\widehat{\Gamma}_p^{-1})_{jj}$ , l'IC approché pour  $\phi_j$  au niveau 95% est  $[\widehat{\phi}_j \pm 1.96 \times \sqrt{\widehat{v}_{jj}/n}]$ .

### 6.1.2 Sélection de l'ordre $p$ d'un $AR$ .

- *A partir des auto-corrélations partielles*

Un  $AR(p)$  est caractérisé par le fait que sa pacf  $\alpha(m)$  vaut 0 si  $m \geq p + 1$ . L'estimation est  $\widehat{\alpha}(m) = \widehat{\phi}_{mm}$  où  $\widehat{\phi}_m = \widehat{\Gamma}_m^{-1}\widehat{\gamma}_m$ . De plus,  $\widehat{\alpha}(m) \sim \mathcal{N}(0, \frac{1}{n})$  si  $m \geq p + 1$ . Donc, si  $X \sim AR(p)$ , les  $\widehat{\phi}_{mm}$  pour  $m > p$  tombent dans l'intervalle  $[-\frac{1.96}{\sqrt{n}}, +\frac{1.96}{\sqrt{n}}]$  avec une probabilité de 95%.

- *A partir de la Final Prediction Error (FPE, Akaike, 1969).*

On suppose que  $X(n) = \{X_1, X_2, \dots, X_n\}$  est la réalisation d'un  $AR(p)$  et que  $Y$  est une réalisation de même loi mais indépendante de  $X$ . Soit  $\widehat{\Phi}_p, \widehat{\sigma}^2$  l'EMV de l' $AR(p)$  sur la base de  $X(n)$ . L'erreur de prédiction finale (FPE) de  $Y$  basée sur cet estimateur est :

$$FPE(p) = E(Y_{n+1} - \{\widehat{\phi}_1 Y_n + \widehat{\phi}_2 Y_{n-1} + \dots + \widehat{\phi}_p Y_{n+1-p}\})^2.$$

Dans un contexte de prévision, minimiser  $FPE(p)$  est raisonnable car on souhaite obtenir une erreur de prévision aussi petite que possible. Il y a deux origines à la FPE :

- l'une, due à l'ajustement sur une régression, est inhérente au bruit du modèle ;
- l'autre est due aux erreurs d'estimations  $\widehat{\Phi}_p$  faites sur  $\Phi_p$ .

Plus  $p$  est grand, meilleur est l'ajustement mais plus grande sera l'erreur d'estimation. FPE est un compromis entre ces deux objectifs antagonistes. Reste à estimer  $FPE(p)$ . Quelques manipulations (cf. exercice (6.1.3)) montrent que si  $X$  est gaussien, un bon estimateur de  $FPE(p)$  est :

$$FPE_p = \widehat{\sigma}_p^2 \times \frac{n+p}{n-p}. \quad (34)$$

Si  $p$  augmente,  $\widehat{\sigma}_p^2$  diminue mais dans le même temps  $\frac{n+p}{n-p}$  augmente. On choisira alors :

$$\widehat{p} = \arg \min_p FPE_p.$$

- *A partir du critère AIC de Akaike*

Une autre approche utilise le critère de Akaike ( $AIC$  pour Akaike Information Criterium) qui retient le modèle réalisant le minimum de :

$$AIC(p) = -2 \log L(\widehat{\phi}_p, \widehat{\sigma}_p^2) + 2(p+1).$$

$L(\widehat{\phi}_p, \widehat{\sigma}_p^2)$  est la vraisemblance calculée à l'estimation du MV  $(\widehat{\phi}_p, \widehat{\sigma}_p^2)$  pour l' $AR(p)$  et  $(p+1)$  est la dimension de ce modèle. Ainsi,  $2 \log L(\widehat{\phi}_p, \widehat{\sigma}_p^2)$  est une mesure de la *qualité d'ajustement* au modèle alors que  $2(p+1)$  est une *pénalisation par la dimension* de ce modèle. Là aussi,  $AIC(p)$  est un compromis entre deux objectifs antagonistes :  $AIC$  est un critère de parcimonie offrant un bon ajustement avec un modèle de taille raisonnable.

### 6.1.3 Exercice et fonction R utiles

*Obtention de l'estimation (34) de la FPE :* (a) - Montrer que :

$$E(Y_{n+1} - \{\widehat{\phi}_1 Y_n + \widehat{\phi}_2 Y_{n-1} + \dots + \widehat{\phi}_p Y_{n+1-p}\})^2 = \sigma^2 + E\{^t(\widehat{\Phi}_p - \Phi_p)\Gamma_p(\widehat{\Phi}_p - \Phi_p)\},$$

où  $\Phi_p = {}^t(\phi_1, \phi_2, \dots, \phi_p)$  et  $\Gamma_p = (E(Y_i Y_j))_{i,j=1}^p$ .

(b) - Sous l'hypothèse  $X \sim AR(p)$  et l'approximation  $\sqrt{n}(\widehat{\Phi}_p - \Phi_p) \sim \mathcal{N}_p(0, \Gamma_p)$ , en déduire que  $FPE(p) \approx \sigma^2(1 + \frac{p}{n})$ .

(c) - L' $AR(p)$  étant un modèle de régression linéaire, expliquer pourquoi, sous hypothèse gaussienne et pour  $n$  grand, l'EMV de  $\sigma^2$  vérifie :  $n \frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$ . En déduire qu'une bonne estimation de  $FPE(p)$  est donnée par (34).

*Fonctions R utiles :*

Package `stats` : `acf2AR` (AR s'ajustant exactement à une acvf), `ar` (par défaut, meilleur AR par AIC), `arima`

## 6.2 Estimation des moments pour un $ARMA(p, q)$

Soit  $X$  un  $ARMA(p, q)$  de paramètres  $(\phi, \theta, \sigma^2)$ . Les équations de YW s'écrivent (cf. (24)) :

$$\text{Pour } 0 \leq k \leq p : \gamma(k) - \sum_{i=1, p} \phi_i \gamma(k-1) = \begin{cases} \sigma^2 \sum_{i \geq 0} \theta_{k+i} \psi_j & \text{si } 0 \leq k \leq q, \\ 0 & \text{si } k \geq q+1. \end{cases}$$

Deux difficultés se présentent : d'une part, il faut commencer par expliciter les coefficients  $\psi_j(\phi)$  en  $\phi$  ; d'autre part ces équations ne sont plus linéaires dans les paramètres dès que  $q \geq 1$ .

Examinons l'estimation des moments dans le cas d'un  $MA(1)$ ,  $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$  où  $\varepsilon \sim BB(0, \sigma^2)$ . Egalant covariances théoriques et empiriques à distance 0 et 1, on obtient :

$$\widehat{\gamma}(0) = \widehat{\sigma}^2(1 + \widehat{\theta}^2) \text{ et } \widehat{\rho}(1) = \frac{\widehat{\theta}}{1 + \widehat{\theta}^2}.$$

Si  $|\widehat{\rho}(1)| > \frac{1}{2}$ , il n'y a pas de solution réelle (on sait qu'il n'y a pas de  $MA(1)$  avec  $|\rho(1)| > \frac{1}{2}$ ). Sinon,  $(\widehat{\sigma}^2, \widehat{\theta})$  s'obtient en résolvant ces deux équations, non-linéaires. La méthode se généralise aux modèles  $MA(q)$ .

## 6.3 Estimation du maximum de vraisemblance

### 6.3.1 Le cas gaussien

Soit  $X(n) = (X_1, X_2, \dots, X_n)$  un *vecteur gaussien* de moyenne 0 et de covariance  $\Gamma_n = Cov(X(n))$ . La *vraisemblance* (la densité de probabilité)  $L_n(X(n))$  de  $X(n)$  est (cf. § 10.1) :

$$L_n(X(n)) = L_n(\Gamma_n) = (2\pi)^{-\frac{n}{2}} (\det \Gamma_n)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \{ {}^t X(n) \Gamma_n^{-1} X(n) \}\right\}. \quad (35)$$

Si la moyenne suit un modèle paramétrique :  $E(X_t) = m_t(\beta), t \in \mathbb{Z}$ , pour  $\beta = {}^t(\beta_1, \beta_2, \dots, \beta_r)$  inconnu et des fonctions  $(m_t)$  connues, on obtiendra la vraisemblance en remplaçant dans (35)  $X(n)$  par  $X^*(n) = X(n) - m(n, \beta)$  où  $m(n, \beta) = {}^t(m_t(\beta), t = 1, n)$ .

Si  $\Gamma_n$  s'exprime à partir de paramètres  $\alpha = {}^t(\alpha_1, \alpha_2, \dots, \alpha_m)$ ,  $L_n$  s'explicitera en  $\theta = (\alpha, \beta)$  et l'estimateur du maximum de vraisemblance (EMV) est une valeur  $\widehat{\xi} = (\widehat{\alpha}, \widehat{\beta})$  maximisant  $L_n(\alpha, \beta)$ . Sous de bonnes conditions générales, cette méthode est asymptotiquement efficace et  $\widehat{\xi} \sim \mathcal{N}_s(\xi, \frac{V(\xi)}{n})$ .

### 6.3.2 Pseudo-vraisemblance gaussienne pour un ARMA

Soit  $X$  un  $ARMA(p, q)$  à bruit générateur *i.i.d.*  $(0, \sigma^2)$  :

$$X_t - (\phi_1 X_{t-1} + \dots + \phi_p X_{t-p}) = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \text{ où } \varepsilon \sim i.i.d.(0, \sigma^2).$$

Notons  $X(n) = (X_1, X_2, \dots, X_n)$  l'observation de  $X$ . Que  $X$  soit gaussien ou non, la méthode précédente basée sur la fonctionnelle "gaussienne"  $L_n$  donnée en (35) peut encore être utilisée : pour des paramètres  $\varphi = (\phi, \theta)$  associés à une écriture causale et inversible et sans facteur commun, une valeur  $(\widehat{\phi}, \widehat{\theta}, \widehat{\sigma}^2)$  maximisant  $L_n$  est une estimation du *maximum de la pseudo-vraisemblance* (EMPV) *gaussienne*. Asymptotiquement (BD) :

$$\sqrt{n}(\widehat{\varphi} - \varphi) \xrightarrow{loi} \mathcal{N}_{p+q}(0, V(\varphi)).$$

*Identification de la covariance asymptotique*  $V(\varphi)$  (BD) : on a, pour  $p > 1$  et  $q > 1$ ,

$$\sigma^2 V(\phi, \theta)^{-1} = \begin{pmatrix} Var_{\phi}(U) & Cov_{\phi, \theta}(U, V) \\ Cov_{\phi, \theta}(U, V) & Var_{\theta}(V) \end{pmatrix}$$

où  $U = {}^t(U_1, U_2, \dots, U_p)$  suit le modèle  $AR(p)$  :  $U_t = \phi_1 U_{t-1} + \dots + \phi_p U_{t-p} + \varepsilon_t$  et  $V = {}^t(V_1, V_2, \dots, V_q)$  l' $AR(q)$  :  $V_t = \theta_1 V_{t-1} + \dots + \theta_q V_{t-q} + \varepsilon_t$ . Pour  $p = 0$ ,  $\sigma^2 V(\theta)^{-1} = Var_{\theta}(V)$ ; pour  $q = 0$ ,  $\sigma^2 V(\phi)^{-1} = Var_{\phi}(U)$ . On retrouve la variance asymptotique des équations de YW.

Une propriété remarquable de l'estimation par PMV pour un  $ARMA$  (BD) est que la loi asymptotique de  $\widehat{\varphi}$  est la même que le bruit  $\varepsilon$  soit gaussien (MV) ou non (PMV). Ceci étant, il est probable que l'EMV est plus efficace que l'EPMV si  $\varepsilon$  n'est pas gaussien.

#### Exemples

1 -  $X$  est un  $AR(p)$  :  $\sigma^2 V(\varphi)^{-1} = Var_{\phi}(U)$  :

- pour un  $AR(1)$  :  $V(\phi) = 1 - \phi_1^2$ .
- pour un  $AR(2)$  :

$$V(\phi) = \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix}.$$

2 -  $X$  est le  $MA(q)$  :  $\sigma^2 V(\varphi)^{-1} = Var_{\theta}(V)$ .

- pour un  $MA(1)$  :  $V(\theta) = 1 - \theta_1^2$ ;
- pour un  $MA(2)$  :

$$V(\theta) = \begin{pmatrix} 1 - \theta_2^2 & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & 1 - \theta_2^2 \end{pmatrix}.$$

3 -  $X$  est un l' $ARMA(1, 1)$ ,  $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$  :

$$V(\phi) = \begin{pmatrix} (1 - \phi^2)^{-1} & (1 + \phi\theta)^{-1} \\ (1 + \phi\theta)^{-1} & (1 - \theta^2)^{-1} \end{pmatrix}.$$

### 6.3.3 Choix de modèle : le critère de Akaike.

Avant de procéder à l'estimation d'un  $ARMA(p, q)$  et sans information a priori sur  $p$  et  $q$ , il faut commencer par *identifier* ces ordres  $p$  et  $q$ . Le critère de Akaike, modifié par Brockwell et Davis, propose le choix de  $(p, q)$  et des EMV des paramètres associés  $((\phi_p, \theta_q), \sigma_{p,q}^2)$  qui minimise le critère  $AIC$  suivant :

$$AIC(p, q; (\phi_p, \theta_q), \sigma_{p,q}^2) = -2 \ln L_n(p, q; (\phi_p, \theta_q), \sigma_{p,q}^2) + 2(p + q + 1) \frac{n}{n - p - q - 2} \quad (36)$$

Pour chaque  $p, q$  fixé, l'EMV fait le choix optimal de  $(\phi_p, \theta_q), \sigma_{p,q}^2$ . On choisit ensuite le couple  $(p, q)$  qui minimise  $AIC$ . C'est un bon compromis entre deux objectifs antagonistes :  $-2 \ln L_n$  qui mesure la qualité de l'ajustement et  $2(p + q + 1) \frac{n}{n - p - q - 2}$  qui pénalise une dimension paramétrique trop importante. D'autres pénalisations existent, telle celle du critère BIC (Bayesian Information Criterion) qui pénalise la dimension par  $2(p + q + 1) \times \log n$ . Le principe du type  $AIC$  est applicable à tout modèle paramétrique.

### 6.3.4 Validation d'un modèle

Pour valider une modélisation  $ARMA(p, q)$  de paramètres  $\psi = (\phi, \theta; \sigma^2)$  sur la base d'observations  $x_1, x_2, \dots, x_n$ , on procède ainsi :

- 1 - On estime  $\hat{\psi}$  par le MV (ou par le MPV, ou toute autre méthode convergente) ;
- 2 - Pour  $t = p + 1, n - 1$  et le modèle  $ARMA(\hat{\psi})$ , on prédit  $X_{t+1}$  à partir de  $X_1, X_2, \dots, X_t$  par  $\hat{X}_{t+1}$  d'EQM  $E(X_{t+1} - \hat{X}_{t+1})^2 = r_t \sigma^2$ . Si le modèle est valide et l'ARMA inversible, alors, on a si  $n \rightarrow \infty$  (BD) :

$$E([X_{t+1} - \hat{X}_{t+1} - \varepsilon_{t+1}]^2) \rightarrow 0 \text{ et } r_t \rightarrow 1.$$

- 3 - On forme les résidus réduits  $\hat{R}_t = \frac{X_{t+1} - \hat{X}_{t+1}}{\sigma}$ ,  $t = 1, n$  et on trace le nuage  $\{(t, \hat{R}_t), t = 1, n\}$ .
- 4 - Si le modèle est valide, ce nuage est "rectangulaire" et les résidus  $\{\hat{R}_t, t = 1, n\}$  se comportent comme :

- un  $BB(0, 1)$  si  $\varepsilon \sim BB(0, \sigma^2)$  ;
- un  $BB \text{ i.i.d.}(0, 1)$  si  $\varepsilon$  est un  $BB \text{ i.i.d.}(0, \sigma^2)$  ;
- un  $BB \text{ i.i.d. } \mathcal{N}(0, 1)$  si  $\varepsilon$  est un  $BB \text{ i.i.d. } \mathcal{N}(0, \sigma^2)$ .

On utilise alors l'un des tests de  $BB \text{ i.i.d.}$  définis en (1.6) (test du portemanteau, de Ljung-Box, de McLeod-Li, du point tournant, différence de signe). En complément, le QQplot ou le test Jarque-Bera permet de juger de la normalité des observations.

### 6.3.5 Exercices et fonctions R utiles

1 - On étudie la série `data(lynx)` donnant la population des lynx au Canada 114 années successives (cycle lynx/lapin).

(a) - Représenter la série, ses acvf et pacf empiriques. Quel modèle  $AR(p)$  proposeriez-vous ? Estimer ce modèle. Quelles sont les racines du polynôme  $AR$  associé (fonction `polyroot`).

(b) - *Validation 1* : vous basant sur ce modèle, estimer les 5 dernières valeurs de la série et juger de la qualité prédictive de ce modèle.

(c) - *Validation 2* : simuler 114 observations de l' $AR(p)$  estimé. Quelle conclusion en tirez-vous ?

(d) - *Validation 3* : tester que le résidu du modèle estimé est un  $BB$ .

2 - Une sous-séquence de taille 100 des données `sunspots` présente les acvf empiriques suivantes :  $\hat{\gamma}(0) = 1382.2$ ,  $\hat{\gamma}(1) = 1114.4$ ,  $\hat{\gamma}(2) = 591.73$  et  $\hat{\gamma}(3) = 96.22$ . En déduire les estimations de  $YW$  de  $\phi_1, \phi_2$  et  $\sigma^2$  du modèle  $AR(2)$ ,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \text{ où } \{Z_t\} \sim BB(0, \sigma^2)$$

où  $Y_t = X_t - 46.93, t = 1, 100$ , est la série corrigée de sa moyenne. Si les données sont la réalisation d'un  $AR(2)$ , déterminer un IC pour  $\phi_1$  et  $\phi_2$ .

3 - Estimation d'un  $AR(2)$  contraint :  $X$  est l' $AR(2)$ ,

$$X_t = \phi X_{t-1} + \phi^2 X_{t-2} + e_t \text{ avec } e \sim BB(0, \sigma^2).$$

(a) - Pour quelles valeurs de  $\phi$  cette représentation est elle stationnaire et causale ?

(b) - Pour un 200-échantillon, on observe  $\hat{\gamma}(0) = 6.06$  et  $\hat{\rho}(1) = 0.687$ . Ecrire les 2 premières équations de YW. En déduire un estimateur de  $\phi$  et  $\sigma^2$ .

Fonctions R utiles :

Package stats : qqplot, Box.test (tests de Box-Pierce et Ljung-Box)

Package tseries : jarque.bera.test

Package fArma : armaFit

Package forecast : auto.arima (meilleur ARIMA avec AIC)

## 7 Régression à résidus corrélés, ARIMA et SARIMA

Nous avons présenté au §2 des modèles  $X$  non-stationnaires, soit parce que leur moyenne  $E(X_t)$  n'est pas constante, soit parce que la variance  $Var(X_t)$  n'est pas constante. Nous supposons ici que  $X$  n'est pas stationnaire en moyenne mais qu'il l'est en variance-covariance, c-à-d que,

$$X_t = m(t) + e_t, t \geq 1,$$

où  $e$  est un bruit centré et stationnaire au second ordre, d'acvf  $\gamma(\cdot)$ . Deux techniques d'étude se présentent :

- (i) la modélisation par régression de  $t \mapsto m(t)$  ;
- (ii) la différenciation de  $X$ .

### 7.1 Régression à résidu ARMA

On modélise  $t \mapsto m(t)$  à partir de variables explicatives : courbe de croissance  $g(t)$  fonction du temps  $t$  ( $t, t^2, \log t, \dots$ ), saisonnières périodiques de période connue  $d$  ( $\forall t : s_t \equiv s_{t+d}$ ), variables exogènes  $u_t \in \mathbb{R}^k$  quantitatives, exogènes  $y_t \in \{1, 2, \dots, I\}$  qualitatives. Une forme très générale pour  $m(\cdot)$  est :

$$m(t) = P_r(t) + s_t + \alpha(y_t) + {}^t\beta u = {}^t\delta z_t, \quad (37)$$

où  $P_r(t) = a_0 + a_1 t + \dots + a_r t^r$ . On devra s'assurer de l'identifiabilité paramétrique du modèle : pour (37), il est nécessaire d'imposer que les saisonnalités ( $s_t$ ) soient centrées ( $s_1 + s_2 + \dots + s_d = 0$ ) ainsi que les paramètres  $\alpha$  ( $\alpha_1 + \alpha_2 + \dots + \alpha_I = 0$ ) ; la dimension du modèle est alors  $r + d + I + k - 1$ .

Supposons que l'on observe  $X(n) = \{X_1, X_2, \dots, X_n\}$  et qu'on note matriciellement (37) :

$$X(n) = Z(n)\delta + e(n), \text{ où } E(e(n)) = 0 \text{ et } Var(e(n)) = \Sigma_n.$$

La démarche d'estimation est la suivante :

1. On estime  $\delta$  par moindres carrés ordinaires (MCO) (cf. §10.3) : sous des conditions assez générales,  $\hat{\delta} \rightarrow \delta$ .
2. On en déduit l'estimation  $\hat{e}_t = X_t - {}^t\hat{\delta} z_t, t = 1, n$  des MCO des résidus.
3. On estime le modèle (i.e. ARMA) des résidus  $e$  sur la base des pseudo-observations  $\hat{e}(n) = (\hat{e}_t, t = 1, n)$ .
4. On calcule  $\hat{\Sigma}_n = Cov(\hat{e}(n))$  (ou on estime empiriquement  $\Sigma_n$  sans modéliser  $e$ ).

5. On réestime  $\delta$  par moindres carrés généralisés (MCG ; cf. §10.3) pour la covariance  $\widehat{\Sigma}_n$ .

Notant  $Z = Z(n)$ , les estimations des MCO et des MCG s'écrivent respectivement :

$$\widehat{\delta} = ({}^tZZ)^{-1} {}^tZX(n) \text{ et } Var(\widehat{\delta}) = ({}^tZZ)^{-1} {}^tZ\Sigma_nZ ({}^tZZ)^{-1}, \text{ et} \quad (38)$$

$$\widetilde{\delta} = ({}^tZ\widehat{\Sigma}_n^{-1}Z)^{-1} {}^tZ\widehat{\Sigma}_n^{-1}X(n) \text{ et } Var(\widetilde{\delta}) = ({}^tZ\widehat{\Sigma}_n^{-1}Z)^{-1}. \quad (39)$$

### Commentaires.

- L'estimation  $\widehat{\delta}$  des MCO est sans biais. Elle est convergente si  $e$  est un *ARMA* et si la matrice du dispositif  $Z(n)$  vérifie  ${}^tZ(n)Z(n) \rightarrow \infty$  (tous ses termes tendent vers l'infini).
- *Attention* :  $Var(\widehat{\delta}) \neq \sigma_e^2({}^tZZ)^{-1}$  et les tests habituels de Fisher sur  $\delta$  ne sont pas valables.
- Sous hypothèse gaussienne, l'estimation des MCG de  $\delta$  est asymptotiquement équivalente à celle du MV.
- Les tests de Fisher pour la norme de la covariance inverse,  $\|u\|_{\widehat{\Sigma}_n^{-1}}$ , sont approximativement valides.
- On peut itérer l'étape (3) en utilisant l'estimation  $\widetilde{e}$  des MCG des résidus.

### 7.1.1 Exercices et fonctions R utiles

1. On modélise les données **LakeHuron** par une tendance affine avec résidu *AR*(1),

$$X_t = a + bt + e_t \text{ où } e \sim AR(1).$$

- (a) - Déterminer l'EMCO de  $(a, b)$  et leurs variances-covariances. Tester  $b = 0$  contre  $b < 0$ .
- (b) - Déterminer l'EMCG de  $(a, b)$ , leurs variances-covariances et estimer le modèle des résidus. Tester  $b = 0$ .

2. La moyenne d'une série  $X_t = m(t) + e_t$  change à un instant connu  $1 < t_0 < n$  suivant,

$$m(t) = m_0 \text{ si } t \leq t_0 \text{ et } m(t) = m_1 \text{ si } t > t_0.$$

- (a) - Tester  $m_0 = m_1$  contre  $m_0 \neq m_1$  si  $e \sim i.i.d. \mathcal{N}(0, \sigma^2)$ .
  - (b) - Tester  $m_0 = m_1$  si  $e \sim AR(1)$  (on supposera  $t_0$  et  $n - t_0$  grands).
3. La série **Seat Belt Law** (cf. §10.5.3 et BD (2002)) donne le nombre de décès  $Y_t$  par accident les mois  $t$  sur les routes d'Angleterre de janvier 1975 à décembre 1984 ( $n = 120$ ). Le port de la ceinture a été rendu obligatoire (Seat Belt Law) en janvier 1983 ( $t = 99$ ). On veut savoir si cette mesure a été efficace. Pour cela, on propose le modèle de régression avec une rupture en  $t = 99$  :

$$Y_t = a + b \times \mathbf{1}(t \geq 99) + e_t, t = 1, 120,$$

où  $e$  est un résidu stationnaire. On veut donc tester :  $(H_0) : b = 0$  contre  $(H_1) : b < 0$ .

- (a) - Constater la saisonnalité mensuelle  $s_t$ . Quel est le "pic", le "creux" des accidents ?
- (b) - Pour faire disparaître  $s_t$ , on considère  $X_t = Y_t - Y_{t-12}$ . Montrer que  $X_t = b \times g(t) + u_t$ ,  $t = 13, 120$ , où  $g(t) = \mathbf{1}(99 \leq t \leq 120)$  est la nouvelle exogène.
  - (b-1) - Estimer  $b$  et les  $(u_t)$  par MCO. Identifier et estimer le modèle des résidus.
  - (b-2) - Estimer  $b$  par MCG. Comparer à l'estimation des MCO. Tester  $(H_0)$  contre  $(H_1)$ . Vérifier qu'une itération supplémentaire des étapes 2-a-b de la démarche d'estimation itérative modifie peu les résultats.

### Fonctions R utiles :

Package **stats** : **lm** (ajustement sur un modèle linéaire), **nls** (non linear Least Squares, régression non-linéaire par MCP), **predict.lm** (prédiction à partir d'un ML).

## 7.2 ARIMA et SARIMA

On a vu au §2.1.4 pourquoi la différenciation  $\nabla X_t = X_t - X_{t-1}$  ou son itérée  $\nabla^k X_t$  absorbe les tendances affines, polynomiales ou localement affines et à variation lente. D'autre part, si  $X_t = m(t) + e_t$  est à résidu stationnaire,  $\nabla X_t = \nabla m(t) + \varepsilon_t$  où  $\varepsilon_t = \nabla e_t$  est aussi stationnaire. Pour garder une structure de covariance simple sur  $\varepsilon$ , on réduira autant que faire se peut les opérations de différenciation. De même, une différenciation à l'ordre  $d$ ,  $\nabla_d X_t = X_t - X_{t-d}$  absorbe les  $d$ -saisonnalités.

### 7.2.1 ARIMA( $p, k, q$ )

C'est un modèle où  $X$  satisfait :

$$\begin{aligned} \nabla^k X_t &\sim ARMA(p, q), \text{ c-à-d :} \\ P(B)(1 - B)^k X_t &= Q(B)e_t, \text{ où } e \sim BB(0, \sigma^2), \end{aligned}$$

$P$  et  $Q$  étant des polynômes de l' $ARMA(p, q)$ . Les paramètres de l' $ARIMA$  sont d'une par les entiers  $p, k$  et  $q$ , d'autre part les paramètres réels  $\Phi$  et  $\Theta$  de l' $ARMA(p, q)$ .

### 7.2.2 Exercices

1. Simuler l' $ARIMA(1, 1, 0)$  de taille  $n = 200$  :  $(1 - B)(1 - 0.8B)X_t = e_t$ ,  $e \sim BB(\mathcal{N}(0, 1))$ . Représenter la trajectoire de  $X$  ainsi que celle de  $Y = \nabla X$ .
2. Déterminer l'acvf et la pacf empiriques de  $X$ . Ajuster un modèle  $AR(2)$  non contraint sur ces observations.
3. Ajuster un modèle  $AR(1)$  sur  $Y_t, t = 2, 200$ .
4. En vous basant sur les  $\{X_t, t \leq 198\}$ , déterminer les prévisions  $\hat{X}_{198,1}$  et  $\hat{X}_{198,2}$  de  $X_{199}$  et  $X_{200}$  pour les modèles ajustés (2) et (3). Répétant 100 fois de façon i.i.d. la simulation de  $\{X_t, t = 1, 200\}$ , évaluer les erreurs des MC pour ces prévisions ainsi que les biais et écarts types des estimations des paramètres de modèles (2) et (3).

### 7.2.3 Test de racine unitaire de Dickey-Fuller

L'exercice précédent donne deux façons de modéliser la série  $X$  : l'une utilise un  $AR(2)$  non-contraint et l'autre un  $AR(2)$  avec la contrainte qu'une racine du polynôme  $AR$  vaut 1. Dickey et Fuller (DF) ont étudié le problème du test de racine unitaire pour les modèles  $ARMA$ , la racine unitaire portant sur la partie  $AR$  ou sur la partie  $MA$  (BD). Tester qu'une représentation  $ARMA$  présente une racine unitaire (ou proche de l'unité) est importante car si une telle racine existe, cela suggère de différencier la série  $X$ . Nous résumons ici la démarche et le test de DF de racine unitaire pour la partie  $AR$ .

Afin de comprendre pourquoi ce test a un comportement non-standard (vitesse d'estimation en variance  $\neq n$ , loi limite non-gaussienne), examinons la situation "frontière" de l' $AR(1)$  causal :  $X_t = \phi X_{t-1} + e_t$ ,  $e \sim i.i.d.(0, \sigma^2)$  lorsque  $0 < \phi < 1$  mais  $\phi \uparrow 1$ . Si on observe  $X_1, X_2, \dots, X_n$ , on sait (cf. §6.1.1) que l'estimation  $\hat{\phi}_n = \hat{\rho}_n(1) = \frac{\hat{\gamma}_n(1)}{\hat{\gamma}_n(0)}$  de  $\phi$  vérifie :

$$\sqrt{n}(\hat{\phi}_n - \phi) \sim \mathcal{N}(0, 1 - \phi^2).$$

Ainsi, si  $\phi \uparrow 1$  et  $n \rightarrow \infty$ ,  $Var(\hat{\phi}_n) \simeq \frac{1-\phi^2}{n} = o(\frac{1}{n})$ . Plus précisément, on peut démontrer que si

$\phi = 1$  :

$$n(\hat{\phi}_n - 1) \xrightarrow{loi} \nu \sim \frac{1}{2} \frac{W(1)^2}{\int_0^1 W(u)du},$$

où  $W$  est le Brownien sur  $\mathbb{R}$ . En particulier,  $\nu$  est de variance finie et  $Var(\widehat{\phi}_n - 1) \propto \frac{1}{n^2}$ .

**Le test de DF.** : Supposons que  $X \sim AR(p)$  de moyenne  $\mu$ ,

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + e_t, e \sim i.i.d.(0, \sigma^2). \quad (40)$$

Quelques manipulations montrent que  $X$  s'écrit encore :

$$\nabla X_t = \phi_0^* + \phi_1^* X_{t-1} + \phi_2^* \nabla X_{t-1} + \cdots + \phi_p^* \nabla X_{t-p+1} + e_t, e \sim i.i.d.(0, \sigma^2). \quad (41)$$

où  $\phi_0^* = \mu(1 - \phi_1 - \cdots - \phi_p)$ ,  $\phi_1^* = -1 + \sum_{i=1}^p \phi_i$  et  $\phi_j^* = -\sum_{i=j}^p \phi_i$  pour  $j = 2, p$ .

Si 1 est racine du polynôme  $AR$ ,  $P(1) = 1 - \sum_{i=1}^p \phi_i = \phi_1^* = 0$  et  $\nabla X$  est un  $AR(p-1)$ . Donc tester l'existence d'une racine unitaire pour le modèle (40) revient à tester  $\phi_1^* = 0$  dans la régression (41). La procédure de DF est la suivante :

1. Effectuer l'EMCO de  $\nabla X_t$  sur  $1, X_{t-1}, \nabla X_{t-1}, \cdots, \nabla X_{t-p+1}$  dans le modèle (41).
2. Calculer  $\widehat{\sigma}^2(\widehat{\phi}_1^*)$  comme pour une EMCO.
3. Former la statistique :  $DF = \frac{\widehat{\phi}_1^*}{\widehat{\sigma}(\widehat{\phi}_1^*)}$ .

DF ont étudié la loi limite de  $DF$  sous l'hypothèse de racine unitaire : pour  $n$  grand, la région de rejet  $\mathcal{R}_\alpha$  au niveau  $\alpha$  du test de racine unitaire est

$$\mathcal{R}_\alpha = \{DF < s_\alpha\},$$

de seuils  $s_\alpha = -3.43$  (resp.  $-2.86$  et  $-2.57$ ) au niveau 1% (resp. 5% et 10%).

#### 7.2.4 Exercices et fonctions de R utiles

1. Tester l'existence d'une racine unitaire pour la série simulée de l'exercice (7.2.2).
2. Modélisant la série *LakeHuron* par un  $AR(2)$ , tester l'existence d'une racine unitaire pour ce modèle. Utilisant un critère de minimum de l'erreur de prévision sur les 10 dernières années observées, lequel des trois modèles suivants préférer :
  - (a) un  $AR(2)$  stationnaire ;
  - (b) un  $ARIMA(1, 1, 0)$  ;
  - (c) un  $AR(1)$  pour les résidus de la régression affine sur les niveaux.
3. Etude de la série journalière  $X$  des  $n = 78$  "*Dow-Jones Utilitie Index*" (cf. §10.5.3).
  - (a) - Constater la non-stationnarité de  $X$ . Tester l'existence d'une racine unitaire AR.
  - (b) - Soit  $Y_t = X_t - X_{t-1}$ . Ajuster un modèle  $ARMA$  à la série différenciée.
  - (c) Valider votre modèle et prévoir les 10 dernières observations.

*Fonctions R utiles :*

Package `tseries` : `adf.test` (test de Dickey-Fuller)

#### 7.2.5 Prévision pour un $ARIMA$

Soit  $X$  un  $ARIMA(p, k, q)$ ,  $(X_1, X_2, \cdots, X_n)$  son observation :  $Y_t = \nabla^k X_t$  est un  $ARMA(p, q)$  et la prédiction  $\widehat{Y}_{n,1}$  s'obtient sur la base des "observations"  $(Y_{k+1}, Y_{k+2}, \cdots, Y_n)$ . Puisque  $X_t = Y_t - \sum_{j=1, k}^k \binom{k}{j} (-1)^j X_{t-j}$ , on obtient :

$$\widehat{X}_{n,1} = \widehat{Y}_{n,1} - \sum_{j=1, k}^k \binom{k}{j} (-1)^j X_{n-j}$$

Les prévisions à pas  $h$ ,  $\widehat{X}_{n,h}$  s'obtiendront de façon récursive à partir des prévisions  $\widehat{Y}_{n,k}$ ,  $k = 1, h$ .

### 7.3 Modèle saisonnier SARIMA

Si il y a une  $d$ -saisonnalité, on peut proposer un modèle  $ARIMA(p, k, q; p', k', q')$  à structure saisonnière :

$$\begin{aligned} ((\nabla_d)^{k'} \times \nabla^k) X_t &= Y_t \sim ARMA(p, q), \text{ c-à-d :} \\ P(B)\mathcal{P}(B^d)(1-B)^k(1-B^d)^{k'} X_t &= Q(B)\mathcal{Q}(B^d)e_t \text{ où } e \sim BB(0, \sigma^2), \end{aligned}$$

où les degrés de  $P$ ,  $Q$ ,  $\mathcal{P}$  et  $\mathcal{Q}$  sont respectivement  $p, q, p'$  et  $q'$ . On remarquera que  $P(B)\mathcal{P}(B^d)$  est un polynôme en  $B$  avec un certain nombre de coefficients nuls, ainsi que  $Q(B)\mathcal{Q}(B^d)$ . Donc  $(Y_t)$  peut être vu comme un  $ARMA$  contraint, la  $d$ -saisonnalité structurant ces contraintes. Par exemple, pour une saisonnalité trimestrielle ( $d = 4$ ) et  $X \sim SARIMA(0, 1, 1; 0, 1, 1)$ , le différencié  $Y_t = \nabla_4 \nabla^1 X_t$  est un  $MA(13)$  contraint :

$$\begin{aligned} Y_t - m &= (1 + \theta_1 B)(1 + \theta_4 B^4) e_t \\ &= e_t + \theta_1 e_{t-1} + \theta_4 e_{t-4} + \theta_1 \theta_4 e_{t-5} \text{ où } e \sim BB(0, \sigma^2). \end{aligned}$$

L'estimation du MV devra tenir compte de ces contraintes. Comme pour  $ARIMA$ , les prévisions sur  $X$  se feront à partir des prévisions  $ARMA$  de  $Y$ .

#### 7.3.1 Exercices

1 - Une série  $(X_t)$  différenciée à l'ordre 12 puis à l'ordre 1 conduit à la série  $(Y_t)$  centrée dont l'acvf empirique satisfait :

$$\begin{aligned} \hat{\rho}(12j) &\approx (.8)^j, j = 0, 1, 2, \dots \\ \hat{\rho}(12j \pm 1) &\approx 0.4 \times (.8)^j, j = 0, 1, 2, \dots \\ \text{et } \hat{\rho}(h) &\approx 0 \text{ sinon.} \end{aligned}$$

Proposer une modélisation  $SARIMA$  et donner une estimation de ses paramètres. Pour  $n$  grand, déterminer les prédictions  $\hat{X}_{n,1}$  et  $\hat{X}_{n,2}$  et leurs erreurs de carrés moyens.

2 - (a) - On considère  $(M1) : X \sim SARIMA(0, 1, 1; 1, 1, 0)$  et  $(M2) : X \sim SARIMA(0, 1, 1; 0, 1, 1)$ . Préciser ces modèles en termes de  $Y_t = \nabla \nabla_{12} X_t$  (paramètres, bruit générateur).

(b) - On considère également :  $(M3) : Y \sim ARMA(12, 1)$  et  $(M4) : Y_t = e_t + \theta_1 e_{t-1} + \theta_{12} e_{t-12} + \theta_{13} e_{t-13}$  où  $e \sim BB(0, \sigma^2)$ . Quelles sont les dimensions paramétriques des 4 modèles ? Quelles inclusions existent entre eux ? Quelles sont les covariances non nulles pour  $(M4)$  ?

(c) - Obtenant :  $\log L(M_1) = -117.5$ ,  $\log L(M_2) = -118.4$ ,  $\log L(M_3) = -112.5$  et  $\log L(M_4) = -117.2$ , quel modèle retenir ?

3 - On cherche à modéliser  $X$ , la série mensuelle **USAccDeaths** donnant le nombre de morts accidentelles aux Etats Unis. Pour éliminer tendance et saisonnalité, on travaille sur la série différenciée  $Y_t = \nabla \nabla_{12} X_t$ . On va ajuster 4 modèles (a), (b), (c) et (d).

(a) -  $Y \sim BB$  : faire la prévision des 10 dernières observations et noter l'AIC.

(b) - Observant l'acvf, expliquer le choix d'un  $MA(12)$  pour  $Y$ . Estimer le modèle, prévision des 10 dernières observations et calculer l'AIC.

(c) - Dans le modèle précédent, identifier les paramètres  $\theta_j$  non-significatifs. Estimer le modèle contraint associé. Donner la prévision des 10 dernières observations et calculer l'AIC.

(d) - Ajuster le  $SARIMA(0, 1, 1; 0, 1, 1) : Y_t - m = (1 - \theta_1 B)(1 - \theta_{12} B^{12}) e_t$ . Estimer le modèle. Prévision des 10 dernières observations et calculer l'AIC. Quel modèle retenir ?

4 - Une série temporelle  $X = (X_t, t = 1, 2, \dots)$  suit le modèle de régression,  $X_t = m(t) + e_t$  où  $m(t) = a + bt + s_t$ ,  $(s_t)$  est une saisonnalité trimestrielle et  $e$  est un résidu stationnaire centré de variance  $\sigma_e^2$ . On observe  $X_1, X_2, \dots, X_{196}$ .

(4-a) - Estimant  $e$  par MCO, on trouve pour acvf et pacf empiriques de  $\hat{e}$  :

$$\hat{\rho}(1) = 0.89; \hat{\rho}(2) = 0.60; \hat{\alpha}(2) = -0.12.$$

(a1) Quel modèle retenir pour  $e$  entre un  $MA(1)$  et un  $AR(1)$  ?

(a2) On retient l' $AR(1)$ ,  $\hat{e}_t = \phi \hat{e}_{t-1} + u_t$  où  $u \sim BB(0, \sigma_u^2)$ , et on estime  $\hat{\sigma}_u^2 = 4.84$ . IC au niveau 95% pour  $\phi$ . Tester :  $\rho(2) = \rho(1)^2$ .

(4-b) Différentiant  $X$  à l'ordre 4, on obtient  $Y_t = \nabla_4 X_t = X_t - X_{t-4}$ ,  $t = 5, 196$ . On suppose de plus que  $e_t = \phi e_{t-1} + u_t$ ,  $u \sim BB(0, \sigma_u^2)$ ,  $|\phi| < 1$ .

(b1) - Vérifier que  $Y_t = \mu + v_t$  est stationnaire. Identifier  $\mu = E(Y_t)$  et la structure  $ARMA$  de  $v$ .

(b2) -  $v$  est-il un  $ARMA$  inversible ? Identifier l'écriture  $v_t = \sum_{i=0}^{\infty} \psi_i u_{t-i}$ , de  $v$ .

(b3) - On a les estimations empiriques  $\hat{\mu} = 0.48$  et  $\hat{c} = \sum_{|h| \leq 192} \hat{\gamma}_Y(h) \{1 - \frac{|h|}{192}\} = 2.44$ . Tester au niveau 5% ( $H_0$ ) :  $b = 0$  contre ( $H_1$ ) :  $b > 0$ .

## 8 Prédiction sans modèle

Observant  $x_1, x_2, \dots, x_n$ , on veut prévoir  $x_{n+h}$ ,  $h > 0$  par  $\hat{x}_{n,h}$  sans connaissance de modèle pour  $X$ . On va présenter trois méthodes linéaires : le lissage exponentiel simple (LES), le LE double (LED) et le lissage de Holt-Winters (LHW) et de Holt-Winters saisonnier (LHWS; cf. POV).

### 8.1 Le lissage exponentiel simple

Soit  $0 < \alpha < 1$ ; la prédiction  $\hat{x}_{n,h}$  du LES (ici indépendante de  $h$ ) est la valeur  $a$  qui minimise le critère des moindres carrés pondérés (MCP) :

$$C(a) = \sum_{j=1, n} \alpha^j (x_{n-j} - a)^2 \text{ soit } \hat{x}_{n,h} = \hat{x}_{n+h}(\alpha) = (1 - \alpha) \sum_{j=1, n-1} \alpha^j x_{n-j}.$$

$\alpha$  est un paramètre de lissage : plus  $\alpha$  est proche de 1, plus la prédiction va loin vers le passé ; plus  $\alpha$  est proche de 0, plus la dépendance est forte dans les valeurs de  $X_t$  pour  $t$  proche de  $n$ . Le choix de  $\alpha$  se fait soit a priori, soit de façon automatique, par exemple en minimisant un critère des MC sur les erreurs de prédiction sur les valeurs observées :

$$\alpha^* = \arg \min \{Q(\alpha) = \sum_{i=2}^{n-1} (x_{i+1} - \hat{x}_{i,1}(\alpha))^2\}. \quad (42)$$

### 8.2 Le lissage exponentiel double

On modifie le LES en ajustant les observations au voisinage de  $n$  sur la droite  $y_t = a_1 + (t - n) \times a_2$  (le LES prend  $a_2 = 0$ ), ceci pour le même critère des MCP :

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

Quelques manipulations donnent :

$$\begin{aligned} \hat{a}_{1,n} &= 2L_1(n) - L_2(n), \hat{a}_{2,n} = \frac{1 - \alpha}{\alpha} \{L_1(n) - L_2(n)\} \text{ où} \\ L_1(n) &= (1 - \alpha) \sum_{j=0, n-1} \alpha^j x_{n-j} \text{ et } L_2(n) = (1 - \alpha) \sum_{j=0, n-1} \alpha^j L_1(n - j). \end{aligned}$$

On en déduit les formules de mises à jour :

$$\begin{aligned}\widehat{a}_1(n) &= \widehat{a}_1(n-1) + \widehat{a}_2(n-1) + (1-\alpha^2)(x_n - \widehat{x}_{n-1,1}), \\ \widehat{a}_2(n) &= \widehat{a}_2(n-1) + (1-\alpha)^2(x_n - \widehat{x}_{n-1,1}),\end{aligned}\tag{43}$$

Pour faire tourner ces formules, on peut prendre comme valeurs initiales  $\widehat{a}_1(0) = x_1$  et  $\widehat{a}_2(0) = x_2 - x_1$ . Ces mises à jour permettent donc un calcul récursif simple des prédictions. Là aussi le paramètre  $\alpha$  est choisi a priori ou calculé en minimisant le critère (42).

### 8.3 La méthode Holt-Winters

C'est une variante du LED par modification des formules de mises à jour (43). Cette variante dépend de deux paramètres  $\beta$  et  $\gamma$  (paramètres qu'il faudra choisir) de la façon suivante :

1. Pour l'ajustement d'une *droite affine*  $x_t = a_1 + (t-n) \times a_2$ , on prend (LHW) :

$$\begin{aligned}\widehat{a}_1(n) &= \beta x_n + (1-\beta)[\widehat{a}_1(n-1) + \widehat{a}_2(n-1)], \\ \widehat{a}_2(n) &= \gamma(\widehat{a}_1(n) - \widehat{a}_1(n-1)) + (1-\gamma)\widehat{a}_2(n-1), \\ \widehat{x}_{n,h} &= \widehat{a}_1(n) + h \times \widehat{a}_2(n).\end{aligned}$$

2. Pour l'ajustement d'une *droite affine + une composante saisonnière* de période  $d$  :

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

on prendra (LHWS) :

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

Avec pour valeurs d'initialisation  $\widehat{a}_1(d+1) = x_{d+1}$ ,  $\widehat{a}_2(d+1) = \frac{x_{d+1}-x_1}{d}$  et  $\widehat{s}_j = x_j - (x_1 + (d-1)\widehat{a}_2(d+1))$  pour  $j = 1, d$ .

Fonctions R utiles

Package `stats` : `HoltWinters` (saisonnier, non saisonnier, lissages exponentiels).

Package `forecast` : `forecast.HoltWinters`, `ses` (prévision par lissage exponentiel)

## 9 Divers

### 9.1 Série temporelle bivariée

Une série bivariée est une série d'observations bidimensionnelles  $X_t = {}^t(X_{1t}, X_{2t})$ ,  $t = 1, 2, 3, \dots$ . Chaque série composante ( $X_{1t}$ ) et ( $X_{2t}$ ) peut être étudiée séparément au second ordre. Mais on perd alors l'information de corrélation entre les deux séries : il faudra donc étudier la série dans sa version bidimensionnelle.

$(X_t, t = 1, 2, \dots)$  est stationnaire au second ordre si la moyenne  $E(X_t) = {}^t(m_1(t), m_2(t)) = {}^t(m_1, m_2)$  est constante et si de plus pour tout  $t$  et  $h$ ,

$$Cov(X_{t+h}, X_t) = \Gamma(h) = \begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{bmatrix}.$$

Les éléments diagonaux de  $\Gamma(h)$  sont les acvf des séries coordonnées ; le terme non-diagonal  $\gamma_{12}(h)$  est l'acvf croisée des deux composantes à distance  $h$ ,

$$\gamma_{12}(h) = Cov(X_{1,t+h}, X_{2,t}).$$

On a  $\gamma_{12}(h) = \gamma_{21}(-h)$  pour tout  $h$ . Par contre, en général,

$$\gamma_{12}(h) \neq \gamma_{21}(h),$$

c-à-d qu'on n'a pas de symétrie par réflexion : la covariance croisée  $\Gamma(h)$  à distance  $h \neq 0$  n'est pas symétrique en général. Ceci disymétrie s'explique bien : si par exemple  $X_{1t}$  est le niveau de pluie le mois  $t$  en un lieu donné et  $X_{2t}$  le rendement d'un verger de légumes en ce même lieu le mois  $t$ , la corrélation  $\rho_{12}(1)$  est  $> 0$  alors que  $\rho_{21}(1)$  est nulle.

Si l'observation est  $X_1, X_2, \dots, X_n$ , les estimations empiriques de la moyenne et des covariances sont :

$$\begin{aligned} \hat{m} &= \bar{X}_n = \frac{1}{n} \sum_{t=1, n} X_t, \\ \hat{\Gamma}(h) &= \frac{1}{n} \sum_{t=1, n-h} (X_{t+h} - \bar{X}_n)^t (X_t - \bar{X}_n) \text{ pour } 0 \leq h \leq n-1, \\ \hat{\Gamma}(h) &= {}^t\hat{\Gamma}(-h) \text{ pour } -n+1 \leq h \leq 0. \end{aligned}$$

Reprenant les résultats unidimensionnels du §1.5,  $\hat{m} \rightarrow m$  en moyenne quadratique et à une vitesse en  $\frac{1}{n}$  si  $\sum_h \{|\gamma_{11}(h)| + |\gamma_{22}(h)|\} < \infty$ . De plus, la formule de Bartlett unidimensionnelle (6) donne les intervalles de confiance des corrélations empiriques  $\hat{\rho}_{11}(h)$  et  $\hat{\rho}_{22}(h)$ , en particulier si  $X$  est un  $BB$ .

Pour les corrélations croisées  $\hat{\rho}_{12}(h)$ , on a le résultat suivant, cas particulier de la formule de Bartlett en situation multidimensionnelle (cf. BD). Supposons que  $X$  est gaussienne et que soit  $(X_{1t})$ , soit  $(X_{2t})$  est un bruit blanc :

$$\text{Si } \rho_{12}(h) = 0 \text{ pour } h \in [a, b], \text{ alors } nVar(\hat{\rho}_{12}(h)) \approx 1 \text{ pour } h \in [a, b].$$

Ceci permet de tracer une bande de confiance pour les  $\hat{\rho}_{12}(h)$  si l'une des composantes est un  $BB$  et si  $X$  est gaussienne.

### 9.1.1 Exercices et fonctions R utiles

1.  $X$  une série bivariable de composantes  $MA(1)$  est engendrée par deux  $BB$  indépendants  $e_1 \sim BB(0, \sigma_1^2)$  et  $e_2 \sim BB(0, \sigma_2^2)$  :

$$X_{1t} = e_{1,t} + 0.8e_{1,t-1} \text{ et } X_{2t} = e_{2,t} - 0.6e_{2,t-1}.$$

On observe  $X_1, X_2, \dots, X_n$ . Déterminer la variance asymptotique de  $n^{-\frac{1}{2}}\hat{\rho}_{12}(h)$  ainsi que la covariance asymptotique de  $n^{-\frac{1}{2}}\hat{\rho}_{12}(h)$  et  $n^{-\frac{1}{2}}\hat{\rho}_{12}(k)$  si  $h \neq k$ .

2. **ldeaths** sont 3 séries bidimensionnelles (population des hommes et des femmes) donnant le nombre mensuel de morts par bronchite, emphysème et asthme en GB sur la période 1974–1979. Etudier l'acvf de ces 3 séries. A-t-on la symétrie par réflexion de l'autocovariance croisée ( $\gamma_{12}(h) = \gamma_{21}(h)$ ) ?

Fonctions R utiles :

Package **stat** : **ccf** (acvf et acf croisées)

Package **fSeries** (Financial Time series) : **cov** et **cor** (acvf. et acf. croisées).

## 9.2 Modèle ARCH

Pour un modèle linéaire stationnaire  $X_t = \sum_{i \geq 0} \phi_i X_{t-i} + e_t$  où  $e \sim BB(0, \sigma^2)$ , la variance conditionnelle au passé  $Var(X_t | x_{t-1}, x_{t-2}, \dots)$  est constante. C'est une limitation si on observe que la *volatilité*  $h$  dépend du passé  $x_{t-1}, x_{t-2}, \dots$  et est fluctuante, où :

$$h(x_{t-1}, x_{t-2}, \dots) = Var(X_t | x_{t-1}, x_{t-2}, \dots).$$

Engle a introduit en 1982 des modèles stationnaires mais non-linéaires, les modèles  $ARCH(p)$  (*AR with Conditional Heterocedasticity*) pour lesquels les variances conditionnelles  $h(x_{t-1}, \dots, x_{t-p})$  s'écrivent comme une  $p$ -régression sur les  $X_{t-i}^2$  de la façon suivante :

$$\begin{aligned} X_t &= \sqrt{h_t} \times e_t \text{ où } e \sim i.i.d. \mathcal{N}(0, 1) \text{ et} \\ h_t &= \alpha_0 + \sum_{i=1, p} \alpha_i X_{t-i}^2, \end{aligned} \quad (44)$$

$\alpha_0 > 0$  et  $\alpha_i \geq 0$  pour  $i = 1, p$ . L'étude de ces modèles non-linéaires et leur généralisation, les  $GARCH(p, q)$ , est mathématiquement plus difficile et a fait l'objet d'une abondante littérature.

Examinons plus précisément le cas d'un modèle  $ARCH(1)$ . Si  $\alpha_1 < 1$ , il existe une unique solution stationnaire au second ordre et causale à (44), centrée, avec  $Var(X_t) = \frac{\alpha_0}{1-\alpha_1}$  et  $\gamma(h) = Cov(X_t, X_{t+h}) = 0$  si  $h \neq 0$  ( $X$  est un BB faible), la loi de  $X_t$  étant symétrique.

Plusieurs caractères différencient cependant ce modèle d'un BB fort ou d'un BB gaussien et peuvent permettre l'identification d'un  $ARCH$  :

1.  $(X_t)$  n'est pas un BB fort :  $E(X_t^2 | X_{t-1}) = (\alpha_0 + \alpha_1 X_{t-1}^2) E(e_t^2 | X_{t-1}) = (\alpha_0 + \alpha_1 X_{t-1}^2)$ ;
2. Si  $3\alpha_1^2 < 1$ ,  $E(X_t^4) < \infty$ , et la série  $(Y_t = X_t^2)$  est un  $AR(1)$  d'acf  $\rho_Y(h) = \alpha_1^{|h|}$  (autre façon de voir que  $X$  est dépendant) ;
3.  $X$  n'est pas gaussien (s'il l'était, étant non-corrélé, ce serait un bruit i.i.d.) ;
4.  $\exists k$  t.q.  $E(X_t^{2k}) = \infty$  : la loi marginale  $X_t$  est à *queue de distribution épaisse* (c.à-d que la densité  $f$  de  $X_t$  vérifie  $\int^{+\infty} t^{2k} f(u) du = \infty$ ).

*Estimation de l'ARCH(1).*

$X$  est *conditionnellement gaussien*, la vraisemblance de  $x_1, \dots, x_n$ , conditionnelle à  $x_1$  est :

$$\begin{aligned} L_n(\alpha_0, \alpha_1 \mid x_1) &= \prod_{t=2, n} p(x_t \mid x_{t-1}, \dots, x_1) \\ &= \prod_{t=2, n} \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 x_{t-1}^2)}} \exp -\frac{x_t^2}{2(\alpha_0 + \alpha_1 x_{t-1}^2)}. \end{aligned}$$

On en déduit l'estimation des paramètres puis de la volatilité stochastique  $\hat{h}_t = \{\hat{\alpha}_0 + \hat{\alpha}_1 x_{t-1}^2\}^{\frac{1}{2}}$ .

Ces résultats se généralisent à l' $ARCH(p)$  et au modèle  $GARCH(p, q)$  pour lequel  $h_t$  vérifie :

$$h_t = \alpha_0 + \sum_{i=1, p} \alpha_i X_{t-i}^2 + \sum_{j=1, q} \beta_j h_{t-j}^2, \quad \alpha_0 > 0, \alpha_i \text{ et } \beta_j \geq 0.$$

Signalons que pour les  $ARCH$  ou les  $GARCH$ , d'autres choix de loi pour le BB générateur  $e$  sont proposés, par exemple celui où  $e$  suit une loi de Student  $T_\nu$  à  $\nu$  d.d.l. et réduite :  $e \sim \sqrt{\frac{\nu}{\nu-2}} T_\nu$ .

### 9.2.1 Exercices et fonctions R utiles

1.  $Y$  est la série des “pourcentages de retours journaliers du Dow Jones Industrial Index” mesurés du 1er juillet 1997 au 9 avril 1999 ( $n = 464$  observations, cf. série E1032.ITSM de ITSM de BD (2002)).
  - (a) - Montrer que  $Y$  peut être modélisé par un  $BB$  faible.
  - (b) - Observer les acvf de  $\{|Y_t|\}$ , de  $\{Y_t^2\}$ , l’histogramme de  $\{Y_t\}$  et le  $QQ$ -plot de  $\{Y_t\}$ . Quelle conséquence en tirez ?
  - (c) - Modéliser  $Y$  par un  $GARCH(1, 1)$ . L’estimer. En déduire une estimation de la volatilité de  $Y$ .
2. Données **sunspot** : un  $ARMA$  à résidu  $GARCH$  (cf. BD, 2002).
  - (a) - Une fois retranchée la moyenne, ajuster  $X$  sur un modèle  $ARMA$ .
  - (b) - Observant son acvf, vérifier que le résidu  $e$  de cet  $ARMA$  est bien un  $BB$  mais que  $|e|$  et  $e^2$  ne le sont pas.
  - (c) - Ajuster  $e$  sur un  $Garch(1, 1)$  et estimer la volatilité  $h(t)$  du résidu. Comparer l’ $ARMA$  standard à celui à résidu  $GARCH$  sur la base de leur  $AIC$ .
  - (d) Simuler le modèle (a) ainsi que le modèle (c). Comparer aux données originales.

Quelques fonctions R utiles :

Package **fGarch** : `garchSim`, `garchFit`, `garchPredictors`  
Package **tseries** : `garch`

## 10 Annexes

### 10.1 Loi gaussienne multidimensionnelle

• La densité de  $X \sim \mathcal{N}(m, \sigma^2)$ , la loi gaussienne réelle de moyenne  $m$  et de variance  $\sigma^2 > 0$  est  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2} \frac{(x-m)^2}{\sigma^2}$ . La fdr  $\Phi$  normale réduite  $\mathcal{N}(0, 1)$  vérifie  $\Phi(-x) = 1 - \Phi(x)$ ; elle est tabulée pour  $x \geq 0$ .

• La loi gaussienne multidimensionnelle  $X$  à valeur dans  $\mathbb{R}^p$ , de moyenne  $m \in \mathbb{R}^p$  et de matrice de covariance régulière  $\Sigma = p \times p$  admet pour densité :

$$f(x) = (2\pi)^{-\frac{p}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp -\frac{1}{2} {}^t(x - m)\Sigma^{-1}(x - m)$$

Cette loi est caractérisée par sa moyenne et sa variance.

Soit  $X = ({}^tX(1), {}^tX(2))$  une décomposition de  $X$  en deux composantes de dimensions respectives  $p_1$  et  $p_2$ . Notons  $m(i)$ ,  $\Sigma_{ij}$ ,  $i, j = 1, 2$  les décompositions de  $m$  et de  $\Sigma$  associées. On a :

- (1) Si  $A : \mathbb{R}^p \rightarrow \mathbb{R}^q$  est linéaire, alors  $AX \sim \mathcal{N}_q(Am, A\Sigma^tA)$ .
- (2) Lois marginales :  $X(1) \sim \mathcal{N}_{p_1}(m(1), \Sigma_{11})$  et  $X(2) \sim \mathcal{N}_{p_2}(m(2), \Sigma_{22})$ .
- (3)  $X(1)$  et  $X(2)$  sont indépendantes si et seulement si elles sont non-corrélées :  $\Sigma_{12} = 0$ .
- (4) La loi de  $X(1)$  conditionnelle à  $X(2) = x_2$  est gaussienne :

$$\mathcal{L}oi(X(1) | X(2) = x_2) \sim \mathcal{N}_{p_1}(m(1) + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - m(2)), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

L’espérance conditionnelle est affine en  $x_2$ ; la variance est indépendante de  $x_2$ . Pour un couple, on a :

- Couple  $(X, Y) \sim \mathcal{N}_2(m, \Sigma)$  où  $Var(X) = \sigma_1^2$ ,  $Var(Y) = \sigma_2^2$ ,  $\rho = Corr(X, Y)$  :

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp -\frac{1}{2(1-\rho^2)} \{\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2\} \text{ où } \tilde{z} = \frac{z - E(z)}{\sigma(z)}$$

$$\mathcal{L}oi(Y | X = x) \sim \mathcal{N}(m_2 + \rho\sigma_2\tilde{x}, \sigma_2^2(1 - \rho^2)) \text{ et } E(Y | X = x) = m_2 + \rho\sigma_2\tilde{x}$$

$X$  et  $Y$  sont indépendantes si et seulement si elles sont non-corrélées. L'espérance conditionnelle définit la droite de régression de  $Y$  en  $X = x$ .

## 10.2 Convergence en moyenne quadratique

Une suite  $(S_n)$  de v.a.r. est dite converger vers  $S$  en moyenne quadratique (m.q.) si  $E(S_n - S)^2 \rightarrow 0$ . On note alors  $S_n \xrightarrow{L^2} S$ .

*Critère de Cauchy* : il permet d'établir la convergence sans avoir à identifier sa limite :

$$“(S_n) \text{ converge dans } L^2 \text{ si et ssi } E(S_n - S_m)^2 \rightarrow 0 \text{ pour } m, n \rightarrow \infty.”$$

*Exemple* : si  $X_n = \sum_{-n}^{+n} a_i e_i$  où  $e$  est un  $BB(0, \sigma^2)$ ,  $(X_n)$  converge dans  $L^2$  si et ssi  $\sum_{m \leq i \leq n} a_i^2 \rightarrow 0$  pour  $m, n \rightarrow \infty$ , c-à-d si et ssi  $\sum_{-\infty}^{+\infty} a_i^2 < \infty$ .

Si  $X_n \rightarrow X$  et  $Y_n \rightarrow Y$  en m.q., alors  $X_n + Y_n \rightarrow X + Y$  et  $a_n X_n \rightarrow aX$  en m.q. si  $(a_n)$  est une suite numérique convergeant vers  $a$ . On a :

$$E(X_n) \rightarrow E(X), E(X_n^2) \rightarrow E(X^2) \text{ et } E(X_n Y_n) \rightarrow E(XY).$$

## 10.3 Modèle linéaire

Supposons que les observations réelles  $Y_i, i = 1, n$ , suivent le modèle paramétrique linéaire :

$$Y_i = {}^t z_i \beta + e_i, i = 1, n$$

où  $\beta \in \mathbb{R}^p$  est un paramètre inconnu,  $z_i \in \mathbb{R}^p$  est une variable exogène (explicative) observée et  $(e_i)$  des résidus centrés et de même variance  $Var(e_i) = \sigma^2 < \infty$ . L'écriture matricielle du modèle est :

$$Y = Z\beta + e,$$

où  $Y$  et  $e \in \mathbb{R}^n$ , et  $Z$  est la matrice exogène  $n \times p$  dont la ligne  $i$  est  ${}^t z_i$ .

- *Estimation des Moindres Carrés Ordinaires (MCO)*.

C'est la valeur (unique) qui minimise  $\|Y - Z\beta\|_2^2$  où  $\|\cdot\|_2$  est la norme euclidienne sur  $\mathbb{R}^n$ . Si  $Z$  est de rang  $p$ , on a :

$$\begin{aligned} \hat{\beta} &= ({}^t Z Z)^{-1} {}^t Z Y \text{ et } E(\hat{\beta}) = \beta \text{ et} \\ Var \hat{\beta} &= ({}^t Z Z)^{-1} {}^t Z Var(e) Z ({}^t Z Z)^{-1}. \end{aligned}$$

- *Cas où  $e \sim BB(0, \sigma^2)$  : le théorème de Gauss- Markov.*

$\hat{\beta}$  est la meilleure estimation (en variance-covariance) parmi celles qui sont linéaires et sans biais, et :

$$Var \hat{\beta} = \sigma^2 ({}^t Z Z)^{-1}.$$

Sous cette hypothèse de  $BB$  pour  $e$ ,  $\sigma^2$  est estimé sans biais par,

$$\hat{\sigma}^2 = S^2 = \frac{\|Y - Z\hat{\beta}\|_2^2}{n - p}.$$

Si de plus  $Y$  est gaussienne,  $\widehat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2({}^tZZ)^{-1})$ ,  $\widehat{\sigma}^2 \sim \sigma^2 \frac{\chi_{n-p}^2}{n-p}$  et  $\widehat{\sigma}^2$  est indépendante de  $\widehat{\beta}$ . Les tests de Student et de Fisher sur les sous-hypothèse linéaires découlent de ces propriétés.

••• *Cas e centré de  $L^2$  mais non BB.*

Il y a hétéroscédasticité des résidus si  $Var(e) = \sigma^2\Omega$  où  $\Omega$  est une matrice  $n \times n$  régulière non proportionnelle à la matrice identité (i.e.  $e \sim ARMA$ ). L'estimation des MCG de  $\beta$  est la valeur  $\widehat{\beta}_{MCG}$  qui minimise  $\|Y - Z\beta\|_{\Omega^{-1}}^2$  où  $\|U\|_{\Omega^{-1}}^2 = {}^tU\Omega^{-1}U$ . On a :

$$\begin{aligned}\widehat{\beta}_{MCG} &= ({}^tZ\Omega^{-1}Z)^{-1} {}^tZ\Omega^{-1}Y, \quad E(\widehat{\beta}) = \beta \text{ et} \\ Var\widehat{\beta}_{MCG} &= \sigma^2({}^tZ\Omega^{-1}Z)^{-1} {}^tZ\Omega Z ({}^tZ\Omega^{-1}Z)^{-1}.\end{aligned}$$

$\widehat{\beta}_{MCG}$  est la meilleure estimation en variance parmi les estimations linéaires et sans biais.

## 10.4 Théorème central limite (TCL)

**Le TCL standard.**

Si  $\{X_n\} \sim i.i.d.(m, \sigma^2)$  et  $\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ , alors

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow{loi} \mathcal{N}(0, \sigma^2) \quad \text{ou encore : } \overline{X}_n \sim \mathcal{AN}(m, \frac{\sigma^2}{n}).$$

**TCL de Lyapunov.**

Il concerne des variables *i.n.i.d* (indépendantes mais non-nécessairement *i.d.*) : soit  $(X_i, i = 1, 2, \dots)$  une suite de v.a.r. indépendantes et centrées de moments d'ordre 3 finis,  $S_n = \sum_1^n X_i$ ,  $s_n^2 = Var(S_n) = \sum_1^n Var(X_i)$  et  $\kappa_n^3 = \sum_1^n E(|X_i|^3) < \infty$ . Alors,

$$\text{Si } \frac{\kappa_n}{s_n} \rightarrow 0 : \frac{S_n}{s_n} \xrightarrow{loi} \mathcal{N}(0, 1).$$

**Le TCL pour un processus stationnaire et  $m$ -dépendant.**

Supposons que  $X = \{X_i\}$  est un processus *strictement stationnaire* et au second ordre, de moyenne  $m$ , d'acvf  $\gamma$ , de variance  $\gamma(0) = \sigma^2 < \infty$ . On dit que  $X$  est  *$m$ -dépendant* si pour tout  $i, j$  tels que  $|i - j| > m$ ,  $X_i$  et  $X_j$  sont indépendants. Notons  $v_m = \gamma(0) + 2 \sum_{j=1, m} \gamma(j)$ . Alors, si  $v_m \neq 0$ , le TCL standard est encore valable avec  $nVar\overline{X}_n = v_m$  :

$$\overline{X}_n \sim \mathcal{AN}(\mu, \frac{v_m}{n}).$$

Signalons qu'il existe des versions plus générales des TCL pour des processus stationnaires et *faiblement dépendants*.

**Loi limite de  $g(\overline{X}_n)$  si  $\overline{X}_n \sim \mathcal{AN}(\mu, \sigma_n^2)$  et  $g \in \mathcal{C}^1$ .**

(a) *Le cas unidimensionnel* : si  $g$  est différentiable en  $\mu$ , si  $\sigma_n^2 \rightarrow 0$  et si  $g'(\mu) \neq 0$  :

$$g(\overline{X}_n) \sim \mathcal{AN}(g(\mu), g'(\mu)^2 \times \sigma_n^2). \quad (45)$$

(b) *Le cas multidimensionnel* : supposons que  $\bar{X}_n \in \mathbb{R}^p$ , que  $\bar{X}_n \sim \mathcal{AN}_p(\boldsymbol{\mu}, \sigma_n^2 \Sigma)$  où  $\boldsymbol{\mu} \in \mathbb{R}^p$ ,  $\sigma_n^2 \rightarrow 0$  et  $\Sigma$  est une matrice  $p \times p$  d.p.. Soit  $g(X) = (g_1(X), g_2(X), \dots, g_m(X))$  où  $g : \mathbb{R}^p \rightarrow \mathbb{R}^m$  est continuellement différentiable au voisinage de  $\boldsymbol{\mu}$ , de matrice jacobienne  $m \times p$  :

$$J(\boldsymbol{\mu}) = \left[ \frac{\partial g_i}{\partial x_j}(\boldsymbol{\mu}) \right]_{i=1, m}^{j=1, p}.$$

Alors si  $J(\boldsymbol{\mu}) \Sigma^t J(\boldsymbol{\mu})$  a tous ses éléments diagonaux non-nuls :

$$g(\bar{X}_n) \sim \mathcal{AN}_p(g(\boldsymbol{\mu}), c_n^2 \times J(\boldsymbol{\mu}) \Sigma^t J(\boldsymbol{\mu})).$$

## 10.5 R et le traitement des séries temporelles

Pour l'installation de R, on consultera par exemple le site de Anne Philippe <sup>(1)</sup> qui donne : (i) le lien avec la **Homepage** de R; (ii) le bouton **download** de chargement en ligne de R à partir d'un site miroir à choisir; (iii) une présentation rapide générale de R.

On trouvera les informations générales sur R à partir de la fenêtre **Rconsole** : aller dans **Aide**  $\rightarrow$  **Aide HTML**. Vous disposer alors de : packages de R, langage R, Introduction à R, auteurs, références .....

Pour utiliser les programmes (ou fonctions) de R, les packages d'intérêt doivent être préalablement chargés s'ils ne le sont pas par défaut :

- pour *installer* un package, aller dans **RGui**, choisir le site miroir et l'option installer le(s) package(s). Une fois installé, un package doit être chargé s'il ne l'est pas par défaut.
- pour charger un package : **Rgui**  $\rightarrow$  **Package**  $\rightarrow$  **charger le package**.

Une fois une fonction d'un package choisie, vous aurez en cliquant dessus l'explication de la fonction, la description de ses paramètres, des références et sites attachés (données, articles, etc...). Le plus souvent, ces descriptions se terminent par un **script** décrivant un exemple simple d'application : une première possibilité est de copier-coller ce **script** et de le faire tourner dans **Rgui** pour voir quelles sont les sorties, en particulier graphiques, de la fonction.

### 10.5.1 Exemples de jeux de données disponibles dans R

#### Dans le package datasets

**AirPassengers** : The classic Box & Jenkins airline data. Monthly totals of international airline passengers, 1949 to 1960.

**austres** : Numbers (in thousands) of Australian residents measured quarterly from March 1971 to March 1994.

**co2** : Loa Atmospheric concentrations of CO2 are expressed in parts per million (ppm). A time series of 468 observations; monthly from 1959 to 1997.

**EuStockMarkets** : Contains the daily closing prices of major European stock indices. A multivariate time series with 1860 observations on 4 variables.

**faithful** : Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. A data frame with 272 observations on 2 variables : [,1] time in mins; [,2] time to next eruption (in mins).

**ldeaths**, **fdeaths**, **mdeaths** : Three time series giving the monthly deaths from bronchitis, emphysema and asthma in the UK, 1974–1979, both sexes (**ldeaths**), males (**mdeaths**) and females (**fdeaths**).

---

<sup>1</sup><http://www.math.sciences.univ-nantes.fr/~philippe/>

**LakeHuron** : Annual measurements of the level, in feet, of Lake Huron 1875–1972 ( $n = 98$ ).

**lynx** : Annual numbers of lynx trappings for 1821–1934 in Canada (BD, 1991).

**Nile** : Measurements of the annual flow of the river Nile at Ashwan 1871–1970 ( $n = 100$ ).

**nottem** : A time series containing average air temperatures at Nottingham Castle in degrees Fahrenheit for 20 years.

**sunspot.month** : Monthly numbers of sunspots. The univariate time series **sunspot.year** and **sunspot.month** contain 289 and 2988 observations, respectively.

**treering** : Contains normalized tree-ring widths in dimensionless units. A univariate time series with 7981 observations. Each tree ring corresponds to one year.

**USAccDeaths** : A time series giving the monthly totals of accidental deaths in the USA.

### Dans le package **tseries**

**bev** : Contains the well-known Beveridge Wheat Price Index which gives annual price data from 1500 to 1869, averaged over many locations in western and central Europe. A univariate time series with 370 observations. The object is of class "ts".

**camp** : Mount Campito Yearly Treering Data, contains annual tree-ring measurements from Mount Campito from 3426 BC through 1969 AD.

**NelPlo** : These are the extended Nelson-Plosser Data, contains the 14 following macroeconomic series : consumer price index, industrial production, nominal GNP, velocity, employment, interest rate, nominal wages, GNP deflator, money stock, real GNP, stock prices (S&P500), GNP per capita, real wages, unemployment.

**ice.river** : Contains the Icelandic river : 4 univariate time series **flow.vat**, **flow.jok**, **prec**, and **temp**, each with 1095 observations.

**nino** : Two univariate time series **nino3** and **nino3.4** with 598 observations and the joint series **nino**.

**Dans ITSM** (BD, 2002) : ITSM est un tutoriel joint au livre de BD (2002). On y trouve également un certain nombre de données.

### 10.5.2 Récapitulatif des fonctions R utiles à l'étude des séries temporelles

- Package **stats** : **rxxx(n,paramètres)**(simulation d'un  $n$ -échantillon de la loi **xxx**), **acf**, **qqplot**, **Box.test** (tests de Box-Pierce et Ljung-Box), **decompose** (en  $T(t)+S(t)+e(t)$ ), **monthplot** (ajustement saisonnier), **arima.sim**, **ARMAacf**, **ARMAtoMA** (écriture  $MA(\infty)$  d'un *ARMA*), **predict.Arima**, **acf2AR** (AR s'ajustant exactement à une acvf), **ar** (par défaut, meilleur *AR* par AIC), **arima**, **lm** (ajustement sur un modèle linéaire), **nls** (Non-linear Least Squares, régression non-linéaire par MCP), **predict.lm** (prédiction à partir d'un ML), **ccf** (acvf et acf croisées).

- Package **tseries** : **jarque.bera.test**, **run.test** (test des runs du caractère *i.i.d.* d'une série binaire), **kpss.test** (test de stationnarité), **adf.test** (test de Dickey-Fuller), **garch**.

- Package **forecast** : **BoxCox**, **InvBoxCox**, **seasadj** (désaisonnalise une série), **forecast.Arima**, **auto.arima** (meilleur ARIMA avec AIC).

- Package **fArma** : **armaSim**, **armaFit**, **armaRoots**, **armaTrueacf**.

- Package **fSeries** (Financial time series) : **cov** et **cor** (acvf. et acf. croisées).

- Package **fGarch** : **garchSim**, **garchFit**, **garchPredictors**.

### 10.5.3 Quelques jeux de données

**Dow Jones Utilities Index** :  $n = 78$  données journalières du 28 août au 18 décembre 1972 (B&D (1991), fichier DOWJ.TSM dans ITSM, BD (2002))

10.94, 10.69, 10.43, 10.56, 10.75, 10.84, 10.46, 10.56, 10.46, 10.05, 9.6, 9.31, 9.31, 9.25, 9.02, 8.54, 8.77, 9.02, 9.44, 9.38, 9.53, 9.89, 10.56, 10.56, 10.72, 11.23, 11.48, 11.58, 11.9, 12.19, 12.06, 11.96, 11.68, 11.36, 11.42, 12, 12.22, 12.7, 13.15, 14.36, 14.65, 15.06, 15.86, 16.4, 16.44, 16.88, 18.07, 18.51, 19.28, 19.79, 19.7, 19.28, 19.66, 20.14, 20.97, 21.13, 21.55, 21.96, 22.26, 23.79, 24.11, 24.14, 23.37, 23.02, 22.86, 23.02, 23.11, 23.05, 23.05, 22.83, 23.18, 22.67, 22.73, 22.86, 22.67, 22.09, 22, 21.23

**Seat Belt Law** : nombre de décès par accident les mois  $t$  sur les routes d'Angleterre de janvier 1975 à décembre 1984 ( $n = 120$ ).

1577	1356	1652	1382	1519	1421	1442	1543	1656	1561	1905	2199	1473	1655	1407
1395	1530	1309	1526	1327	1627	1748	1958	2274	1648	1401	1411	1403	1394	1520
1528	1643	1515	1685	2000	2215	1956	1462	1563	1459	1446	1622	1657	1638	1643
1683	2050	2262	1813	1445	1762	1461	1556	1431	1427	1554	1645	1653	2016	2207
1665	1361	1506	1360	1453	1522	1460	1552	1548	1827	1737	1941	1474	1458	1542
1404	1522	1385	1641	1510	1681	1938	1868	1726	1456	1445	1456	1365	1487	1558
1488	1684	1594	1850	1998	2079	1494	1057	1218	1168	1236	1076	1174	1139	1427
1487	1483	1513	1357	1165	1282	1110	1297	1185	1222	1284	1444	1575	1737	1763

### Références

- Akaike H. (1969), Fitting autoregressive models for prediction, *Ann. Inst. Stat. Math.*, 21, 243-247.  
Box G.E.P. et Jenkins G.M. (1976), *Time Series Analysis : Forecasting and Control*, Holden-Day.  
Brockwell P.J. et Davis R.A. (1991), *Time Series : Theory and Methods*, Springer.  
Brockwell P.J. et Davis R.A. (2002), *Introduction to Time Series and Forecasting*, Springer (avec tutoriel ITSM).  
Dickey D.A. et Fuller W.A. (1979), Distribution of the estimators for autoregressive time series with a unit root, *J. Amer. Stat. Soc.*, 74, 427-431.  
Engle R.F. (1982), Autoregressive conditional heterocedasticity with estimates of the variance of UK inflation, *Econometrica*, 50, 145-157.  
Gourieroux C. et Monfort A. (1995), *Séries temporelles et modèles dynamiques*, Economica.

Quelques photocopies “Séries Temporelles” en ligne (entre autres)

Viano, Philippe, *Maîtrise d'Econométrie* de Lille.

Charpentier, *DESS Actuariat et Maths de la décision*, Dauphine-ENSAE.

Aragon, *Introduction aux ST*, GREMAQ, Toulouse.

Cabanal-Duvillard, *Maîtrise MASS* de Paris 5.