

## Modélisation des séries stationnaires

### 1. Auto-corrélation partielle

DÉFINITION 1. Pour un processus stationnaire  $(X_t)$ , le coefficient de corrélation partielle (on dit aussi corrélation partielle) entre  $X_n$  et  $X_1$  est

$$r_{X_2, \dots, X_n}(X_1, X_n) = \frac{1}{\sigma(0)} \text{Cov}(X_1 - \mathbb{E}(X_1 | X_2, \dots, X_{n-1}), X_n - \mathbb{E}(X_n | X_2, \dots, X_n)),$$

le coefficient d'auto-corrélation partielle en  $h$  est défini par

$$\begin{cases} r(h) = r_{X_2, \dots, X_{h-1}}(X_1, X_h) & \text{pour } h \geq 2, \\ r(1) = \rho(1), \\ r(0) = 1, \\ r(-h) = r(h) & \text{pour } h \leq -1. \end{cases}$$

Pour un processus stationnaire,

$$r(h) = r_{X_{n+1}, \dots, X_{n+h-1}}(X_n, X_{n+h}) \text{ pour tout } n.$$

Donc on va pouvoir estimer ces coefficients en faisant des moyennes empiriques (c'est ce que fait la fonction **pacf** de R)

### 2. Les processus linéaires générales

Soit  $\{Y_t\}$  une série que est observée et une série de bruit  $\{e_t\}$  qui n'est pas observée. Le bruit satisfait les hypothèses classiques.

Nous voudrions écrire  $\{Y_t\}$  comme une combinaison linéaire, c'est-à-dire que nous expliquons la série observée par les perturbations aléatoires passées.

$$(1) \quad Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$$

Si c'est une somme infinie, conditions spécifiques doivent être placés sur les coefficients pour que la somme soit convergent.

EXAMPLE 1. Soit le cas que 1 soit sur la forme

$$(2) \quad Y_t = e_t + \psi e_{t-1} + \psi^2 e_{t-2} + \dots = \sum_{i=0}^{\infty} \psi^i e_{t-i} \quad \psi^0 = 1$$

Ici, la condition pour 2 soit convergent est  $|\psi| < 1$

- $\mathbb{E}(Y_t) = 0$
- $\text{Var}(Y_t) = \frac{\sigma_e^2}{1-\psi^2}$
- $\text{Cov}(Y_t, Y_{t-1}) = \frac{\psi \sigma_e^2}{1-\psi^2}$  et  $\text{Cov}(Y_t, Y_{t-k}) = \frac{\psi^k \sigma_e^2}{1-\psi^2}$
- $\text{Corr}(Y_t, Y_{t-1}) = \psi$  et  $\text{Corr}(Y_t, Y_{t-k}) = \psi^k$

La processus est stationnaire, l'autocovariance ne dépend pas au temps  $t$ , seulement à décalage du temps (lag)  $k$

Avant de nous lancer dans le reste de la section, nous avons besoin de définir quelques notions. On définit l'opérateur de décalage sur les suites stationnaires:

$$L : (Y_t)_{t \geq 0} \mapsto \left( L(Y)_t = \begin{cases} Y_{t-1} & \text{si } t \geq 1 \\ 0 & \text{si } t = 0 \end{cases} \right)_{t \geq 0}$$

(ici,  $(Y_t)$  est une suite stationnaire). Pour un polynôme  $P \in \mathbb{C}[X]$  ( $P(X) = \alpha_0 + \alpha_1 X + \dots + \alpha_n X^n$ ,  $n$  étant le degré de  $P$ ), nous définissons l'opérateur  $P(L)$  par :

$$P(L) : (Y_t)_{t \geq 0} \mapsto (P(L)(Y)_t) \text{ avec } P(L)(Y)_t = \begin{cases} \alpha_0 Y_t + \alpha_1 Y_{t-1} + \dots + \alpha_n Y_{t-n} & \text{si } t \geq n, \\ 0 & \text{sinon.} \end{cases}$$

Nous remarquons que pour deux polynômes  $P$  et  $Q$  :

$$P(L)Q(L) = (PQ)(L),$$

où le produit de gauche est la composition.

LEMMA 1. Pour  $\lambda$  un nombre complexe de module  $< 1$ , l'opérateur  $1 - \lambda L$  admet pour inverse

$$(1 - \lambda L)^{-1} = \sum_{k=0}^{+\infty} (\lambda L)^k.$$

PROOF. Pour toute suite stationnaire  $(Y_t)_{t \geq 0}$ , nous regardons à  $t$  fixé :

$$\sum_{k=0}^{+\infty} ((\lambda L)^k Y)_t = \sum_{k=0}^t \lambda^k Y_{t-k},$$

qui est une somme finie. Donc la suite

$$\left( \sum_{k=0}^{+\infty} (\lambda L)^k Y \right)_{t \geq 0}$$

est bien définie. Nous avons

$$\begin{aligned} Y &= (Y_0, Y_1, \dots) \xrightarrow{\sum_{k=0}^{+\infty} (\lambda L)^k} (Y_0, Y_1 + \lambda Y_0, Y_2 + \lambda Y_1 + \lambda^2 Y_0, \dots) \\ &\xrightarrow{1 - \lambda L} (Y_0, (Y_1 + \lambda Y_0) - \lambda Y_0, (Y_2 + \lambda Y_1 + \lambda^2 Y_0) - \lambda(Y_1 + \lambda Y_0) - \lambda^2 Y_0, \dots) \\ &= (Y_0, Y_1, Y_2, \dots). \end{aligned}$$

Donc  $(1 - \lambda L) \sum_{k=0}^{+\infty} (\lambda L)^k = \text{Id}$ . On montre facilement de même que  $\sum_{k=0}^{+\infty} (\lambda L)^k (1 - \lambda L) = \text{Id}$ .  $\square$

### 3. Les processus auto-régressifs

DÉFINITION 2. Un processus  $(X_t)$  est dit auto-régressif d'ordre  $p$  centré s'il vérifie

$$(3) \quad X_t = \epsilon_t + \sum_{j=1}^p a_j X_{t-j}, \text{ pour tout } t \geq 0$$

(avec  $p \in \mathbb{N}^*$ ,  $a_p \neq 0$ ) avec des  $\epsilon_t$  qui forment un bruit blanc centré de variance  $\sigma^2$  ( $p \in \mathbb{N}^*$ ,  $a_1, \dots, a_p \in \mathbb{R}$ ), tels que  $\epsilon_t$  est indépendant de  $X_{t-1}, X_{t-2}, \dots$  pour tout  $t$ . Par convention :  $X_{-k} = 0$  pour tout  $k$  dans  $\mathbb{N}^*$ . On dira aussi que  $(X_t)$  est un processus AR( $p$ ).

On dit que le  $X_t$  "s'explique" par les  $p$  observations précédentes  $(X_{t-1}, X_{t-2}, \dots, X_{t-p})$ . Dans cette définition, le processus  $(\epsilon_t)$  s'appelle processus des innovations.

EXAMPLE 2.  $AR(1)$

$$(4) \quad Y_t = aY_{t-1} + e_t$$

*Nous supposons que la série est stationnaire, et de moyenne zéro*

- $Var(Y_t) = \frac{\sigma_e^2}{1-a^2}$
- $Cov(Y_t, Y_{t-1}) = \frac{a\sigma_e^2}{1-a^2}$  et  $Cov(Y_t, Y_{t-k}) = a^k \frac{\sigma_e^2}{1-a^2}$
- $Corr(Y_t, Y_{t-1}) = a$  et  $Corr(Y_t, Y_{t-k}) = a^k$

RÉMARK 1. De la variance on pourrait suspecter la condition  $|a| < 1$  mais pour le justifier rigoureusement on doit d'aller plus loin sur les propriétés d'un modèle AR.

***L'inversion d'un modèle  $AR(1)$  et la condition pour la stationnarité***

*Nous rendrons le modèle  $AR(1)$  sous la forme récursif (c'est à dire, nous chercherons à faire la recursion sur les valeurs précédents de la série)*

$$Y_t = aY_{t-1} + e_t = e_t + a \underbrace{(aY_{t-2} + e_{t-1})}_{Y_{t-1}}$$

*Si nous répétons la substitution  $k-1$  fois, nous arrivons à*

$$(5) \quad Y_t = e_t + ae_{t-1} + a^2e_{t-2} + \dots + a^{k-1}e_{t-k+1} + a^k e_{t-k}$$

*Ce qui ressemble à le modèle 2, et en plus, si nous laisserions  $k \rightarrow \infty$ , alors nous récupérerons modèle 2 et la condition pour la stationnarité.*

EXAMPLE 3.  $AR(2)$

$$(6) \quad Y_t = a_1Y_{t-1} + a_2Y_{t-2} + e_t$$

*en utilisant l'opérateur de décalage  $L$ , avec  $L(Y_t) = Y_{t-1}$  et en general  $L^k(Y_t) = Y_{t-k}$  alors 6 pourrait être écrit:*

$$(7) \quad (1 - a_1L - a_2L^2)Y_t = e_t$$

*par 7 on appelle polynôme caractéristique:*

$$(8) \quad 1 - a_1z - a_2z^2 = 0$$

*Pour que  $(1 - \phi_1L - \phi_2L^2)$  soit invertible, il faut placer les conditions sur les racines de 8. Après, changer le côté de 7*

$$(9) \quad \psi(L) = \frac{1}{1 - a_1L - a_2L^2}$$

$$(10) \quad Y_t = \psi(L)e_t$$

*Alors pour inverser le modèle AR à MA on a besoin de developper 9 aux puissants de  $L$  par la série géométrique,*

$$\frac{1}{1 - a_1L - a_2L^2} = 1 + a_1L + a_2L^2 + \dots$$

*Ainsi de 10*

$$(11) \quad Y_t = e_t + a_1e_{t-1} + a_2e_{t-2} + \dots$$

*Nous arrivons à la forme inversée infini. Pour preciser la forme des coefficients  $\psi_i$  à 11, soit  $r_1, r_2$  les racines du polynôme caractéristique et  $R_{1,2} = \frac{1}{r_{1,2}}$ , on a la factorisation*

$$\frac{1}{1 - a_1z - a_2z^2} = \frac{1}{(1 - R_1z)(1 - R_2z)} = \frac{A}{1 - R_1z} + \frac{B}{1 - R_2z}$$

$$A = \frac{r_2}{r_2 - r_1} \quad B = \frac{-r_1}{r_2 - r_1}$$

$$(12) \quad \frac{1}{1 - a_1 z - a_2 z^2} = \sum_{i=0}^{\infty} (AR_1^i + BR_2^i) z^i$$

C'est facile de vérifier par 12 que  $\psi_0 = 1$

Par cette dérivation, il est clair que on a besoin d' étudier les racines du polynôme caractéristique.

**Etude des racines du polynôme caractéristique**

Pour  $AR(2)$  d'être invertible et alors stationnaire il faut que ses racines sont hors du cercle unitaire. Nous montrerons quelques conditions pour les coefficients de 6 Nous savons que.

$$r_{1,2} = \frac{a_2 \sqrt{a_1 + 4a_2}}{-2a_2}$$

Et

$$R_1 = \frac{1}{r_2} = \frac{a_1 - \sqrt{a_1 + 4a_2}}{2} \quad R_2 = \frac{a_1 + \sqrt{a_1 + 4a_2}}{2}$$

Par 12 est claire qu'il faut  $|R_i| < 1$  for  $i = 1, 2$

: **Racines reels** Pour avoir deux solutions reels il faut  $a_1^2 + 4a_2 \geq 0$

$$-1 < \frac{a_1 - \sqrt{a_1 + 4a_2}}{2} < \frac{a_1 + \sqrt{a_1 + 4a_2}}{2} < 1$$

en faisant les calculs nécessaires on arrive à les conditions

$$a_2 - a_1 < 1 \quad \text{et} \quad a_2 + a_1 < 1$$

: **Racines complexes** Pour avoir deux solutions complexes il faut  $a_1^2 + 4a_2 < 0$  Ici, les  $R_{1,2}$  sont complex conjugates et

$$|R_1| = |R_2| < 1 \Leftrightarrow |R_1|^2 < 1$$

$$|R_1|^2 = \frac{a_1^2 + (-a_1^2 - 4a_2)}{4} = -a_2$$

Et alors  $|a_2| < 1$

PROPOSITION 1. On associe le polynôme de  $\mathbb{R}[X]$  suivant à l'équation (3)

$$A(X) = 1 - a_1 X - \dots - a_p X^p.$$

Si les racines (dans  $\mathbb{C}$ ) de ce polynôme sont toutes de module strictement supérieur à 1 alors il existe un processus stationnaire  $(X_t)$  vérifiant (3) et tel que  $\epsilon_t$  est le bruit d'innovation pour ce processus (c'est à dire que, pour tout  $t$ ,  $\epsilon_t$  est indépendant de  $X_{t-1}, X_{t-2}, \dots$ ).

RÉMARK 2. La condition " racines de  $A$  de module  $> 1$  " n'est pas nécessaire à l'existence d'une suite stationnaire vérifiant (3). Voir le chapitre 5 de l'édition en anglais de [?] pour plus de détails.

PROOF. On peut toujours développer la fraction rationnelle  $1/A(z)$  ( $z \in \mathbb{C}$ ) dans un voisinage de 0

$$\begin{aligned} \frac{1}{A(z)} &= \frac{1}{1 - a_1 z - \dots - a_p z^p} = 1 - (a_1 z + \dots + a_p z^p) - (a_1 z + \dots + a_p z^p)^2 - \dots \\ &= 1 + \alpha_1 z + \alpha_2 z^2 + \dots \end{aligned}$$

pour certains coefficients  $\alpha_1, \alpha_2, \dots$  (on remarque que ces coefficients sont dans  $\mathbb{R}$ ). Plus précisément, nous avons pour tout  $z$ ,

$$\frac{1}{A(z)} = \prod_{i=1}^p \left( \frac{1}{1 - u_i z} \right),$$

où l'ensemble des racines de  $A$  est  $\{1/u_1, \dots, 1/u_p\}$  (les racines apparaissent avec leur multiplicité). Donc, pour  $z$  tel que  $|z| < \inf_i(1/|u_i|)$ , nous pouvons développer en produit de séries entières

$$\frac{1}{A(z)} = \prod_{i=1}^p \left( \sum_{k=0}^{+\infty} u_i^k z^k \right).$$

Quand on développe le produit ci-dessus, le coefficient de  $z^n$  est

$$\alpha_n = \sum_{\substack{j_1, \dots, j_p \geq 0 \\ j_1 + \dots + j_p = n}} u_1^{j_1} \dots u_p^{j_p}.$$

Donc

$$(13) \quad |\alpha_n| \leq (\sup_i |u_i|)^n \frac{(n+1)^p}{p!}$$

(nous utilisons :  $\#\{(j_1, \dots, j_p) \in (\mathbb{N})^p : j_1 + \dots + j_p = n\} = \frac{(n+1)^p}{p!}$ , qui nécessite un petit calcul).

On suppose que l'on dispose d'un bruit blanc  $(\epsilon_k)_{k \in \mathbb{Z}}$ . On fixe  $t$  dans  $\mathbb{N}$  et on s'intéresse à la suite  $(Y_k = \epsilon_t + \alpha_1 \epsilon_{t-1} + \dots + \alpha_k \epsilon_{t-k})_{k \geq 1}$ . La série  $\sum_{n \geq 1} |\alpha_n|$  est convergente (petit exercice sur la convergence des séries, c'est ici que l'on utilise que  $\sup_i |u_i| < 1$ ) donc la série  $\sum_{n \geq 1} \alpha_n^2$  aussi. En particulier, pour tout  $\delta > 0$ , il existe  $N$  dans  $\mathbb{N}$  tel que

$$n \geq N \Rightarrow \forall k \geq 0, \alpha_n^2 + \alpha_{n+1}^2 + \dots + \alpha_{n+k}^2 < \delta.$$

Pour tout  $n > N$  et  $k \geq 0$ , nous avons alors

$$\begin{aligned} \mathbb{E}((Y_n - Y_{n+k})^2) &= \mathbb{E}((\alpha_{n+1} \epsilon_{t-n-1} + \dots + \alpha_{n+k} \epsilon_{t-n-k})^2) \\ &= (\alpha_{n+1}^2 + \dots + \alpha_{n+k}^2) \sigma^2 \\ &< \delta \end{aligned}$$

Donc la suite  $(Y_n)_{n \geq 1}$  est une suite de Cauchy dans  $L^2(\mathbb{P})$  (l'espace des variables aléatoires réelles de carré intégrable, muni de la norme  $L^2$ ). Comme cette espace est complet, la suite  $(Y_n)_{n \geq 1}$  converge dans  $L^2$  vers une limite que nous noterons  $X_t$ . Nous pouvons écrire

$$(14) \quad X_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots$$

(en considérant le terme de droite comme une limite dans  $L^2$ ).

Nous aimerions montrer que cette limite a aussi lieu p.s. Soit  $\delta > 0$ . For all  $n \geq 1$ ,

$$\begin{aligned} \mathbb{P}(|Y_n - X_t| \geq \delta) &= \mathbb{P}((\alpha_{n+1} \epsilon_{t-n-1} + \alpha_{n+2} \epsilon_{t-n-2} + \dots)^2 \geq \delta^2) \\ (\text{inégalité de Markov}) &\leq \frac{\mathbb{E}((\alpha_{n+1} \epsilon_{t-n-1} + \alpha_{n+2} \epsilon_{t-n-2} + \dots)^2)}{\delta^2} \\ &= \frac{\sigma^2(\alpha_{n+1}^2 + \alpha_{n+2}^2 + \dots)}{\delta^2}. \end{aligned}$$

Et donc

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}(|Y_n - X_t| \geq \delta) &\leq \sum_{n \geq 1} \frac{\sigma^2(\alpha_{n+1}^2 + \alpha_{n+2}^2 + \dots)}{\delta^2} \\ &= \frac{\sigma^2}{\delta^2} \sum_{n \geq 2} (n-1) \alpha_n^2 \\ &< \infty \end{aligned}$$

(petit exercices sur les séries à partir de l'équation (13)). Le lemme de Borel-Cantelli nous dit que, p.s., il existe un  $N$  tel que, pour  $n > N$ ,  $|Y_n - X_t| < \delta$ . Ceci est valable pour tout  $\delta > 0$ , donc (attention, nous sautons une étape de la démonstration)  $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X_t$ .

Montrons que les  $X_t$  que nous venons de définir vérifient la relation de récurrence (3). Nous avons, pour tout  $t$ ,

$$\begin{aligned} X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} &= \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots \\ &\quad - a_1 (\epsilon_{t-1} + \alpha_1 \epsilon_{t-2} + \alpha_2 \epsilon_{t-3} + \dots) \\ &\quad - \dots \\ &\quad - a_p (\epsilon_{t-p} + \alpha_1 \epsilon_{t-p-1} + \alpha_2 \epsilon_{t-p-2} + \dots) \\ \text{(convention } \alpha_0 = 1) &= \epsilon_t + \epsilon_{t-1} (\alpha_1 - a_1) + \epsilon_{t-2} (\alpha_2 - a_1 \alpha_1 - a_3) \\ &\quad + \dots \\ &\quad + \epsilon_{t-k} (\alpha_k - \sum_{1 \leq i \leq k \wedge p} \alpha_{k-i} a_i) \\ &\quad + \dots \end{aligned}$$

Nous avons, pour tout  $z$  dans un voisinage adéquat de 0,

$$\begin{aligned} 1 &= A(z) \times \frac{1}{A(z)} \\ &= (1 - a_1 z - \dots - a_p z^p) \times (1 + \alpha_1 z + \alpha_2 z^2 + \dots). \end{aligned}$$

Si nous développons ce dernier produit, nous trouvons une série entière dans laquelle le coefficient de  $z^n$  ( $n \geq 1$ ) est

$$\alpha_n - \sum_{1 \leq i \leq n \wedge p} \alpha_{n-i} a_i,$$

qui doit être nul puisque la série vaut 1. Donc

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \epsilon_t.$$

On montre facilement que la suite  $(X_t)_{t \geq 1}$  est stationnaire (exercice) (voir la définition ??).  $\square$

**PROPOSITION 2.** *S'il existe un processus stationnaire  $(X_t)_{t \geq 0}$  satisfaisant la relation de récurrence de l'équation (3) alors sa fonction d'auto-covariance vérifie :*

$$(15) \quad \text{pour } h > 0, \sigma(h) = a_1 \sigma(h-1) + a_2 \sigma(h-2) + \dots + a_p \sigma(h-p),$$

$$\text{et } \sigma(0) = \sigma^2 + a_1 \sigma(1) + a_2 \sigma(2) + \dots + a_p \sigma(p),$$

et sa fonction d'auto-corrélation vérifie

$$\text{pour } h > 1, \rho(h) = a_1 \rho(h-1) + a_2 \rho(h-2) + \dots + a_p \rho(h-p).$$

**PROOF.** Nous avons pour tout  $t$  et tout  $h \geq 1$ ,

$$\begin{aligned} \text{Cov}(X_t X_{t+h}) &= \mathbb{E}(X_t X_{t+h}) \\ \sigma(h) &= \mathbb{E}(X_t (\epsilon_{t+h} + \sum_{i=1}^p a_i X_{t+h-i})) \\ \sigma(h) &= 0 + \sum_{i=1}^p a_i \sigma(h-i) \end{aligned}$$

et

$$\begin{aligned}
\sigma(0) = \mathbb{V}(X_t) &= \mathbb{E}\left(X_t\left(\epsilon_t + \sum_{i=1}^p a_i X_{t-i}\right)\right) \\
&= \mathbb{E}(X_t \epsilon_t) + \sum_{i=1}^p a_i \sigma(i) \\
&= \mathbb{E}\left(\left(\epsilon_t + \sum_{i=1}^p a_i X_{t-i}\right)\epsilon_t\right) + \sum_{i=1}^p a_i \sigma(i) \\
(\text{car } \epsilon_t \text{ ind. de } X_j \text{ pour } j < t) &= \sigma^2 + \sum_{i=1}^p a_i \sigma(i).
\end{aligned}$$

□

Le polynôme caractéristique de la relation de récurrence de l'équation (15) est

$$\begin{aligned}
B(X) = X^p - a_1 X^{p-1} - \dots - a_p &= X^p \left(1 - \frac{a_1}{X} - \dots - \frac{a_p}{X^p}\right) \\
&= X^p A \left(\frac{1}{X}\right).
\end{aligned}$$

Donc, si les racines de  $A$  sont de modules  $> 1$ , alors celle de  $B$  sont de module  $< 1$ . Si les racines de  $B$  sont distinctes égales à  $\lambda_1, \lambda_2, \dots, \lambda_p$ , les solutions de (15) sont de la forme

$$\sigma(h) = \sum_{i=1}^p c_i \lambda_i^h$$

(voir [?], p. 196 pour plus de détails). On en déduit le résultat suivant.

**LEMMA 2.** *S'il existe un processus stationnaire  $(X_t)_{t \geq 0}$  satisfaisant la relation de récurrence de l'équation (3) et si le polynôme de la relation de récurrence a des racines qui sont toutes de module  $> 1$ , alors la fonction d'auto-covariance du processus décroît exponentiellement quand  $h \rightarrow +\infty$  (et donc sa fonction d'auto-corrélation a le même comportement).*

**PROPOSITION 3.** *S'il existe un processus stationnaire  $(X_t)_{t \geq 0}$  satisfaisant la relation de récurrence de l'équation (3) alors sa fonction d'auto-corrélation partielle vérifie :*

$$r(h) = 0 \text{ si } h \geq p + 1.$$

**PROOF.** Soit  $h \geq p + 1$ . Nous calculons pour  $t$  quelconque :

$$\begin{aligned}
\mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}) &= \mathbb{E}(\epsilon_{t+h} + a_1 X_{t+h-1} + \dots + a_p X_{t+h-p} | X_{t+1}, \dots, X_{t+h-1}) \\
&= a_1 X_{t+h-1} + \dots + a_p X_{t+h-p}, \\
X_{t+h} - \mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}) &= \epsilon_{t+h},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1}) &= \\
\mathbb{E}\left(-\frac{1}{a_p}(X_{t+p} - \epsilon_{t+p} - a_1 X_{t+p-1} - \dots - a_{p-1} X_{t+1}) | X_{t+1}, \dots, X_{t+h-1}\right) &= \\
-\frac{1}{a_p}(X_{t+p} - a_1 X_{t+p-1} - \dots - a_{p-1} X_{t+1}), & \\
X_t - \mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1}) &= \frac{\epsilon_{t+p}}{a_p}.
\end{aligned}$$

Donc

$$\text{Cov}(X_{t+h} - \mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}), X_t - \mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1})) = 0.$$

□

#### 4. Les processus en moyenne mobile

DÉFINITION 3. Un processus en moyenne mobile d'ordre  $q$  est un processus de la forme

$$(16) \quad X_t = \epsilon_t + b_1\epsilon_{t-1} + \cdots + b_q\epsilon_{t-q}, \text{ pour tout } t \geq 0$$

(avec  $q \in \mathbb{N}^*$ ,  $b_q \neq 0$ ) avec des  $(\epsilon_t)$  qui forment un bruit blanc centré de variance  $\sigma^2$ . Par convention :  $\epsilon_{-k} = 0$  pour tout  $k$  dans  $\mathbb{N}^*$ . On dira aussi que  $(X_t)$  est un processus  $MA(q)$ .

Pour un bruit blanc centré  $(\epsilon_t)$ , un tel processus existe toujours et est toujours stationnaire.

PROPOSITION 4. L'auto-covariance d'un processus  $MA(q)$  vérifiant l'équation (16) vérifie, pour  $h \geq 0$ ,

$$\sigma(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} b_k b_{k+h} & \text{si } h \leq q, \\ 0 & \text{si } h > q, \end{cases}$$

sous la convention  $b_0 = 1$ .

PROOF. Nous calculons

$$\begin{aligned} \mathbb{E}(X_t X_{t+h}) &= \mathbb{E} \left( \left( \sum_{k=0}^q \epsilon_{t-k} b_k \right) \times \left( \sum_{k=0}^q \epsilon_{t+h-k} b_k \right) \right) \\ &= \begin{cases} 0 & \text{si } h > q, \\ \sum_{k=q}^{q-h} \mathbb{E}(\epsilon_{t-k} b_k \epsilon_{t-k} b_{k+h}) & \text{sinon.} \end{cases} \end{aligned}$$

□

De la proposition 1 et de l'équation (14), nous déduisons le corollaire suivant.

COROLLAIRE 1. Sous les hypothèses de la proposition 1, le processus  $AR(p)$  stationnaire vérifiant l'équation de récurrence (3) peut s'écrire comme un processus  $AM$  d'ordre infini (au sens de l'équation (14) (dans laquelle la limite a lieu p.s. et dans  $L^2$ ).

La démonstration du résultat suivant étant similaire à la démonstration de la proposition 1, nous le citons comme un simple corollaire.

COROLLAIRE 2. Si  $(X_t)_{t \geq 0}$  est un processus  $AM(q)$  satisfaisant l'équation (16) et si le polynôme  $B(X) = 1 + b_1X + \cdots + b_qX^q$  n'admet que des racines de module  $> 1$ , alors  $(X_t)$  peut s'écrire comme un processus  $AR$  d'ordre infini et tel que  $\epsilon_t$  est le bruit d'innovation pour ce processus.

En particulier, sa fonction d'auto-corrélation partielle vérifie :

$$r(h) \xrightarrow{h \rightarrow +\infty} 0.$$

En effet, en partant de l'équation (16), on peut exprimer  $\epsilon_t$  en fonction de  $X_t, X_{t-1}, \dots$  pour tout  $t$  (de même qu'en partant de l'équation (3), on peut exprimer  $X_t$  en fonction de  $\epsilon_t, \epsilon_{t-1}, \dots$  pour tout  $t$ ). Pour plus de détail voir [?] (ces questions sont développées au chapitre 5 de l'édition en anglais de ce livre).

### 5. Les processus mixtes ARMA(p,q).

DÉFINITION 4. Un processus auto-régressif en moyenne mobile d'ordres  $p, q$  (tels que  $p \times q \neq 0$ ) est un processus qui peut s'écrire

$$(17) \quad X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=0}^q b_j \epsilon_{t-j}, \text{ pour tout } t \geq 0,$$

où les  $(\epsilon_j)$  sont des bruits blancs centrés de variance  $\sigma^2$ . Par convention :  $X_{-k} = \epsilon_{-k}$  pour tout  $k$  dans  $\mathbb{N}^*$ . On dira aussi que  $(X_t)$  est un processus ARMA(p,q).

On peut toujours écrire

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q},$$

d'où la relation

$$\begin{aligned} \text{Cov}(X_{t+h} - a_1 X_{t+h-1} - \dots - a_p X_{t+h-p}, X_t) &= \sigma(h) - a_1 \sigma(h-1) - \dots - a_p \sigma(h-p) \\ &= \text{Cov}(\epsilon_{t+h} + b_1 \epsilon_{t+h-1} + \dots + b_q \epsilon_{t+h-q}, X_t). \end{aligned}$$

Cette dernière quantité est nulle dès que  $h > q$ . Les auto-covariances d'un processus ARMA(p,q) vérifie donc la même relation de récurrence que celles d'un processus AR(p) à partir d'un certain rang. Donc ces auto-covariances (ainsi que les auto-corrélations correspondantes) convergent exponentiellement vite vers 0 à l'infini.

PROPOSITION 5. On s'intéresse à la relation de récurrence (17) ci-dessus. On définit les polynômes

$$A(X) = 1 - a_1 X - \dots - a_p X^p, \quad B(X) = 1 + b_1 X + \dots + b_q X^q.$$

Si les racines de  $A$  et  $B$  sont de modules  $> 1$ , alors il existe un processus stationnaire vérifiant la relation (17) et tel que  $\epsilon_t$  est le bruit d'innovation pour ce processus.

Si les racines de  $A$  et  $B$  sont de module  $> 1$  alors : "  $(X_t)$  ne vérifie pas de relation de récurrence plus courte que (17) " implique "  $A$  et  $B$  n'ont pas de racine commune ". Si on suppose, de plus, que les  $(\epsilon_t)$  ont des densités par rapport à la mesure de Lebesgue, alors les deux propositions sont équivalentes.

RÉMARK 3. Pour plus de détail voir [?] (ces questions sont développées au chapitre 5 de l'édition en anglais de ce livre).

DÉMONSTRATION PARTIELLE DE LA PROPOSITION 5. Supposons que  $A$  et  $B$  n'ont que des racines de module  $> 1$ . Supposons, de plus, que  $A$  et  $B$  ont une racine commune, nous allons montrer que  $(X_t)$  et  $(\epsilon_t)$  vérifie une relation plus courte que (17). Notons  $1/\lambda$  une racine commune à  $A$  et  $B$ . Nous avons donc  $p \geq 1$  et  $q \geq 1$ . Nous pouvons écrire

$$A(X) = (1 - \lambda X)A_1(X), \quad B(X) = (1 - \lambda X)B_1(X)$$

avec  $A_1$  polynôme de degré  $p - 1$  et  $B_1$  polynôme de degré  $q - 1$ . Nous avons

$$\begin{aligned} A(L)X &= B(L)\epsilon \\ (1 - \lambda L)A_1(L)X &= (1 - \lambda L)B_1(L)\epsilon \\ (1 - \lambda L)^{-1}(1 - \lambda L)A_1(L)X &= (1 - \lambda L)^{-1}(1 - \lambda L)B_1(L)\epsilon \\ A_1(L)(X) &= B_1(L)\epsilon. \end{aligned}$$

□

EXAMPLE 4. Soient

$$A(X) = \left(1 - \frac{X}{2}\right) \left(1 - \frac{X}{3}\right), \quad B(X) = \left(1 - \frac{X}{2}\right) \left(1 + \frac{X}{4}\right).$$

Nous avons

$$\begin{aligned} A(X) &= 1 - \frac{5}{6}X + \frac{X^2}{6}, \\ B(X) &= 1 - \frac{X}{4} - \frac{X^2}{8}. \end{aligned}$$

On s'intéresse au processus ARMA(2,2) (stationnaire) vérifiant la relation de récurrence :

$$(18) \quad X_t - \frac{5}{6}X_{t-1} + \frac{1}{6}X_{t-2} = \epsilon_t - \frac{1}{4}\epsilon_{t-1} - \frac{1}{8}\epsilon_{t-2}.$$

Les polynômes caractéristiques de cette relation sont les  $A, B$  ci-dessus. Ils ont une racine commune (2). Nous allons montrer que  $(X_t), (\epsilon_t)$  vérifient une relation de récurrence plus courte que (18). Soit la suite

$$\begin{aligned} (Y_0, Y_1, Y_2, \dots, Y_t, \dots) &= (X_0, X_1 - \frac{5}{6}X_0, X_2 - \frac{5}{6}X_1 + \frac{1}{6}X_0, \dots, X_t - \frac{5}{6}X_{t-1} + \frac{1}{6}X_{t-2}, \dots) \\ &= (\epsilon_0, \epsilon_1 - \frac{1}{4}\epsilon_0, \epsilon_2 - \frac{1}{4}\epsilon_1 - \frac{1}{8}\epsilon_0, \dots, \epsilon_t - \frac{1}{4}\epsilon_{t-1} - \frac{1}{8}\epsilon_{t-2}, \dots). \end{aligned}$$

Par convention,  $X_{-k} = \epsilon_{-k} = 0$  pour tout  $k$  dans  $\mathbb{N}^*$ . Et donc :

$$X_0 = X_0 - \frac{5}{6}X_{-1} + \frac{1}{6}X_{-2}, \quad X_1 = X_1 - \frac{5}{6}X_0 + \frac{1}{6}X_{-1}.$$

Calculons maintenant

$$(Y_0, Y_1 + \frac{1}{2}Y_0, Y_2 + \frac{1}{2}Y_1 + \frac{1}{2^2}Y_0, \dots, Y_t + \frac{1}{2}Y_{t-1} + \dots + \frac{1}{2^t}Y_0, \dots).$$

Le terme général de cette suite est

$$Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 = \sum_{k=0}^{+\infty} \frac{1}{2^k} Y_{t-k},$$

puisque les termes de la somme sont nuls à parti d'un certain rang. Donc, pour tout  $t \geq 0$  :

$$\begin{aligned} Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 &= \sum_{k=0}^{+\infty} \frac{1}{2^k} \left( X_{t-k} - \frac{5}{6}X_{t-k-1} + \frac{1}{6}X_{t-k-2} \right) \\ &= X_t + \left( -\frac{5}{6}X_{t-1} + \frac{1}{2}X_{t-1} \right) + \sum_{j=2}^{+\infty} X_{t-j} \left( \frac{1}{2^j} - \frac{1}{2^{j-1}} \frac{5}{6} + \frac{1}{2^{j-2}} \frac{1}{6} \right) \\ &= X_t - \frac{1}{3}X_{t-1} + \sum_{j=2}^{+\infty} X_{t-j} \frac{1}{2^j} A(2) \\ &= X_t - \frac{1}{3}X_{t-1}. \end{aligned}$$

De même :

$$\begin{aligned} Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 &= \sum_{k=0}^{+\infty} \frac{1}{2^k} \left( \epsilon_{t-k} - \frac{1}{4}\epsilon_{t-k-1} - \frac{1}{8}\epsilon_{t-k-2} \right) \\ &= \epsilon_t + \left( -\frac{1}{4}\epsilon_{t-1} + \frac{1}{2}\epsilon_{t-1} \right) + \sum_{j=2}^{+\infty} \epsilon_{t-j} \left( \frac{1}{2^j} - \frac{1}{2^{j-1}} \frac{1}{4} - \frac{1}{2^{j-2}} \frac{1}{8} \right) \\ &= \epsilon_t + \frac{1}{4}\epsilon_{t-1} + \sum_{j=2}^{+\infty} \epsilon_{t-j} \frac{1}{2^j} B(2) \\ &= \epsilon_t + \frac{1}{4}\epsilon_{t-1}. \end{aligned}$$

Nous avons donc la relation :

$$X_t - \frac{1}{3}X_{t-1} = \epsilon_t + \frac{1}{4}\epsilon_{t-1}, \text{ pour tout } t \in \mathbb{N}^* .$$

## 6. Tableau des propriétés

Modèle	$MA(q)$	$AR(p)$	$ARMA(p, q)$
auto-corrélation	$\rho(h) = 0$ si $h > q$	$\rho(h) \xrightarrow{h \rightarrow +\infty} 0$	$\rho(h) \xrightarrow{h \rightarrow +\infty} 0$
auto-corrélation partielle	$r(h) \xrightarrow{h \rightarrow +\infty} 0$	$r(h) = 0$ si $h > p$	$r(h) \xrightarrow{h \rightarrow +\infty} 0$

TABLE 1. Tableau des propriétés

Voir à la fin du chapitre pour des illustrations de ce tableau (pages tirées du polycopié [?]).

Ces propriétés servent à identifier la nature des séries temporelles. Sous **R**, on utilisera les fonctions **acf**, **pacf** qui tracent, respectivement, les  $\hat{\rho}(h)$  et les  $\hat{r}(h)$  (les auto-corrélations empiriques et les auto-corrélations partielles empiriques). Le logiciel trace en plus un niveau bleu horizontal en  $y = m_\alpha$  tel que pour tout  $h$ ,  $\mathbb{P}(|\hat{\rho}(h)| \geq m_\alpha | \rho(h) = 0) = \alpha$  (en général, le niveau  $\alpha$  est fixé à 0,05) (la situation est la même pour les auto-corrélations partielles empiriques). On peut fixer  $\alpha$  en ajoutant l'option : **acf(..., ci=alpha)**. Un  $\hat{\rho}(h)$  sous la courbe bleue est donc non significatif (au niveau  $\alpha$ ). Le raisonnement est le suivant : on suppose  $\rho(h) = 0$ , si  $|\hat{\rho}(h)| < m_\alpha$ , on considère qu'il n'est pas nécessaire de revenir sur cette hypothèse de départ. La probabilité de rejeter à tort l'hypothèse " nulle " ( $\rho(h) = 0$ ) est  $\alpha$ .

Mais quand on veut utiliser le tableau ci-dessus, on cherche à savoir à partir de quel indice les  $\rho(h)$  sont nuls (par exemple), pas si l'un d'eux est nul<sup>1</sup>. Si on s'intéresse à la nullité de  $(\rho(n), \rho(n+1), \dots, \rho(n+l-1))$ , on pourrait vouloir trouver  $\beta$  tel que

$$\mathbb{P}(\exists h \in \{n, n+1, \dots, n+l-1\}, |\hat{\rho}(h)| \geq m_\beta | (\rho(n), \dots, \rho(n+l-1)) = (0, \dots, 0)) \leq \alpha .$$

Les  $\hat{\rho}(h)$  sont supposés indépendants donc la probabilité ci-dessus est

$$\begin{aligned} 1 - \mathbb{P}(\forall h, |\hat{\rho}(h)| < m_\beta | \dots) &= 1 - (1 - \beta)^l \\ &= l\beta + o(\beta) \end{aligned}$$

(quand  $\beta$  est petit). Donc on prend, en général,  $\beta = \alpha/l$ .

Si on appelle les fonctions **acf** ou **pacf** sans préciser le **lag.max**, le logiciel le fixe par défaut à  $10 \log_{10}(n)$  (où  $n$  est la longueur de la série). La raison est que l'on ne veut pas prendre le **lag.max** trop grand parce que les moyennes empiriques ne convergent pas bien pour  $h$  grand. On conseille en général de prendre  $n \geq 50$  et  $h \leq n/4$  (voir [?], p. 60 et [?], p. 32).

## 7. Les modèles des séries non-stationnaires

N'importe quand une série n'a pas moyenne constant, elle est non-stationnaire. Pour une série d'être stationnaire il faut que son moyen et sa covariance ne dépend pas au temps  $t$ . Nous commençons par quelques exemples des processus non-stationnaires. Dans toutes les exemples essayez vous-mêmes de calculer l'espérance, covariance etc pour devenir à l'aise avec les sommes infinies et calculs.

<sup>1</sup>Voir <https://xkcd.com/882/> sur le problème des tests multiples.

EXAMPLE 5. Soit un processus  $X_t$ ,  $\mathbb{E}[X_t] = 0$  et une fonction de moyenne non-constant  $\mu_t$ ,  $\mathbb{E}[\mu_t] = c(t)$

$$Y_t = \mu_t + X_t$$

est une série non-stationnaire. On peut aller plus loin dans cette "catégorie" des modèles non-stationnaires en écrivant l'exemple à la forme

$$(19) \quad Y_t = a + \delta t + \psi(L)e_t$$

$\psi(L)$  comme à 9, une partie de la série a une tendance linéaire et l'autre est stationnaire.

**Devoir 1**

Calculer : l'espérance, la variance, la covariance et la corrélation de la processus

$$Y_t = a + \delta t + e_t$$

EXAMPLE 6. Soit un processus pour lequel le condition de la stationnarité est violé.

$$Y_t = 3Y_{t-1} + e_t$$

En transformant la série en fonction des bruits précédents

$$Y_t = e_t + 3e_{t-1} + 3^2e_{t-2} + \dots + 3^{k-1}e_1 + 3^kY_0$$

On remarque que les coefficients ne se tend pas vers zéro.

EXAMPLE 7. Marche Aléatoire avec dérive

$$(20) \quad Y_t = Y_{t-1} + \delta + e_t$$

Ici,  $\delta \in \mathbb{R}$  est la dérive, si  $\delta = 0$  on reviens à la cas classique d'une marche aléatoire.

**Devoir 2**

Calculer : l'espérance, la variance, la covariance et la corrélation du marche aléatoire avec dérive. Vérifier qu'il s'agit d'un processus non-stationnaire.

Toutes les exemples montrent différents cas de non-stationnarité et toujours ) En gross, on pourrais identifier deux catégories de pathologie: soit que il y a En général, dans une série non-stationnaire, l'influence des valeurs précédents augmente. On peut appeler cette comportement explosif.

**7.1. Les processus ARIMA.** Dans cette section nous nous occupons avec les cas où la source de la non-stationnarité est la violation des conditions du polynôme caractéristique comme il était montré par les exemples précédents. Pour un processus  $AR(1)$  il faut que la racine du polynôme soit entre 1, -1 strictement. On peut écrire l'exemple 3 en utilisant l'opérateur de décalage du temps (lag) comme

$$(1 - L)Y_t = \delta + e_t$$

Ici, il est clair qu'il s'agit d'une racine unitaire, et alors ce n'est pas possible de tourner le modèle en forme rétroactif convergente. Mais si on prendre les premières différences on voit que

$$(21) \quad \Delta Y_t = Y_t - Y_{t-1} = \delta + e_t$$

qui donne une série stationnaire. Le même argument s'applique aussi à la cas générale de la forme

$$(22) \quad (1 - L)Y_t = \delta + \psi(L)e_t$$

et on appelle cette catégorie des processus non-stationnaires, *processus de racine unitaire (unit root processes)* et le processus qui sort des différences, *processus intégré d'ordre 1, I(1)*

On peut aussi imaginer un processus ARMA qui contient les racines unitaires en multiple fois.

$$(23) \quad (1-L)^k(1-R_1L)(1-R_2L)\cdots(1-R_{p-k}L)Y_t = \psi(L)e_t$$

ou  $R_1, \dots, R_{p-k}$  sont les reciprocaux des racines différents, et la multitude de racine unitaire est  $k$ . On s'appelle alors 23 un processus  $ARIMA(p, d, q)$  avec  $d = k$  et  $q$  dépendant à la représentation du  $\psi(L)$

Une autre catégorie de pathologie qui peut rendre une série non-stationnaire est la tendance linéaire, comme à l'exemple 5 et l'équation 19.

Pour soulager la non-stationnarité on prend les différences en appliquant l'opérateur des difference  $\Delta$  autant des fois il est nécessaire pour rendre la série stationnaire. Il faut prêter attention d'être économe en appliquant l'opérateur car trop des différences créent des problèmes au partie MA, la raison précis est hors du sujet de ces notes Pour la même raison, si on soupçonne une tendance linaire, il est conseillé d'essayer l'enlever avant prendre les premiers différences.

### 7.2. IMA(1,1).

$$(24) \quad Y_t = Y_{t-1} + e_t - \theta e_{t-1}$$

Ici, c'est un modèle avec une racine unitaire et un lag au bruit, alors  $ARIMA(0, 1, 1)$  Pour le rendre stationnaire nous prenons les premières différences

$$W_t = \Delta Y_t = e_t - \theta e_{t-1}$$

Maintenant,  $W_t$  est un processus  $MA(1)$  que nous avons déjà étudié

Équation 24 pourrait être transformé en forme rétroactif

$$(25) \quad Y_t = e_t + (1-\theta)e_{t-1} + \cdots + (1-\theta)e_{t-k+1} + \theta e_{t-k}$$

#### Devoir

Avec l'aide de 25 montrez

- $\mathbb{E}[Y_t] = 0$
- $Var(Y_t) = [1 + \theta^2 + (1-\theta)^2(t+k)]\sigma_e^2$
- $Cov(Y_t, Y_{t-h}) = 1 + \theta^2 + (1-\theta)^2(t+k-h)$

### 7.3. ARI(1,1).

$$(26) \quad Y_t = (1-\phi)Y_{t-1} - \phi Y_{t-2} + e_t$$

ou

$$W_t = \Delta Y_t = \phi W_{t-1} + e_t$$

Maintenant,  $W_t$  est un processus  $AR(1)$  comme aux chapitres précédents. Si nous voulons écrire le polynôme characteristic pour cette processus, nous arrivons à la formule

$$\frac{1}{(1-z)(1-\phi z)} = 1 + \psi_1 z + \psi_2 z^2 + \cdots$$

ou

$$(27) \quad (1 - (1 + \phi)z + \phi z^2)(1 + \psi_1 z + \psi_2 z^2 + \cdots) = 1$$

pour définir les  $\psi_i$  nous arrivons au system

$$\begin{aligned} -(1 + \phi) + \psi_1 &= 0 \\ \phi - (1 + \phi)\psi_1 + \psi_2 &= 0 \end{aligned}$$

et

$$(28) \quad \psi_k = (1 + \phi)\psi_{k-1} - \phi\psi_{k-2}, \quad \psi_0 = 1$$

**Devoir** Calculer la variance, la covariance et la correlation du processus ARI(1,1)

## 8. Specification du Modèle, la méthodologie Box-Jenkins

Après l'étude théorique des modèles ARMA et ARIMA maintenant nous nous occupons avec les questions pratiques.

- Comment modéliser les données comme un processus  $ARIMA(p, d, q)$ ?
- Est-ce que la série est stationnaire?
- Comment choisir  $p, d, q$ ?

Pour répondre à ces questions à la manière pratique et pragmatique, Box et Jenkins<sup>2</sup> ont développé une methodology en 4 étapes:

- (1) Identification
- (2) Estimation
- (3) Diagnostique
- (4) Prévision

**8.1. Identification.** En cet étape on s'occupe de déterminer l'ordre du modèle, et le comportement stationnaire (ou non) et saisonnier (ou non). Il y a deux approches pour traiter cette partie

- L'approche graphique: On plot la série, les ACF, PACF en cas que ACF tend vers zéro très lentement ou pas du tout on soupçonne nonstationnarité. Box et Jenkins conseillent à prendre les premières différences et refaire les graphes en cas qu'ils indiquent stationnarité. Sinon, on reprendre les différences jusqu'à le graphe a l'air stationnaire.

Pour déterminer l'ordre du modèle  $p, q$  il y a quelques règles empiriques à partir de graph de ACF

- (1) Décroissance exponentielle vers zéro : Modèle AR (PACF indique l'ordre  $p$ )
- (2) Oscillations amorties décroissantes rapidement vers zéro: AR (PACF indique l'ordre  $p$ )
- (3) Quelques pics, le reste nul: MA (l'ordre est le nombre des pics)
- (4) Décroissance exponentielle après quelques pas de temps (lags)
- (5) pattern régulière en écarts de temps : saisonnalité

Une remarque générale, ACF nous aide à déterminer l'ordre MA du modèle et PACF l'ordre AR.

- L'approche statistique. Il y a quelques tests statistiques que on peut essayer pour reconnaître la stationnarité ou non de notre modèle, voir la prochaine section. Une fois que la question de stationnarité sera répondu on tourne sur la détermination d'ordre AR ou MA par la voie statistique. On choisit entre plusieurs modèles en regardant :
  - l'ajustement à la série de données,
  - la complexité du modèle (il est plus facile d'estimer un nombre réduit de paramètres).

Pour concilier ces deux critères, on minimise une des deux quantités suivantes

$$AIC = -2\log(L(\theta)) + 2\nu,$$

$$BIC = -2\log(L(\theta)) + n\log(\nu),$$

où  $\nu$  est le nombre de paramètres ( $n = p + q$  si l'on s'agit d'un ARMA sans constant ou  $n = p + q + 1$ ),  $\theta$  est un vecteur contenant les paramètres,  $n$  est le nombre d'observations,  $L$  est la vraisemblance (dans laquelle on a omis les observations).

<sup>2</sup>G. Box, G. Jenkins, *Time Series Analysis: forecasting and control*, 1970

AIC (Akaike's Information Criterion) est un estimateur d'espérance de la divergence de Kullback-Leibler entre le modèle estimé et le vrai modèle. Soit  $p(y_1, \dots, y_n)$  la vraie densité de probabilité de  $Y_1, \dots, Y_n$  et soit  $q_\theta(y_1, \dots, y_n)$  la densité sous le modèle de paramètre  $\theta$  la divergence de Kullback-Leibler de  $q_\theta$  à  $p$  est donnée par la formule suivante

$$D(p, q_\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(y_1, \dots, y_n) \log \left( \frac{p(y_1, \dots, y_n)}{q_\theta(y_1, \dots, y_n)} \right) dy_1 \dots dy_n$$

On veut estimer l'espérance  $\mathbb{E}[D(p, q_{\hat{\theta}})]$  où  $\hat{\theta}$  est l'estimateur du vecteur des paramètres, qui sort de la formule du début.

**EXEMPLE 8.** *Le calcul des quantités AIC et BIC se fait facilement en R. Calculons une vraisemblance dans un cas simple.*

*On suppose  $X_t = at + b + \epsilon_t$  avec des  $\epsilon_t$  i.i.d. de loi  $\mathcal{N}(0; 1)$ . Les paramètres du modèle sont  $a$  et  $b$ . On dispose d'observations  $x_1, \dots, x_n$ . La densité de  $(X_1, \dots, X_n)$  est la fonction*

$$(u_1, \dots, u_n) \mapsto \prod_{t=1}^n \left\{ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (u_t - at - b)^2 \right) \right\}.$$

*La vraisemblance est donc*

$$(a', b') \mapsto L(a', b') = L((a', b'); (x_1, \dots, x_n)) = \prod_{t=1}^n \left\{ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (x_t - a't - b')^2 \right) \right\}.$$

*Si on veut estimer  $(a, b)$  à partir de  $(x_1, \dots, x_n)$ , l'estimateur du maximum de vraisemblance est*

$$(\hat{a}, \hat{b}) = \arg \max_{a', b'} L(a', b').$$

Pour la saisonnalité, nous allons apprendre plus au chapitre 5

**8.2. Estimation du modèle.** On suppose que  $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t$  (avec des  $\epsilon_t$  bruits blancs centrés, de variance  $\sigma^2$ ) et que  $(X_t)$  est stationnaire. On calcule les auto-covariances :

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

D'où

$$\begin{aligned} a_1 &= \frac{\sigma(1)}{\sigma(0)} \times \frac{\sigma(0)^2 - \sigma(0)\sigma(1)}{\sigma(0)^2 - \sigma(1)^2}, \\ a_2 &= \frac{\sigma(0)\sigma(2) - \sigma(1)^2}{\sigma(0)^2 - \sigma(1)^2}. \end{aligned}$$

On peut donc estimer  $a_1, a_2$  en remplaçant  $\sigma(0), \sigma(1), \sigma(2)$  par leurs estimateurs empiriques dans les formules ci-dessus. Puisque  $\sigma(0) = \sigma^2 + a_1 \sigma(1) + a_2 \sigma(2)$ , on peut aussi estimer  $\sigma^2$ .

Dans le cas général, on estime par maximum de vraisemblance (voir cours de statistiques et l'exemple ci-dessous).

**8.3. Diagnostique du modèle.** Ici, nous nous intéressons principalement au comportement des résidus du modèle. Quelles sont les propriétés statistiques des résidus? Hypothèse de normalité, bruit blanc etc. Box-Pierce ou Ljung-Box sont les outils principaux ici. Deux remarques importantes

- (1) En testant la corrélation des résidus, il n'est possible que de dévoiler un modèle sousparamétrisé. Il y a rien à dire pour un modèle surparamétrisé.
- (2) Autocorrélation des résidus est possible de faire les autres tests statistiques donner les faux résultats.

**8.4. Prédiction.** Si le modèle est un processus *ARMA*, la prédiction pour  $X_{n+h}$ , sachant  $X_1, \dots, X_n$  est

$$\widehat{X}_{n,h} = c_1 X_1 + \dots + c_n X_n,$$

où les coefficients sont choisis de manière à minimiser l'erreur quadratique

$$\mathbb{E}((X_{n+h} - c_1 X_1 - \dots - c_n X_n)^2).$$

PROPOSITION 6. *Ce choix de  $\widehat{X}_{n,h}$  entraîne l'égalité*

$$\widehat{X}_{n,h} = \mathbb{E}(X_{n+h} | X_1, \dots, X_n).$$

Ce qui n'est pas surprenant si on se rappelle que l'espérance conditionnelle de  $X_{n+h}$  sachant  $X_1, \dots, X_n$  est la projection orthogonale de  $X_{n+h}$  sur  $\sigma(X_1, \dots, X_n)$  dans l'espace des variables  $L^2$  (muni de la norme  $L^2$ ).

PROPOSITION 7. *L'erreur de prédiction à l'horizon 1 ( $X_{n+1} - \widehat{X}_{n,1}$ ) est le bruit d'innovation  $\epsilon_{n+1}$ .*

*La variance de l'erreur de prédiction ( $\mathbb{E}((X_{n+h} - \widehat{X}_{n,h})^2)$ ) est croissante avec  $h$  et tend vers  $\mathbb{V}(X_1)$  quand  $h \rightarrow +\infty$  (on rappelle que le processus  $(X_t)$  est supposé stationnaire).*

Puisque les  $\epsilon_t$  sont gaussiens, les  $X_t$ ,  $\widehat{X}_{n,h}$  et  $\widehat{X}_{n,h} - X_{n+h}$  sont aussi gaussiens (nous sautons une petite démonstration). Ceci permet de construire facilement des intervalles de confiance (c'est inclus dans **R**). Nous faisons ici un rappel dans un cas simple. Supposons  $\widehat{X}_{n,h} - X_{n+h} \sim \mathcal{N}(0; \sigma^2)$ . Soit  $\alpha = 0,01$ . Nous cherchons  $\Delta$  tel que

$$\mathbb{P}(X_{n+h} \in [\widehat{X}_{n,h} - \Delta; \widehat{X}_{n,h} + \Delta]) \geq 1 - \alpha$$

(ici, toutes les probabilités sont conditionnelles à  $X_1, \dots, X_n$ ). Nous calculons

$$\begin{aligned} \mathbb{P}(X_{n+h} \in [\widehat{X}_{n,h} - \Delta; \widehat{X}_{n,h} + \Delta]) &= \mathbb{P}(|X_{n+h} - \widehat{X}_{n,h}| \leq \Delta) \\ &= \mathbb{P}\left(\frac{|X_{n+h} - \widehat{X}_{n,h}|}{\sigma} \leq \frac{\Delta}{\sigma}\right) \end{aligned}$$

$$\text{(en utilisant les symétries de la gaussienne)} = 1 - 2\mathbb{P}\left(\frac{X_{n+h} - \widehat{X}_{n,h}}{\sigma} \geq \frac{\Delta}{\sigma}\right).$$

Nous voulons donc  $\Delta$  tel que

$$\mathbb{P}\left(\frac{X_{n+h} - \widehat{X}_{n,h}}{\sigma} \geq \frac{\Delta}{\sigma}\right) \leq \frac{\alpha}{2} = 0,005.$$

Puisque  $(X_{n,h} - \widehat{X}_{n,h})/\sigma$  est de loi  $\mathcal{N}(0; 1)$ , on lit dans une table de la loi normale qu'il suffit de prendre  $\Delta/\sigma = 2,58$  pour que l'inégalité ci-dessus soit vérifiée.

**8.5. Tests statistiques pour la stationnarité.** Comme nous avons vu si le polynôme d'une série contient une racine unitaire, la série n'est pas stationnaire. Dès premières étapes de la méthodologie Box-Jenkins on peut soupçonner la non-stationnarité par les graphes mais ici notre but c'est de développer des tests statistiques pour tester rigoureusement les séries. Souvent il y a deux cas de non-stationnarité et ce n'est pas toujours facile à les distinguer, trend-stationnaire (c'est-à-dire la série n'est pas stationnaire par rapporte de définition mais elle est stationnaire autour une tendance deterministic) et racine unitaire non-stationnaire. Les tests qui suivent sont développés pour la deuxième catégorie.

8.5.1. *Augmented Dickey-Fuller test de racine unitaire.***: Dickey-Fuller test**

$$(29) \quad Y_t = aY_{t-1} + e_t$$

Avec les hypothèse:

$$H_0 : a = 1$$

$$H_1 : a < 1$$

sous l'hypothèse  $H_0$  nous avons un marche aléatoire. DF-test est le t-test de  $H_0$  et on construit la statistique pour l'estimateur  $\hat{a}$  de  $a$

$$\hat{\tau} = \frac{\hat{a} - 1}{se(\hat{a})}$$

$\hat{\tau}$  ne suis pas Gaussian et la distribution analytic pourrait être rétablie par Hamilton p489 et table B6 *Remarques*

(1)  $e_t$  doit être bruit blanc

(2) La distribution de  $\hat{\tau}$  reste la même aussi pour un modèle ARMA

**: Augmented Dickey-Fuller test**

$$(30) \quad \begin{aligned} Y_t &= aY_{t-1} + X_t \\ X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_k X_{t-k} + e_t \end{aligned}$$

Avec les hypothèse:

$$H_0 : a = 1$$

$$H_1 : a < 1$$

Sous  $H_0$  nous avons  $\Delta Y_t$  d'être un processus stationnaire  $AR(k)$

$$Y_t - Y_{t-1} = X_t$$

et  $Y_t$  un  $AR(k+1)$  avec polynôme caractéristique

$$(1 - az)(1 - \phi_1 z - \dots - \phi_k z^k) = 0$$

**: Augmented Dickey-Fuller avec contant et dérive**

$$(31) \quad \begin{aligned} Y_t &= \mu + \delta t + aY_{t-1} + X_t \\ X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_k X_{t-k} + e_t \end{aligned}$$

Avec les hypothèse:

$$H_0 : a = 1$$

$$H_1 : a < 1$$

Sous  $H_0$

$$\Delta Y_t = X_t + \delta$$

*Remarques*

- (1) C'est ADF avec dérive et constant que  $R$  utilise comme method standard pour la command `adf.test()` dans la librairie `tseries`.
- (2)  $k$  paramètre en  $R$  est précisément l'ordre du modèle AR pour  $X_t$
- (3) Si on met  $k = 0$  en  $R$  elle comptera le test DF
- (4) La distribution pour la statistique  $\hat{\tau}$  de chaque modèle change mais il sont disponibles analytiquement à Hamilton p.502 et p.528-529

Il y a encore les tests Philips-Perron et KPSS qui pourront être utilisés en complément de ADF pour verifier la stationnarité ou non d'une série.

FIGURE 1. Bruit Blanc

EXAMPLE 9. *Un bruit blanc On fait la simulation de 1000 bruit blanc comme représenté à l'image*

*Pour tester notre série pour une racine unitaire on frappe*

*Qui donne **p-value** <0.01 pour un nombre des lags 7 et alors on rejette l'hypothèse zero de non-stationnarité*

*Par contre pour une série dérivée par un bruit blanc qui est donne par la com-mende R*

FIGURE 2. L'inverse des premiers differences

*le test ADF donne **p-value** =0.7258 pour un nombre des lags 9 et alors on ne peut pas rejeter l'hypothèse zéro de la non-stationnarité.*

*Codes:*

```
x <- rnorm(1000)
plot(x)
adf.test(x)
```

*et*

```
y <- diffinv(x)
plot(y)
adf.test(y)
```

### 8.6. Deux exemples détaillés.

**Deere:** Le fichier de données nommé « deere3 » contient 57 mesures consécutives enregistrées à partir d'une machine-outil complexe chez Deere & Co. Les valeurs données sont des écarts par rapport à une valeur cible en unités de dix millièmes de pouce. Le processus utilise un mécanisme de contrôle qui réinitialise certains des paramètres de la machine-outil en fonction de l'ampleur de l'écart par rapport à la cible du dernier article produit

Pour étudier cette série nous suivons les pas de la méthodologie Box-Jenkins.

**Identification:** (1) **Stationnarité** On voit le graph et aussi les fonctions ACF et PACF

Graphiquement, la série a l'air stationnaire, les autocorrélations tendent vers zéro rapidement et seulement le premier lag est important des autocorrélations partielles (c'est qui indique un modèle AR1). Pour le vérifier on utilise encore le test ADF qui nous donne **p-value**  $< 0.01$  au niveau 3-lags et alors nous pouvons vérifier la stationnarité de la série.

(2) **Déterminer l'ordre AR et ou MA**

Pour choisir le modèle ARMA, on voit les ACF, PACF et on teste progressivement de AR1, jusqu'à AR4 et les combinaisons avec MA1, .. pour trouver que celui qui maximise le critère AIC, c'est ARMA(2,1)

**Pour resumer cette analyse, la série initial deere3 pourrait être décrit comme un modèle ARMA(2,1)**

**Wages:** Le fichier de données nommé "wages" contient mensuel valeurs de moyenne salaire par heure pour les travailleurs aux Etats-Unis dans l'industrie textile de juillet 1981 jusqu'à juin 1987

**Identification:** On commence par une plot de la série

Par le graph on peut soupçonner une tendance linéaire avec le temps, on essaie de l'enlever et afficher les résidus de la regression.

les résidus semblent un peu pathologiques et on affiche aussi le ACF pour prendre plus d'informations.

On soupçonne une non-stationnarité persistente et on prendre le premières difference pour rendre les résidus stationnaires.

maintenant ils ont l'air stationnaire. On peut vérifier les résultats par la voie statistique en faisant à chaque étape un test Box-Pierce sur les résidus. Pour le premier après la regression on rejette l'hypothèse zero

d'être bruit blanc et on peut vérifier l'existence de non stationnarité par un test ADF avant prendre la première différence.

**Ordre du modèle AR, MA :** on voit sur le dernier plot de la série stationnaire un seul lag statistiquement significatif et alors on décide de commencer avec un modèle AR1

**Pour résumer cette analyse, la série initial wage pourrait être modélisée comme un processus ARIMA(1,1,0) avec tendance linéaire.**

**Codes :** Pour cette code on a besoin de la librairie `tseries`.

```
data = read.csv("deere3.csv", col.names = 'x' )
```

```
deere3= ts(data)
```

```
plot(deere3)
```

```
acf(deere3)
pacf(deere3)
library(tseries)
adf.test(deere3)
outAR1 = arima(deere3, order = c(1,0,0))
outAR2 = arima(deere3, order = c(2,0,0))
outAR3 = arima(deere3, order = c(3,0,0))
outAR4 = arima(deere3, order = c(4,0,0))
outARMA21 = arima(deere3, order = c(2,0,1))
Box.test(outAR1$resid,lag=1)
Box.test(outAR2$resid,lag=2)
Box.test(outAR3$resid,lag=3)
Box.test(outAR4$resid,lag=4)
prediction = predict(outAR1, n.ahead=5)
plot(deere3)
points(prediction$pred, type = "l", col = 2 )
```

et

```
data2 = read.csv("wages.csv", col.names = "Wages")
wages = ts(data2)
plot(wages)
t=1:72
reg = lm(wages~t)
plot(reg$fitted.values,t, type = "l")
plot(reg$residuals, type = "l")
acf(reg$residuals)
Box.test(reg$residuals, lag = 1)
adf.test(reg$residuals)
wagesRes = reg$residuals
dwagesRes = diff(wagesRes)
plot(dwagesRes)
acf(dwagesRes)
out1= arima(dwagesRes, order = c(1,0,0))
Box.test(out1$residuals, lag = 1)
```

### 9. Processus non stationnaires : ARIMA et SARIMA

On veut revenir à la série d'origine une fois que l'on a étudié la partie stationnaire

EXEMPLE 10. Nous disposons d'une série temporelle  $(x_1, \dots, x_n)$  qui a une saisonnalité de période 12. On étudie  $y_t = x_t - x_{t-12}$  (pour supprimer la saisonnalité). L'ajustement d'un modèle ARMA et les prévisions sont réalisées sur la série  $(y_t)$ . On écrit ensuite les  $(x_t)$  en fonction des  $(y_t)$  :

$$\begin{aligned} x_t &= y_t + x_{t-12} \\ &= y_t + y_{t-12} + x_{t-24} \\ &= \dots \\ &= y_t + y_{t-12} + \dots + y_{r+12} + x_r, \end{aligned}$$

avec  $r = t$  modulo 12 (le reste de la division euclidienne de  $t$  par 12).

On connaît les  $x_1, \dots, x_n$  ( $n$  supposé plus grand que 12), et donc aussi  $y_1, \dots, y_n$ . On peut calculer les prévisions  $\hat{y}_{n,h}$  ( $h \geq 1$ ). On en déduit les prévisions  $\hat{x}_{n,h}$ . Par exemple, si  $h \in \{1, 2, \dots, 11\}$ , nous utilisons l'égalité ci-dessus avec  $t = n + h$  pour calculer la prévision

$$\hat{x}_{n,h} = \hat{y}_{n,h} + y_{t-12} + \dots + \dots + y_{r+12} + x_r.$$

**Les processus.** Ce sont des généralisation des processus ARMA aux cas non stationnaires, avec tendance polynômiale (ARIMA) ou avec une saisonnalité (SARIMA). Ce sont les processus directement utilisés par R.

DÉFINITION 5. Le processus  $(X_t)_{t \geq 0}$  est un processus ARIMA( $p, d, q$ ) si le processus  $Y_t = \Delta_1^d X_t$  est une processus ARMA( $p, q$ ).

Les processus ARIMA( $p, d, q$ ) sont donc bien adaptés à l'étude des séries temporelles présentant une tendance polynômiale de degré  $d - 1$ .

DÉFINITION 6. Le processus  $(X_t)_{t \geq 0}$  est un processus SARIMA( $p, d, q, T$ ) si le processus  $Y_t = \Delta_T \circ \Delta_1^d X_t$  est un processus ARMA( $p, q$ ).

Les processus SARIMA( $p, d, q, T$ ) sont donc bien adaptés à l'étude des séries temporelles qui présentent une saisonnalité de période  $T$  et qui ont une tendance polynômiale de degré  $d - 1$ .

RÉMARK 4. Attention, il existe dans la littérature des définitions de processus SARIMA plus complexes.

#### 9.1. Exercices supplémentaires.

(1) Devoir 1: Calculer la variance de la série:

$$Y_t = e_t + \psi e_{t-1} + \psi^2 e_{t-2} + \dots$$

PROOF.

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}\left(\sum_{i=0}^{\infty} \psi^i e_{t-i}\right) \\ &= \sum_{i=0}^{\infty} \text{Var}(\psi^i e_{t-i}) + \sum_{\substack{j=0 \\ j \neq i}}^{\infty} \text{Cov}(\psi^i e_{t-i}, \psi^j e_{t-j}) \\ &= \sum_{i=0}^{\infty} \psi^{2i} \text{Var}(e_{t-i}) = \sigma_e \frac{1}{1 - \psi^2} \end{aligned}$$

l'équation finale viens de la série géométrique de  $\psi^2$

□

(2) Calculer la variance de la série

$$Y_t = e_t + \psi e_{t-1} + \psi^2 e_{t-2} + \dots + \psi^k e_{t-k}$$

PROOF.

$$Var(Y_t) = \sigma_e \sum_{i=0}^k \psi^{2i} = \sigma_e \frac{1 - \psi^{2(k+1)}}{1 - \psi^2}$$

Pour arriver à cette dernière équation il faut écrire

$$\sum_{i=0}^k \psi^{2i} = \sum_{i=0}^{\infty} \psi^{2i} - \sum_{i=k+1}^{\infty} \psi^{2i} = \frac{1}{1 - \psi^2} - \psi^{2(k+1)} \sum_{n=0}^{\infty} \psi^{2n}$$

□

(3) Pour le modèle suivant, calculer l'espérance, la variance, la covariance et la corrélation.

$$Y_t = -Y_{t-1} + 2Y_{t-2} + e_t$$

On fait l'hypothèse que  $e_t$  est un bruit blanc centré de variance  $\sigma_e$

PROOF. On va réécrire le modèle à la forme des lags pour définir le polynôme caractéristique:

$$Y_t = -Y_{t-1} + 2Y_{t-2} + e_t \\ (-2L^2 + L + 1)Y_t = e_t$$

qui a les racines  $r_1 = 1$ ,  $r_2 = -\frac{1}{2}$ . Il est clair que nous avons une racine unitaire et alors il s'agit d'un modèle genre ARI(1,1,0).

$$\frac{1}{(1-z)(1+2z)} = \psi_1 z + \psi_2 z^2 + \dots$$

En utilisant l'équation (28) on peut réécrire l'équation précédente comme

$$\psi_k = 3\psi_{k-1} - 2\psi_{k-2} \quad \psi_0 = 1 \\ \text{ou} \quad \psi_k = \frac{1 - \phi^{k+1}}{1 - \phi}$$

Par quoi on peut calculer facilement l'espérance

$$\mathbb{E}[Y_t] = 0$$

et la variance

$$Var(Y_t) = \sigma_e \left[ 1 + 2^2 + \dots + \left( \frac{1 - 2^{t-k}}{-1} \right)^2 \right]$$

□