

Introductions à la modélisation des séries temporelles

Farouk Mhamdi

Janvier 2023

1 Introduction

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps. Elles ont été utilisées en astronomie ('on the periodicity of sunspots', 1906), en météorologie ('time-series regression of sea level on weather', 1968), en théorie du signal ('Noise in FM receivers', 1963), en biologie ('the autocorrelation curves of schizophrenic brain waves and the power spectrum', 1960), en économie ('time-series analysis of imports, exports and other economic variables', 1971)...etc.

2 Objectifs de l'études des séries temporelles

2.1 Description et modélisation

Le but est ici de déterminer les différentes composantes d'une série (X_t) , en particulier, obtenir la série corrigée des variations saisonnières (méthodes de désaisonnalisation). Pour les séries stationnaires, on peut aussi chercher à modéliser la série à l'aide d'un modèle ARMA, par exemple dans le but de faire de la prévision.

Dans le cas des modèles *ARMA*, de nombreuses relations existent afin de faire de la prévision, avec un intervalle de confiance. Nous verrons comment ces intervalles de confiance sont modifiés si une modélisation ARCH est retenue, ou du type mémoire longue.

2.2 Filtrage

Le lissage consiste à transformer une série de façon à détecter (pour éliminer ou au contraire conserver) certaines caractéristiques (composante saisonnière, points abérants...). Cette méthode permet également de détecter des ruptures au sein d'une série.

2.3 Prédiction

Sur la base d'observation X_1, \dots, X_T le but est de faire une prédiction, à la date T , de la réalisation en $T + h$, notée $\hat{X}_T(h)$. Une première méthode est le lissage exponentiel, basé sur une formule de récurrence de la forme $\hat{X}_T(1) = \alpha X_T + (1 - \alpha)\hat{X}_{T-1}(1)$, où α , compris entre 0 et 1, est généralement choisi de façon à minimiser la somme des carrés des erreurs de prédiction.

3 Propriétés des processus univariés en temps discret

La pratique de l'analyse des séries temporelles vise à modéliser une série d'observations x_1, \dots, x_n par un processus aléatoire à temps discret, c'est à dire une suite (X_n) de variables aléatoires définies sur un espace de probabilité (Ω, \mathcal{A}, P) , tel que l'on puisse penser que la série observée soit une réalisation du processus. En d'autres termes, x_1, \dots, x_n doit être obtenu comme tirage aléatoire de X_1, \dots, X_n suivant la probabilité P , c'est à dire que se réalise un événement ω tel que $x_i = X_i(\omega)$ pour $i = 1, \dots, n$. Le but est alors, étant donnée une trajectoire x_1, \dots, x_n de reconstruire la dynamique du modèle sous-jacent à dire de comprendre la liaison entre X_i et son passé $X_{i-1}, X_{i-2}, \dots, X_1$.

3.1 Rappels sur les Chaînes de Markov

Définition : Le processus $\{X_t, t \in N\}$ est une chaîne de Markov d'ordre 1 si et seulement si, pour tout t ,

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) = \mathcal{L}(X_t | X_{t-1})$$

Autrement dit, compte tenu de la trajectoire $(X_{T-1} = x_{T-1}, X_{T-2} = x_{T-2}, \dots)$ d'un processus (X_t) , la loi de X_T à l'instant T est entièrement déterminée par le fait que la valeur en $T - 1$ soit x_{T-1} .

Théorème : Le processus $\{X_t, t \in N\}$ est une chaîne de Markov d'ordre 1 si et seulement si il existe une fonction $g(\cdot)$ mesurable et un processus ε_t tel que $X_t = g(X_{t-1}, \varepsilon_t)$ avec (ε_t) une suite de variables aléatoires, indépendantes et de même loi.

Lorsque l'application g ne dépend pas de t , la chaîne de Markov est dite homogène.

Exemple 1 : Les processus $AR(1) : X_t = \alpha + \beta X_{t-1} + \varepsilon_t$, où (ε_t) est un bruit blanc indépendant du passé du processus, sont markoviens.

Exemple 2 : En particulier, les processus de la forme $X_t = X_{t-1} + \varepsilon_t$ correspondent à une marche aléatoire : - si $X_0 \in Z$ et $P(\varepsilon_t = -1) = P(\varepsilon_t = +1) = 1/2$, on obtient la marche aléatoire symétrique sur Z (jeu du pile ou face), - si ε_t suit une loi normale centrée, on obtient une discrétisation du mouvement brownien, ou un processus ARIMA $(0, 1, 0)$ comme nous l'appellerons ici.

Définition : Le processus $\{X_t, t \in N\}$ est une chaîne de Markov d'ordre p si et seulement si, pour tout t ,

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots) = \mathcal{L}(X_t | X_{t-1}, \dots, X_{t-p}).$$

3.2 Notions de 'stationnarité'

Définition : Un processus (X_t) est stationnaire au second ordre si

- (i) pour tout $t, E(X_t^2) < +\infty$,
- (ii) pour tout $t, E(X_t) = \mu$, constante indépendante de t ,
- (iii) pour tout t et pour tout $h, cov(X_t, X_{t+h}) = E([X_t - \mu][X_{t+h} - \mu]) = \gamma(h)$, indépendante de t .

Définition : La fonction $\gamma(\cdot)$ sera appelée fonction d'autocovariance

On peut montrer aisément que $\gamma(\cdot)$ est une fonction paire, au sens où $\gamma(h) = \gamma(-h)$ pour tout h .

Remarque : Une des conséquences est que variance $V(X_t)$ est constante, indépendante de $t, V(X_t) = \gamma(0)$.

Proposition : Si $(X_t, t \in Z)$ est un processus stationnaire, et si $(a_i, i \in Z)$ est une suite de réels absolument convergente, i.e. $\sum_{i \in Z} |a_i| < +\infty$, alors, le processus (Y_t) défini par :

$$Y_t = \sum_{i \in Z} a_i X_{t-i}, \text{ pour tout } t \in Z,$$

est un processus stationnaire.

Corollaire : En particulier, si $(a_i, i \in Z)$ est une suite de réels finie, la suite Y_t est stationnaire. Par exemple, si $a_0 = a_1 = 1/2$, et $a_i = 0$ pour $i \notin \{0, 1\}$:

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

est stationnaire dès lors que (X_t) est stationnaire. De même pour $Y_t = X_t - X_{t-1}$.

Définition : Un processus (X_t) est stationnaire au sens fort si pour tous t_1, \dots, t_n et h on a l'égalité en loi

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{\mathcal{L}}{=} (X_{t_1+h}, \dots, X_{t_n+h}).$$

Remarque : Cette notion revient à dire que la loi temporelle est invariante en temps. Cette stationnarité est beaucoup plus forte que la stationnarité du second ordre, puisqu'on ne recherche pas la stabilité de la loi, mais seulement la stabilité des deux premiers moments.

Définition : On appelle bruit blanc (parfois appelé bruit blanc faible) un

processus (ε_t) stationnaire dont les autocovariances sont toutes nulles : $\gamma(h) = 0$ pour $h \neq 0$.

Définition : Un processus stationnaire (X_t) sera dit ergodique si pour tout $p \in \mathbb{N}^*$, et pour toute fonction borélienne de \mathbb{R}^p à valeurs dans \mathbb{R} , on a :

$$\frac{1}{N} \sum_{i=1}^N f(X_{i+1}, X_{i+2}, \dots, X_{i+p}) \rightarrow E(f(X_1, X_2, \dots, X_p)), \text{ quand } N \rightarrow \infty$$

qui peut être vu simplement comme une généralisation de la loi de grand nombre.

Remarques : La notion de stationnarité (faible, ou au second ordre) se définit par une invariance des moments d'ordre 1 et 2 au cours du temps. Par opposition, on dira qu'une série est non-stationnaire si elle n'est pas stationnaire.

On peut noter que la classe des processus non-stationnaires est alors relativement vaste, et surtout hétérogène : il existe différentes sources de non-stationnarité, et à chaque origine de non-stationnarité est associée une méthode propre de stationnarisation.

Nelson et Plosser ont retenu, en 1982, deux classes de processus non-stationnaires : les processus TS (trend stationary) et les processus DS (difference stationary). Les premiers correspondent à une non-stationnarité de type déterministe, alors que les seconds correspondent à une non-stationnarité de type stochastique.

Définition : (X_t) est un processus non-stationnaire TS s'il peut s'écrire sous la forme $X_t = f(t) + Z_t$ où $f(t)$ est une fonction (déterministe) du temps, et (Z_t) est un processus stationnaire.

L'exemple le plus simple est celui de la tendance linéaire bruitée :

$$X_t = \alpha + \beta t + \varepsilon_t.$$

Ce processus est en effet non-stationnaire puisque son espérance vaut $\alpha + \beta t$ à la date t , et donc, dépend de t .

Une des propriétés importantes de ce type de processus est qu'il n'y a pas persistance des chocs : l'influence d'un choc subit à un instant τ aura tendance à s'estomper au cours du temps, et la variable rejoint alors sa dynamique de long-terme, déterminée par $f(t)$.

Définition : (X_t) est un processus non-stationnaire DS - ou intégré d'ordre d , noté $I(d)$ - si le processus obtenu après d différenciation est stationnaire : $Z_t = \Delta^d X_t = (1 - L)^d X_t$ est stationnaire.

Comme nous le verrons par la suite, le fait qu'il faille différencier d fois, c'est à dire multiplier par $(1 - L)^d$, polynôme de l'opérateur retard L , revient à chercher la présence de racines unitaires : si le processus $\Phi(L)X_t$ est stationnaire, si 1 est une racine du polynôme Φ , alors (X_t) sera non-stationnaire. C'est pour cela que la plupart des tests de nonstationnarité sont des tests de détection de racine unitaire.

Remarque : Pour obtenir les racines d'un polynôme, on peut utiliser la commande sous R `Mod(polyroot)`, `library(polynom)`, par exemple pour $\Phi(L) = (1 + 0.7L - 0.5L^2)$ ou $\Phi(L) = (1 + 0.7L - 0.2L^2)$.

3.3 Fonction d'autocovariance et autocorrélation

Définition : Pour une série stationnaire (X_t) , on définit la fonction d'autocovariance, pour tout t , par

$$h \mapsto \gamma_X(h) = \text{cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) - E(X_t) \cdot E(X_{t-h}).$$

Définition : Pour une série stationnaire (X_t) , on définit la fonction d'autocorrélation, pour tout t , par

$$h \mapsto \rho_X(h) = \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Cette fonction $\rho_X(\cdot)$ est à valeurs dans $[-1, +1]$, et $\rho_X(0) = 1$.

Définition : Un processus (ε_t) sera appelé bruit blanc (faible) s'il est stationnaire, centré et non-autocorrélé :

$$E(\varepsilon_t) = 0, V(\varepsilon_t) = \sigma^2 \text{ et } \rho_\varepsilon(h) = 0 \text{ pour } h \neq 0.$$

On parlera de bruit blanc fort s'il est indépendant et identiquement distribué (i.i.d.) : la notion d'indépendance est plus forte que la nullité des autocorrélations, et le fait que le processus soit identiquement distribué est plus fort que la stabilité des deux premiers moments.

Exemple : Processus MA(1) : $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ où (ε_t) est un bruit blanc centré de variance σ^2 ,

$$\begin{cases} \gamma(0) = [1 + \theta^2] \sigma^2 \\ \gamma(1) = \theta\sigma^2 \\ \gamma(h) = 0 \text{ si } |h| \geq 2 \end{cases}, \text{ soit } \rho(1) = \frac{\theta}{1 + \theta^2} \text{ et } \rho(h) = 0 \text{ pour } |h| \geq 2$$

Estimation de la fonction d'autocorrélation

Considérons un ensemble d'observations X_1, \dots, X_T . La moyenne empirique est donnée par :

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t.$$

La fonction d'autocovariance empirique est donnée par :

$$\hat{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T),$$

et la fonction d'autocorrélation empirique est donnée par :

$$\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}_T(0)}.$$

Si ces estimateurs sont biaisés (à distance finie), ils sont malgré tout asymptotiquement sans biais.

Proposition : Les moments empiriques convergent vers les moments théoriques : $\bar{X}_T \rightarrow m$, $\hat{\gamma}_T(h) \rightarrow \gamma(h)$ et $\hat{\rho}_T(h) \rightarrow \rho(h)$ quand $T \rightarrow \infty$

Remarque : Bien que ces fonctions soient définies pour tout h tel que $-T < h < T$, la fonction d'autocovariance empirique fournit un estimateur très pauvre de $\gamma(h)$ pour des valeurs h proches de n . A titre indicatif, Box et Jenkins recommandent de n'utiliser ces quantités que si $T > 50$ et $h \leq T/4$.

Proposition : Si (X_t) est un processus linéaire, au sens où il satisfait $X_t = \sum_{j \in \mathbb{Z}} \phi_j \varepsilon_{t-j}$ où (ε_t) est une suite de variables i.i.d. centrées, telle que $E(\varepsilon_t^4) = \eta E(\varepsilon_t^2)^2 < +\infty$, et $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, et où les ϕ_j définissent une série absolument convergente, et où η est une constante positive, alors, on a, pour tout $p \geq 0$,

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_T(0) \\ \vdots \\ \hat{\gamma}_T(p) \end{pmatrix} \rightarrow \mathcal{N} \left(\begin{pmatrix} \gamma(0) \\ \vdots \\ \gamma(p) \end{pmatrix}, V \right)$$

où V est la matrice de variance-covariance définie par

$$V = \left[\eta \gamma(h) \gamma(k) + \sum_{i=-\infty}^{+\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h) \right]_{h,k=0,\dots,p}$$

4 Désaisonnalisation par régression linéaire

4.1 Le modèle linéaire

On considère une série X_t contenant 2 composantes déterministes : une tendance Z_t , d'une saisonnalité S_t et d'une composante aléatoire ε_t

$$X_t = Z_t + S_t + \varepsilon_t.$$

On suppose que Z_t et S_t sont des combinaisons linéaires de fonctions connues dans le temps, Z_t^i et S_t^j , i.e.

$$\begin{cases} Z_t = Z_t^1 \beta_1 + Z_t^2 \beta_2 + \dots + Z_t^m \beta_m \\ S_t = S_t^1 \gamma_1 + S_t^2 \gamma_2 + \dots + S_t^n \gamma_n \end{cases}$$

Le but est d'estimer les β_1, \dots, β_m et $\gamma_1, \dots, \gamma_n$ à partir des T observations.

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^n S_t^j \gamma_j + \varepsilon_t \text{ pour } t = 1, \dots, T$$

Hypothèses sur les erreurs :

On supposera l'hypothèse suivante vérifiée, à savoir que les erreurs sont centrées: $E(\varepsilon_t) = 0$, de même variance $V(\varepsilon_t) = \sigma^2$ et non-corrélées $cov(\varepsilon_t, \varepsilon_{t-h}) = 0$ pour tout $h > 0$.

Composante saisonnière du modèle

La forme de S_t dépend du type de données, et de la forme de la saisonnalité. On peut citer par exemple les fonctions S_t^i indicatrices,

$$S_t^i = \begin{cases} 0 & \text{si } t \neq \text{mois } i \\ 1 & \text{si } t = \text{mois } i \end{cases}$$

Pour des données trimestrielles, on a $S_t = S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + S_t^4 \gamma_4$ où S_t^j est la fonction indicatrice du trimestre j .

Composante tendancielle

Cette composante a généralement une forme simple, reflétant la croissance moyenne.

Pour une tendance linéaire, $Z_t = \beta_1 + \beta_2 t$ on pose $Z_t^1 = 1$ et $Z_t^2 = t$.

Plusieurs types de composantes tendancielle existent:

- (i) linéaire: $Z_t = \beta_0 + \beta_1 t$,
- (ii) exponentielle: $Z_t = \alpha \beta^t$, ou $Z_t = \alpha(1+r)^t$ ou encore $Z_t = \alpha \exp(rt)$,
- (iii) quadratique $Z_t = \beta_0 + \beta_1 t + \beta_2 t^2$,
- (iv) de Gompertz $Z_t = \exp(\alpha \beta^t + \gamma)$,
- (v) logistique $Z_t = [\alpha \beta^t - \gamma]^{-1}$.

4.2 Applications sous R

Séries utilisées :

- (i) le trafic voyageur de la SNCF en France.
- (ii) le trafic sur l'autoroute A7.

5 Désaisonnalisation par moyennes mobiles

On considère une série temporelle (X_t) admettant une décomposition selon le schéma additif suivant :

$$X_t = Z_t + S_t + t \text{ pour } t = 1, \dots, T.$$

Le but est de trouver une transformation du processus X_t qui annule la composante saisonnière S_t : on cherche un "filtre" ϕ tel que $Y_t = \phi(X_t) = Z_t + t$

5.1 Généralités sur les moyennes mobiles

a. Notion d'opérateur retard L

Définition : On appellera opérateur retard L (=lag, ou B = backward) l'opérateur linéaire défini par

$$L : X_t \mapsto L(X_t) = LX_t = X_{t-1},$$

et opérateur avance F (=forward)

$$F : X_t \mapsto F(X_t) = FX_t = X_{t+1},$$

Remarque : $L \circ F = F \circ L = I$ (opérateur identité) et on notera par la suite $F = L^{-1}$ et $L = F^{-1}$.

(i) Il est possible de composer les opérateurs: $L^2 = L \circ L$, et plus généralement,

$$L^p = \underbrace{L \circ L \circ \dots \circ L}_{p \text{ fois}} \text{ où } p \in \mathbb{N}$$

avec la convention $L^0 = I$. On notera que $L^p(X_t) = X_{t-p}$.

(ii) Soit A le polynôme, $A(z) = a_0 + a_1z + a_2z^2 + \dots + a_pz^p$.

On notera $A(L)$ l'opérateur

$$A(L) = a_0I + a_1L + a_2L^2 + \dots + a_pL^p = \sum_{k=0}^p a_kL^k.$$

Soit (X_t) une série temporelle. La série (Y_t) définie par $Y_t = A(L)X_t$ vérifie

$$Y_t = A(L)X_t = \sum_{k=0}^p a_kX_{t-k}.$$

Par passage à la limite, on peut aussi définir des séries formelles

$$A(z) = \sum_{k=0}^{\infty} a_kz^k \text{ et } A(L) = \sum_{k=0}^{\infty} a_kL^k.$$

Proposition : Pour toutes moyennes mobiles A et B, alors

$$\begin{cases} A(L) + B(L) = (A + B)(L) \\ \alpha \in R, \alpha A(L) = (\alpha A)(L) \\ A(L) \circ B(L) = (AB)(L) = B(L) \circ A(L). \end{cases}$$

La moyenne mobile $C = AB = BA$ vérifie alors

$$\left(\sum_{k=0}^{\infty} a_k L^k \right) \circ \left(\sum_{k=0}^{\infty} b_k L^k \right) = \left(\sum_{i=0}^{\infty} c_i L^i \right) \text{ où } c_i = \sum_{k=0}^i a_k b_{i-k}.$$

b. Les moyennes mobiles :

Définition : Une moyenne mobile est un opérateur linéaire, combinaison linéaire d'opérateurs retard

$$M = \sum_{i=-m_1}^{m_2} \theta_i L^{-i}, m_1, m_2 \in N,$$

qui peut s'écrire :

$$M = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} L^{-i} = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} F^i = L^{m_1} \Theta(F),$$

où $\Theta(.)$ est un polynôme appelé polynôme caractéristique de M , de degré $m_1 + m_2$, et $m_1 + m_2 + 1$ sera appelé ordre de M (correspondant au nombre (théorique) de terme de M).

Définition : Si $m_1 = m_2 = m$, la moyenne mobile sera dite centrée. De plus, si M est centrée, et que pour tout $i, \theta_i = \theta_{-i}$ alors la moyenne mobile est dite symétrique.

Exemple : La moyenne mobile $M_1(X_t) = (X_t + X_{t-1})/2$, soit $M_1 = (L + I)/2 = L[I + F]/2$ est de degré 1, d'ordre 2 et n'est pas centrée (ni symétrique).

Exemple : La moyenne mobile $M_2(X_t) = (X_{t+1} + 2X_t + X_{t-1})/4$, soit $M_2 = (L^{-1} + 2I + L)/4 = L[I + 2F + F^2]/4$ est de degré 2, d'ordre 3, est centrée et symétrique.

On peut déjà noter, pour les moyennes centrées symétriques, sont nécessairement d'ordre impair (pour être centrées). Pour m impair, on considèrera les moyennes mobiles d'ordre $m = 2p + 1$ définie par :

$$M_m(X_t) = \frac{1}{m} [X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + X_{t+p}].$$

Exemple : La moyenne mobile d'ordre 3

- Cette moyenne mobile a pour coefficients $1/3, 1/3, 1/3$,

$$M_3(X_t) = \frac{1}{3} [X_{t-1} + X_t + X_{t+1}].$$

Exemple : La moyenne mobile d'ordre 9.

- Cette moyenne mobile a pour coefficients $1/9, 1/9, \dots, 1/9$

$$M_9(X_t) = \frac{1}{9} [X_{t-4} + X_{t-3} + \dots + X_t + \dots + X_{t+4}].$$

De manière générale, le filtre

$$M_{2p+1}(X_t) = \frac{1}{2p+1} [X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + X_{t+p}]$$

Remarque : Moyennes mobiles centrées et symétriques d'ordre pair

Toutefois, il est possible de construire des moyennes mobiles centrées et symétriques d'ordre pair, de façon artificielle. Pour cela, pour $m = 2p$ on considérera les moyennes mobiles définies par

$$M_m(X_t) = \frac{1}{m} [X_{t-p+1/2} + \dots + X_{t-1/2} + X_{t+1/2} + \dots + X_{t+p-1/2}],$$

où $X_{t-1/2}$ est obtenue comme valeur intermédiaire entre X_{t-1} et X_t . Cette moyenne mobile peut donc se réécrire comme suit :

$$\begin{aligned} M_m(X_t) &= \frac{1}{m} \left[\frac{1}{2} (X_{t-p} + X_{t-p+1}) + \dots + \frac{1}{2} (X_{t-1} + X_t) + \frac{1}{2} (X_t + X_{t+1}) + \dots + \frac{1}{2} (X_{t+p-1} + X_{t+p}) \right] \\ &= \frac{1}{m} \left[\frac{1}{2} X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + \frac{1}{2} X_{t+p} \right]. \end{aligned}$$

Cette moyenne mobile d'ordre pair est en fait une moyenne mobile d'ordre impair, que l'on notera $M_{2 \times p}$, définie par

$$M_{2 \times p}(X_t) = \frac{1}{2m} [X_{t-p} + 2X_{t-p+1} + \dots + 2X_{t-1} + 2X_t + 2X_{t+1} + \dots + 2X_{t+p-1} + X_{t+p}].$$

Exemple : La moyenne mobile 2×4 - Cette moyenne mobile permet d'estimer des tendances dans le cas de données trimestrielles, elle est d'ordre 5 et de coefficients $1/8, 1/4, 1/4, 1/4, 1/8$

$$M_{2 \times 4}(X_t) = \frac{1}{8} [X_{t-2} + 2X_{t-1} + 2X_t + 2X_{t+1} + X_{t+2}].$$

Comme nous le verrons par la suite, elle élimine les saisonnalités trimestrielles des séries trimestrielles, elle conserve les tendances linéaires, et elle réduit de 75% la variance d'un bruit blanc.

Exemple : La moyenne mobile 2×12 - Cette moyenne mobile permet d'estimer des tendances dans le cas de données mensuelles, elle est d'ordre 13 et de coefficients $1/24, 1/12, 1/12, \dots, 1/12, 1/24$

Comme nous le verrons par la suite, elle élimine les saisonnalités annuelles des séries mensuelles, elle conserve les tendances linéaires, et elle réduit de plus de

90% la variance d'un bruit blanc.

C. L'espace des opérateurs moyenne-mobile

Définition : Soient M_1 et M_2 deux moyennes mobiles. Le produit de M_1 et M_2 est obtenu par composition des moyennes mobiles

$$M_1 M_2 (X_t) = M_1 \circ M_2 (X_t).$$

Proposition : Ce produit est commutatif et associatif

$$M_1 M_2 = M_2 M_1 \text{ et } M_1 (M_2 M_3) = (M_1 M_2) M_3.$$

De plus, le produit est distributif par rapport à l'addition.

Proposition : La composée de deux moyennes mobiles symétriques est symétrique.

5.2 Vecteurs propres associés à une moyenne mobile

Définition : Soit M une moyenne mobile. S'il existe λ et (X_t) non nul tels que $M(X_t) = \lambda X_t$, (X_t) sera vecteur propre associé à la valeur propre λ .

A. Les séries absorbées : $\lambda = 0$

Définition : Une suite (X_t) est dite absorbée par M si et seulement si $M(X_t) = 0$ pour tout t .

Exemple : Soit M la moyenne mobile définie par $M(X_t) = X_t + X_{t-1} + X_{t-2}$. La série chronologique définie récursivement par $Y_t = -[Y_{t-1} + Y_{t-2}]$ est absorbée par M .

Proposition : Les vecteurs propres associés à la valeur propre $\lambda = 0$ forment un espace vectoriel de dimension $m_1 + m_2$, dont une base est constituée des $Z_t^k = (\alpha_k r^t)$ pour $k = 0, 1, \dots, p-1$, où r est racine non nulle du polynôme Θ . Illustration sur un exemple Dans l'exemple précédent, on peut chercher à construire une base de la forme $Z_t = r^t$, qui devra satisfaire

$$r^t + r^{t-1} + r^{t-2} = 0 \text{ pour tout } t$$

c'est à dire $r^2 + r + 1 = 0$. Aussi, r est une racine du polynôme caractéristique de M si et seulement si

$$r = \frac{-1 \pm i\sqrt{3}}{2} \text{ soit } r_1 = \exp\left(\frac{2i\pi}{3}\right) \text{ et } r_2 = \exp\left(-\frac{2i\pi}{3}\right)$$

Aussi, les suites absorbées sont nécessairement de la forme

$$X_t = \lambda r_1^t + \mu r_2^t, \text{ pour tout } t.$$

Or

$$\begin{cases} r_1^t = \cos(2t\pi/3) + i \sin(2t\pi/3) \\ r_2^t = \cos(2t\pi/3) - i \sin(2t\pi/3). \end{cases}$$

Et donc, l'espace vectoriel des suites absorbées par M admet pour base réelle

$$\mathcal{B} = \left\{ \cos\left(2t\frac{\pi}{3}\right), \sin\left(2t\frac{\pi}{3}\right) \right\},$$

ce qui correspond à des séries chronologiques de la forme

$$X_t = \lambda \cos\left(2t\frac{\pi}{3}\right) + \mu \sin\left(2t\frac{\pi}{3}\right) \text{ pour tout } t.$$

B. Les séries invariantes : $\lambda = 1$

Définition : Une suite (X_t) est dite invariante par M si et seulement si $M(X_t) = 0$ pour tout t

Une suite (X_t) est dite invariante par M si elle est absorbée par $(M - I)$.

Proposition : (i) Les suites constantes sont invariantes par M si et seulement si la somme de ses coefficients vaut 1 ,

(ii) Les polynômes de degré k sont invariantes par M si et seulement si 1 est racine d'ordre au moins $k + 1$ de $\Phi = \Theta(z) - z^{m_1}$, où $M = L^{m_1}\Theta(F)$,

(iii) Si M est symétrique et conserve les constantes, alors M conserve les polynômes de degré 1.

Démonstration

Remarque :

1) Les propriétés (i) et (iii) montrent dans quel cas la tendance de la série reste invariante : ces séries peuvent servir à enlever la composante saisonnière, pour récupérer la tendance linéaire.

2) Le but des moyennes mobiles est :

- (i) d'absorber les composantes saisonnières en laissant invariantes les tendances,
- (ii) de réduire la variance des perturbations.

Définition L'indice de réduction de la moyenne mobile M est donné par :

$$\tau = \frac{E(MX_t^2)}{E(X_t^2)} = \sum_j \theta_j^2.$$

Exercice : Déterminer l'indice de réduction et les composantes absorbées par :

- $M_{2 \times 4}$.

- $M_{2 \times 12}$.

- Dans le cas d'une moyenne mobile définie par $M(X_t) = [X_t + X_{t-1}]/2$, alors $\tau = 1/2$.

5.3 Applications sous R

6 Introduction aux modèles linéaires ARMA

On considère le processus (X_t) défini sur l'espace de probabilité (Ω, \mathcal{A}, P) , à valeurs dans R .

Définition L'espace $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ est l'espace des variables de carré intégrable (variances-covariances finies).

6.1 Rappels : Fonction et matrices autocorrélations

Pour rappels, un processus (X_t) est stationnaire (au second ordre) si pour tout t , $E(X_t^2) < +\infty$, pour tout t , $E(X_t) = \mu$, constante indépendante de t et, pour tout t et pour tout h , $cov(X_t, X_{t+h}) = \gamma(h)$, indépendante de t .

6.1.1 Autocovariance et autocorrélation

Pour une série stationnaire (X_t) , on définit la fonction d'autocovariance $h \mapsto \gamma_X(h) = cov(X_t, X_{t-h})$ pour tout t , et on définit la fonction d'autocorrélation $h \mapsto \rho_X(h) = \gamma_X(h)/\gamma_X(0)$ pour tout t , soit

$$\rho_X(h) = corr(X_t, X_{t-h}) = \frac{cov(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}$$

Définition : On appellera matrice d'autocorrélation du vecteur $(X_t, X_{t-1}, \dots, X_{t-h+1})$

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-3) \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(1) & 1 \end{bmatrix}$$

Remarque : Ces fonctions sont estimées, pour un échantillon X_1, \dots, X_T , de la façon suivante:

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t-h} \text{ et } \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},$$

(quand le processus est centré, sinon, il faut considérer $(X_t - \mu)(X_{t-h} - \mu)$).

6.2 Autocorrélations partielles

Les deux précédentes mesures de dépendance entre X_t et X_{t+h} ne faisaient intervenir que les variables X_t et X_{t+h} . Nous allons introduire ici une notion faisant intervenir les variables intermédiaires. Nous supposons, sans perte de généralité que le processus (X_t) est centré : $E(X_t) = 0$ pour tout t .

Définition : Pour une série stationnaire (X_t) , on définit la fonction d'autocorrélation partielle $h \mapsto \psi_X(h)$ par

$$\psi_X(h) = \text{corr} \left(\widehat{X}_t, \widehat{X}_{t-h} \right),$$

où

$$\begin{cases} \widehat{X}_{t-h} = X_{t-h} - EL(X_{t-h} \mid X_{t-1}, \dots, X_{t-h+1}) \\ \widehat{X}_t = X_t - EL(X_t \mid X_{t-1}, \dots, X_{t-h+1}). \end{cases}$$

On regarde ici la projection (ou l'espérance linéaire) les deux valeurs extrêmes X_t et X_{t-h} sur l'ensemble des valeurs intermédiaires.

On a le résultat suivant, permettant d'obtenir les coefficients de façon récursive

Proposition : Soit (X_t) un processus stationnaire, alors $\psi_X(0) = 1$, et, pour $h \geq 1$, $\psi_X(h)$ est le coefficient relatif à X_{t-h} dans la projection de X_t sur $X_{t-1}, \dots, X_{t-h+1}, X_{t-h}$, soit $a_h(h)$.

Théorème : Il est équivalent de connaître la fonction d'autocorrélation $(\rho_X(h))$ ou la fonction d'autocorrélation partielle $(\psi_X(h))$.

Exemple : En particulier, on peut noter que

$$\psi_X(1) = \rho_X(1) \text{ et } \psi_X(2) = \frac{[\rho_X(2) - \rho_X(1)^2]}{[1 - \rho_X(1)^2]}$$

Autre formulation :

Une autre formulation consiste à dire que la fonction d'autocorrélation partielle mesure la corrélation entre X_t et X_{t-h} une fois retirée l'influence des variables antérieures à X_{t-h} . En reprenant les notations de la partie précédente,

$$\mathcal{R}(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-3) & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-4) & \rho(h-3) & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-5) & \rho(h-4) & \rho(h-3) \\ & & \ddots & \ddots & \ddots & & \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \ddots & 1 & \rho(1) & \rho(2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & & \rho(1) & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(2) & \rho(1) & 1 \end{bmatrix}$$

et on introduit de façon analogue la matrice $\mathcal{R}^*(h)$ obtenue en remplaçant la dernière colonne de $\mathcal{R}(h)$ par le vecteur $[\rho(1), \dots, \rho(h)]'$,

$$\mathcal{R}^*(h) = \begin{bmatrix} 1 & \rho(1) & \rho(2) & & \rho(h-3) & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & & \rho(h-4) & \rho(h-3) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \ddots & \rho(h-5) & \rho(h-4) & \rho(3) \\ & & \ddots & \ddots & \ddots & & \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \ddots & 1 & \rho(1) & \rho(h-2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & & \rho(1) & 1 & \rho(h-1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & & \rho(2) & \rho(1) & \rho(h) \end{bmatrix}$$

Il est alors possible de montrer simplement que

$$\psi_X(h) = \frac{|\mathcal{R}^*(h)|}{|\mathcal{R}(h)|}$$

pour tout h

6.3 Représentation de Wald

Propriété : Si $Y_t, t \in Z$ est un processus stationnaire, et si $(a_i, i \in Z)$ est une suite de nombres réels absolument sommable $\left(\sum_{i=-\infty}^{+\infty} |a_i| < +\infty\right)$, alors :

$$Z_t = \sum_{i=-\infty}^{+\infty} a_i Y_{t-i}, t \in Z$$

est un nouveau processus stationnaire. On parle de représentation moyenne mobile infinie, notée $MA(\infty)$

Les processus pouvant s'écrire comme moyennes mobiles infinies de bruits blancs sont appelés processus linéaires. Le théorème qui suit justifie l'utilisation des représentations moyennes mobiles infinies dites unilatérales vers le passé, c'est-à-dire dans lesquelles $a_k = 0$ pour tout $k < 0$.

Le Théorème de Wold (ou décomposition de Wold) :

Tout processus stationnaire du second ordre peut être représenté sous la forme:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + m$$

où les paramètres ψ_j satisfont $\psi_0 = 1, \psi_j \in R, \forall j \in N^*, \sum_{j=0}^{\infty} \psi_j^2 < \infty$ et où ε_t est un bruit blanc i.i.d. $(0, \sigma_\varepsilon^2)$

Remarques : On montre que le processus $BB\varepsilon_t$ est le processus d'innovation de Y_t .

On appelle innovation d'un processus du 2nd ordre Y_t , la variable :

$$\varepsilon_t = Y_t - Y_t^* = Y_t - E(Y_t \mid 1, Y_{t-1}, Y_{t-2}, \dots)$$

Y_t^* est la prévision optimale de Y_t fonction de son passé (meilleure approximation linéaire) et l'erreur de prévision correspondante à 1 pas est appelée innovation.

Ainsi, l'innovation est la partie de Y_t non corrélée au passé de la série.

6.4 Rappels sur les opérateurs retards

Nous avons défini précédemment l'opérateur retard L par $L :$

$X_t \mapsto L(X_t) = LX_t = X_{t-1}$ et l'opérateur avance F par $F : X_t \mapsto F(X_t) = FX_t = X_{t+1}$. On notera alors :

$$L^p = \underbrace{L \circ L \circ \dots \circ L}_{p \text{ fois}} \text{ où } p \in \mathbb{N},$$

avec la convention $L^0 = I$ et $L^{-1} = F$. Et de façon analogue, $L^{-p} = F^p$ pour $p \in \mathbb{N}$.

Soit $A(\cdot)$ un polynôme, on cherche $B(\cdot)$ tel que $A(\cdot) \circ B(\cdot) = B(\cdot) \circ A(\cdot) = 1$.

6.4.1 inversibilité de $P(L) = 1 - \lambda L$

Proposition (i) Si $|\lambda| < 1$ alors $1 - \lambda L$ est inversible, et de plus,

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

(ii) Si $|\lambda| > 1$ alors $1 - \lambda L$ est inversible, et de plus,

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

(iii) Si $|\lambda| = 1$, alors $1 - \lambda L$ n'est pas inversible.

Démonstration :

Exemple. Soit (X_t) et (Y_t) deux processus stationnaires tels que $Y_t = X_t - \lambda X_{t-1} = (1 - \lambda L)X_t$, où $\lambda < 1$. Cette relation s'inverse en

$$X_t = (1 - \lambda L)^{-1}Y_t = Y_t + \lambda Y_{t-1} + \dots + \lambda^k Y_{t-k} + \dots$$

Exemple Dans le cas où $\lambda = 1$ (racine unité) on se retrouve en présence d'une marche aléatoire $Y_t = X_t - X_{t-1}$ (non stationnaire).

6.4.2 Inversibilité des polynômes en L

Tout polynôme $A(L) = 1 + a_1L + \dots + a_nL^n$ (normalisé tel que $A(0) = 1$), peut s'écrire :

$$A(z) = a_n (z - z_1)(z - z_2) \dots (z - z_n),$$

correspondant à la décomposition en éléments simples ($z_i =$ racines du polynôme). On peut écrire

$$A(L) = \prod_{i=1}^n (1 - \lambda_i L) \text{ où } \lambda_i = \frac{1}{z_i}$$

Proposition : Si pour tout $i, |\lambda_i| \neq 1$, alors $A(L)$ est inversible.

6.5 Modèles ARMA et représentations canoniques

On s'intéresse aux processus ARMA, ceux-ci donnant des représentations plus parcimonieuses que la représentation de Wold (qui nécessite un nombre infini de paramètres!) :

$$\begin{aligned} Y_t - \sum_{i=1}^p \phi_i Y_{t-i} &= \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \\ \left(1 - \sum_{i=1}^p \phi_i L^i\right) Y_t &= \left(1 - \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \\ \Phi(L) Y_t &= \Theta(L) \varepsilon_t \end{aligned}$$

On ne s'intéressera qu'aux **représentations canoniques**, c'est-à-dire aux écritures des modèles ARMA telles que ε_t soit le processus d'innovation de $Y_t \iff$ la prédiction optimale est $\hat{Y}_T(k) = E(Y_{T+k}/Y_T, Y_{T-1}, \dots)$ et l'erreur de prévision à 1 pas est l'innovation \Rightarrow on ne peut pas faire mieux que la prédiction optimale.

Il s'agit donc de trouver des contraintes sur les paramètres des modèles ARMA. **Quelles sont ces conditions?**

- il faut que la représentation soit stationnaire, donc que Φ admette une inverse (\Rightarrow module des racines de Φ différent de 1) :

$$Y_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t = \sum_{j=-\infty}^{\infty} h_j \varepsilon_{t-j} \text{ avec } \sum |h_j| < \infty$$

- il faut que la représentation soit inversible, c'est-à-dire que la forme $AR(\infty)$ soit tournée vers le passé (\Rightarrow module des racines de Θ strictement supérieur à 1) :

$$\Theta(L)^{-1} \Phi(L) Y_t = \varepsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}$$

Y_t peut s'écrire en fonction de son passé et de ε_t

- il faut que la représentation soit causale, c'est-à-dire que la forme $MA(\infty)$ soit tournée vers le passé (\Rightarrow module des racines de Φ strictement supérieur à 1) :

$$Y_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t = \sum_{j=0}^{\infty} h_j \varepsilon_{t-j}$$

Y_{t-i} non corrélé à $\varepsilon_t \Rightarrow$ Alors, ε_t est le processus d'innovation de Y_t .

6.6 Les processus autorégressifs : $AR(p)$

6.6.1 Définition

On appelle processus autorégressif d'ordre p , noté $AR(p)$, un processus stationnaire $\{Y_t\}_{t \in \mathbb{Z}}$ vérifiant une relation du type :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t \quad \forall t \in \mathbb{Z}$$

où les ϕ_i sont des réels, $\phi_p \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

$$\iff (1 - \phi_1 L - \dots - \phi_p L^p) Y_t = \varepsilon_t \iff \Phi(L) Y_t = \varepsilon_t$$

6.6.2 Représentation stationnaire

Ce processus est pour l'instant défini sous forme implicite et en particulier il n'est pas certain que cette dernière équation admette toujours une solution stationnaire.

Si le polynôme Φ a toutes ses racines de module différent de 1, on peut inverser l'opérateur $\Phi(L)$. On en déduit que l'équation admet une solution unique, avec une écriture $MA(\infty)$:

$$Y_t = \Phi(L)^{-1} \varepsilon_t = \sum_{i=-\infty}^{\infty} h_i \varepsilon_{t-i}$$

On peut alors montrer que l'on a $\sum_{i=-\infty}^{\infty} |h_i| < \infty$ et donc que la représentation est stationnaire.

6.6.3 Représentation inversible

La représentation $AR(p)$ est inversible par définition.

6.6.4 Représentation causale

- Si le polynôme Φ a toutes ses racines de module strictement supérieur à 1, l'opérateur inverse $\Phi(L)^{-1}$ admet un développement ne faisant intervenir que les puissances positives de L . On a alors :

$$Y_t = \sum_{i=0}^{\infty} h_i \varepsilon_{t-i}$$

Dans ce cas, on montre que l'on a :

$$\sum_{i=0}^{\infty} |h_i| < \infty, \quad h_0 = 1$$

Dans ce cas, Y_{t-1}, Y_{t-2}, \dots sont fonctions linéaires de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ et en particulier sont non corrélés avec ε_t .

Projetant la relation AR sur le passé de X , on obtient:

$$E(Y_t / Y_{t-1}, Y_{t-2}, \dots) = \sum_{i=1}^p \phi_i Y_{t-i}$$

ainsi, le bruit blanc est aussi l'innovation puisque :

$$Y_t - E(Y_t / Y_{t-1}, Y_{t-2}, \dots) = \varepsilon_t$$

- Lorsque certaines racines de $\Phi(z) = 0$ sont de module inférieur à 1, alors ε_t n'est pas l'innovation, puisque la forme $MA(\infty)$ sera tournée vers le futur (et peut-être aussi vers le passé si certaines racines étaient bien supérieur à 1 en module). Dans ce cas, le passé de Y_t dépend du passé et du futur de ε_t : on ne peut plus dire qu'il n'y a pas de corrélation entre le passé de Y_t et ε_t ! Ainsi, ce n'est pas ε_t qui est l'innovation du processus.

Exemple de l'AR(1): $Y_t = \phi Y_{t-1} + \varepsilon_t$ avec $\varepsilon_t \text{BB}(0, \sigma_\varepsilon^2)$
La racine de $\Phi(z) = 1 - \phi z$ est $z = 1/\phi$.

- Si $|z| = 1$ et donc $|\phi| = 1$, par exemple $\phi = 1$, alors :

$$Y_t = Y_{t-1} + \varepsilon_t = \dots = Y_{t-n} + \varepsilon_{t-n+1} + \dots + \varepsilon_{t-1} + \varepsilon_t$$

et la variance dépend de t ! donc il n'existe pas de solution stationnaire. Il s'agit d'un processus non stationnaire de type stochastique, appelé marche aléatoire (random walk).

- Si $|z| > 1$ et donc $|\phi| < 1$ alors la forme $MA(\infty)$ est donnée par :

$$Y_t = (1 - \phi L)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \phi^i L^i \varepsilon_t = \varepsilon_t + \sum_{i=1}^{\infty} \phi^i \varepsilon_{t-i}$$

La représentation est alors causale, en plus d'être stationnaire et inversible \implies elle est donc canonique. ε_t est le processus d'innovation de Y_t , puisque le passé de Y_t dépend du passé de ε_t et donc n'est pas corrélé à ε_t .

- Si $|z| < 1$ et donc $|\phi| > 1$ alors :

$$Y_t = (1 - \phi L)^{-1} \varepsilon_t = - \sum_{i=1}^{\infty} \phi^{-i} L^{-i} \varepsilon_t = - \left(\frac{\varepsilon_{t+1}}{\phi} + \frac{\varepsilon_{t+2}}{\phi^2} + \dots \right)$$

La forme $MA(\infty)$ n'est pas tournée vers le passé et **la représentation n'est donc pas canonique.**

Remarque : Cependant, dans ce cas, on peut toujours (tant que $|\phi| \neq 1$) se ramener à une représentation canonique, en changeant la représentation, c'est-à-dire en changeant de BB et en inversant les racines.

6.6.5 Représentation $MA(\infty)$

Proposition Tout processus $AR(p)$ de représentation canonique

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in Z,$$

admet la représentation $MA(\infty)$, i.e.

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}.$$

6.6.6 Caractéristiques des processus AR: autocorrélation simple et partielle

Corrélogramme simple :

On rappelle que la fonction d'autocorrélation simple est donnée par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{E(Y_t Y_{t-h})}{E(Y_t^2)}$$

Soit le processus $AR(p)$:

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t \quad \forall t \in Z$$

Pour calculer sa variance, on multiplie cette équation par Y_t et on en prend son espérance :

$$E(Y_t^2) - \sum_{i=1}^p \phi_i E(Y_t Y_{t-i}) = E(Y_t \varepsilon_t)$$

on obtient alors :

$$\gamma(0) = \sum_{i=1}^p \phi_i \gamma(i) + \sigma^2$$

Ainsi, puisque $\gamma(i) = \gamma(0)\rho(i)$

$$\gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \phi_i \rho(i)}$$

Pour obtenir les autocovariances, on multiplie l'équation (2) par Y_{t-h} et on en prend son espérance. On obtient alors les équations suivantes :

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \quad \forall h > 0$$

Les $p+1$ premières équations sont appelées équations de Yule-Walker. En divisant par $\gamma(0)$, on obtient alors :

$$\rho(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 0 \quad \forall h > 0$$

soit une équation de récurrence linéaire homogène d'ordre p dont le polynôme caractéristique est :

$$z^h - \sum_{i=1}^p \phi_i z^{h-i} = z^h (1 - \phi_1 z^{-1} - \dots - \phi_p z^{-p}) = 0$$

Si le polynôme $\Phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$ admet p racines distinctes $1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_p$, alors les racines de l'équation (3) sont $\lambda_1, \lambda_2, \dots, \lambda_p$ et le coefficient d'autocorrélation d'ordre h est alors donné par :

$$\rho(h) = a_1 (\lambda_1)^h + a_2 (\lambda_2)^h + \dots + a_p (\lambda_p)^h$$

où les $a_i, i = 1, \dots, p$ sont des constantes déterminées par les conditions initiales.

Puisque la représentation est canonique, les racines $1/\lambda_i, i = 1, \dots, p$ sont alors de module supérieur à 1 donc $|\lambda_i| < 1$.

Ainsi, la fonction d'autocorrélation (d'un processus stationnaire) décroît, soit de manière exponentielle (si les racines sont réelles), soit selon des cycles amortis (si les racines sont complexes).

Application sous R : Simulation de AR(p)

Résumé : principaux résultats :

1- Le développement précédent nous permet de constater que les autocorrélations d'un AR(p) **sous sa représentation canonique** vérifient une équation linéaire récurrente d'ordre p . Elles sont donc entièrement déterminées par la donnée des p premières et celles-ci vérifient **l'équation matricielle dite de Yule-Walker** :

$$\begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(p) \end{pmatrix} = R(p) \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_X(p-1) \\ \rho_X(1) & 1 & \ddots & \rho_X(p-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho_X(p-1) & \cdots & \rho_X(1) & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}$$

et

$$\gamma_X(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho_X(i)}.$$

2- Les autocorrélations simples d'un processus $AR(p)$ décroissent, de manière exponentielle ou sinusoidale amortie, vers 0.

3- Les autocorrélations partielles d'un processus $AR(p)$ sont nulles à partir du rang $p + 1$. Plus précisément, on a : $r_X(p) = \varphi_p \neq 0$ et $r_X(k) = 0$, pour $k > p$. (Démonstration (faite en exercice))

Démonstration : Exercice

On rappelle que le coefficient $r_X(k)$ est donné par le coefficient $\alpha_k(k)$ dans la projection de X_t sur \mathcal{H}_{t-k}^{t-1} , i.e.

$$EL(X_t | \mathcal{H}_{t-k}^{t-1}) = \alpha_1(k)X_{t-1} + \cdots + \alpha_k(k)X_{t-k}.$$

- En déduire que $r_X(p) = \varphi_p \neq 0$.
- En déduire également que $r_X(k) = 0$ dès que $k > p$.

Remarques :

- La propriété $r_X(p) = \varphi_p$ n'est vraie que dans le cas d'une représentation canonique.
- On peut montrer que la propriété d'autocorrélations partielles nulles à partir d'un certain rang $p + 1$ est caractéristique d'un $AR(p)$.

6.7 Les processus : $MA(q)$

6.7.1 Définition

On introduit la définition des processus MA (Moving Average).

On dit qu'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation Moyenne Mobile d'ordre q (noté $MA(q)$) s'il vérifie l'équation:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc $bb(0, \sigma^2)$ et $\theta_1, \dots, \theta_q$ sont des coefficients réels, avec $\theta_q \neq 0$. Autrement dit, en utilisant le polynôme en L

$$\Theta(L) = I + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q,$$

un processus $MA(q)$ vérifie l'équation :

$$X_t = \Theta(L)\varepsilon_t,$$

pour tout t dans \mathbb{Z} .

Cette représentation est dite canonique si les racines du polynôme $\Theta(z)$ sont toutes à l'extérieur du disque unité.

Remarques :

- Il n'existe qu'une seule représentation canonique, celle donc où toutes les racines du polynôme sont à l'extérieur du disque unité.
- Un processus MA , quelque soit sa représentation, est stationnaire, puisque obtenu par filtrage linéaire d'un bruit blanc, donc stationnaire.
- Un processus MA est toujours centré.

6.7.2 Représentation $AR(\infty)$

Proposition : Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $MA(q)$ de représentation canonique

$$X_t = \Theta(B)\varepsilon_t$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc $bb(0, \sigma^2)$. Le processus des innovations correspond alors au bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ de sa représentation canonique. De plus, le processus $(X_t)_{t \in \mathbb{Z}}$ possède une représentation $AR(\infty)$:

$$\varepsilon_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i}$$

avec $\pi_0 = 1$

6.7.3 Caractéristiques des processus MA : autocorrélation simple et autocorrélation partielle

Proposition (Autocorrélations simples d'un $MA(q)$)

Les autocorrélations simples d'un processus $MA(q)$ sont nulles à partir du rang $q + 1$. Plus précisément, on a : $\rho_X(q) \neq 0$ et $\rho_X(h) = 0$, pour $h > q$ et

$$\rho_X(h) = \frac{\theta_h + \sum_{i=1}^{q-h} \theta_i \theta_{i+h}}{1 + \sum_{j=1}^q \theta_j^2}.$$

Remarque : On peut montrer que la propriété d'autocorrélations simples nulles à partir d'un certain rang $q + 1$ est caractéristique d'un $MA(q)$.

Proposition (Autocorrélations partielles d'un $MA(q)$) :

Les autocorrélations partielles d'un processus $MA(q)$ sont solutions d'une équation linéaire récurrente d'ordre q . Elles décroissent, de manière exponentielle ou sinusoidale amortie, vers 0.

6.8 Processus ARMA

Nous allons maintenant introduire un modèle de processus stationnaire comportant une partie AR et une partie MA. C'est pourquoi il porte le nom de processus ARMA (AutoRegressive Moving Average). Les processus ARMA sont très importants en pratique car on peut montrer que tout processus stationnaire peut être approché par un processus ARMA.

6.8.1 Définition

On dit qu'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $ARMA(p, q)$ s'il vérifie l'équation :

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc $bb(0, \sigma^2)$ et $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ sont des coefficients réels, avec $\varphi_p \neq 0$ ainsi que $\theta_q \neq 0$. Autrement dit, en utilisant les polynômes en L

$$\Phi(L) = I - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$$

et

$$\Theta(L) = I + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q,$$

un processus $ARMA(p, q)$ vérifie l'équation:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t,$$

pour tout t dans \mathbb{Z} .

Naturellement un processus $AR(p)$ est un $ARMA(p, 0)$ et un processus $MA(q)$ est un $ARMA(0, q)$.

Définition La représentation d'un processus $ARMA(p, q)$

$$\Phi(L)X_t = \Theta(L)\varepsilon_t,$$

est dite :

- **minimale** si les polynômes $\Phi(z)$ et $\Theta(z)$ n'ont pas de racine commune ;
- **causale** si le polynôme $\Phi(z)$ a toutes ses racines à l'extérieur du disque unité ;
- **inversible** si le polynôme $\Theta(z)$ a toutes ses racines à l'extérieur du disque unité ;
- **canonique** si elle est causale et inversible.

6.8.2 Écritures $MA(\infty)$ et $AR(\infty)$

Proposition (Écriture $MA(\infty)$). Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ de représentation minimale et causale. Il admet alors la représentation $MA(\infty)$

$$X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i},$$

où les coefficients $(\psi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\psi_i - \sum_{j=1}^p \varphi_j \psi_{i-j} = \theta_i, \text{ pour } i \in \mathbb{N},$$

avec $\psi_i = 0$, pour $i < 0$, $\psi_0 = 1$, $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$.

6.8.3 Proposition (Écriture $AR(\infty)$)

. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ de représentation minimale et inversible. Il admet alors la représentation $AR(\infty)$

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i},$$

où les coefficients $(\pi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\pi_i + \sum_{j=1}^q \theta_j \pi_{i-j} = -\varphi_i, \text{ pour } i \in \mathbb{N},$$

avec $\pi_i = 0$, pour $i < 0$, $\varphi_0 = -1$ et $\varphi_i = 0$ pour $i > p$.

6.8.4 Caractéristiques temporelles du modèle ARMA(p,q) : auto-corrélation simple et autocorrélation partielle

Exercice : On peut obtenir les autocorrélations simples d'un processus $ARMA(p, q)$ de deux expressions différentes, l'une en utilisant l'écriture $ARMA(p, q)$, l'autre en utilisant l'écriture $MA(\infty)$.

1- En utilisant l'écriture $ARMA(p, q)$, montrer que l'on a

$$\gamma_X(h) - \varphi_1 \gamma_X(h-1) - \dots - \varphi_p \gamma_X(h-p) = Cov(\varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \theta_2 \varepsilon_{t+h-2} + \dots + \theta_q \varepsilon_{t+h-q}, \varepsilon_t) + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}.$$

2- En déduire, pour $h > q$, la relation de récurrence :

$$\gamma_X(h) - \varphi_1 \gamma_X(h-1) - \dots - \varphi_p \gamma_X(h-p) = 0.$$

3- Donner l'expression de la relation de récurrence pour les valeurs $0 \leq h \leq q$.

4- Quelle expression de $\gamma_X(h)$ obtient-on en utilisant cette fois l'écriture $MA(\infty)$ du processus $(X_t)_{t \in \mathbb{Z}}$.

- On montre ainsi que les autocorrélations simples décroissent de manière exponentielle ou sinusoidale amortie vers 0 avec h . On peut montrer le même genre de résultat pour les autocorrélations partielles.
- On constate aussi, qu'à la différence des cas particuliers des processus $AR(p)$ ou $MA(q)$, il n'existe pas de caractérisation aisée pour les modèles $ARMA(p, q)$. Les autocorrélations simples ou partielles ne s'annulent pas à partir d'un certain rang.

autocovariance d'un $ARMA(p, q)$ si X_t est un processus stationnaire de représentation $ARMA(p, q)$ minimale:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

sa fonction d'autocovariance vérifie:

$$\gamma(h) + \sum_{j=1}^p \phi_j \gamma(h-j) = 0, \quad \forall h \geq q+1$$

pour obtenir les valeurs initiales des covariances on exploite le développement $MA(\infty)$:

$$X_t = \sum_{j=0}^{+\infty} h_j \varepsilon_{t-j}$$

et la définition de l'ARMA(p,q):

$$X_t + \sum_{j=1}^p \phi_j X_{t-j} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

En multipliant par X_{t-h} et en prenant l'espérance on a:

$$\gamma(h) + \sum_{j=1}^p \phi_j \gamma(h-j) = \text{cov} \left(\varepsilon_t + \sum_{j=1}^q \varepsilon_{t-j}, \sum_{j=0}^{+\infty} h_j \varepsilon_{t-j} \right)$$

et donc:

$$\gamma(h) + \sum_{j=1}^p \phi_j \gamma(h-j) = \sigma^2 (\theta_h + h_1 \theta_{h+1} + \dots + h_{q-h} \theta_q), \quad 0 \leq h \leq q$$

6.8.5 Résumé

<i>Type de processus</i>	<i>Définition</i>	<i>autocorrélations</i>	<i>autocorrélations partielles</i>
<i>AR(p)</i>	$\Phi(L)X_t = \varepsilon_t$	$\rho(h) \searrow 0$	$r(h) = 0$ pour $h \geq p+1$
<i>MA(q)</i>	$X_t = \Theta(L)\varepsilon_t$	$\rho(h) = 0$ pour $h \geq q+1$	$r(h)$ rien de particulier
<i>ARMA(p, q)</i>	$\Phi(L)X_t = \Theta(L)\varepsilon_t$	$\rho(h) \searrow 0, h \geq q+1$	$r(h) \searrow 0, h \geq \max(q+1, p+1)$

6.9 Processus ARIMA

Les processus étant dans la pratique rarement stationnaires, on a introduit une généralisation des processus ARMA vus précédemment de manière à les étendre à des processus non stationnaires. L'idée générale, essentiellement conçue pour les processus non stationnaires à tendance polynomiale, est de différencier suffisamment le processus initial afin d'obtenir un processus sans tendance et sur la partie différenciée appliquer un modèle ARMA. Cette classe de modèle est connue sous le nom de processus ARIMA (AutoRegressive Integrated Moving Average).

6.9.1 Définition

On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $ARRMA(p, d, q)$ s'il vérifie l'équation :

$$\Phi(B) \nabla^d X_t = \Theta(B) \varepsilon_t,$$

pour tout t dans \mathbb{Z} , où

$$\begin{aligned} \nabla^d &= (I - B)^d \\ \Phi(B) &= I - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \\ \Theta(B) &= I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \end{aligned}$$

6.10 Estimation, choix de modèle et prévisions

La marche à suivre est la suivante.

- Choisir un modèle.
- Estimer les paramètres.
- Vérifier que le modèle est crédible.
- Faire des prévisions.

6.10.1 Estimation (sur un exemple)

. On suppose que $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t$ (avec des ϵ_t bruits blancs centrés, de variance σ^2) et que (X_t) est stationnaire. On calcule les autocovariances :

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

D'où

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

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 $\sigma(0) = \sigma^2 + a_1 a_2 \sigma(1) + (a_1^2 + a_2^2) \sigma(0)$, on peut aussi estimer σ^2 .

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

6.10.2 Choix du modèle.

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

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

où ν est le nombre de paramètres, Θ est un vecteur contenant les paramètres, n est le nombre d'observations, L est la vraisemblance (dans laquelle on a omis les observations).

6.11 Stratégie du test de stationnarité de Dickey-Fuller

6.11.1 Les tests de la non stationnarité

Il existe plusieurs tests relatifs à la détection de la non stationnarité. Nous aborderons les plus utilisés dans les travaux empiriques, à savoir les tests présentés par Dickey et Fuller, soit sous la forme simple ou sous la forme augmentée, ces deux tests représentent le premier fruit des recherches sur les tests de la non stationnarité. Additivement aux tests cités au-dessus on trouve le test de Phillips et Perron.

6.11.2 Le test de Dickey-Fuller simple et sa stratégie d'application

En 1979 Dickey et Fuller ont présenté un test de la non stationnarité (de racine unitaire) noté DF qui est largement utilisé et répandu est le test de racine unitaire. Ce test est fondé sur l'estimation de trois modèles, chacun représentant un processus AR d'ordre 1 mais avec des formes différentes de la tendance déterministe (nulle, constante ou linéaire), ces modèles sont les suivants: les suivants:

1. $\Delta Y_t = \phi Y_{t-1} + \mu_t$ (01)
2. $\Delta Y_t = \phi Y_{t-1} + c + \mu_t$ (02)
3. $\Delta Y_t = \phi Y_{t-1} + c + \beta t + \mu_t \dots\dots\dots$ (03)

Ce test adopte comme hypothèse nulle la présence de racine unitaire, soit la non stationnarité stochastique. Le test consiste à tester : l'hypothèse nulle $H_0 : \phi = 0$ (présence de racine unitaire, non stationnarité) contre l'hypothèse alternative $H_1 : \phi < 0$ (absence de racine unitaire, stationnarité).

$H_0 : \phi = 0$ Présence de racine unitaire, non stationnaité
 $H_1 : \phi < 0$ Absence de racine unitaire, stationnnaité

Le principe général de la stratégie de ce test est le suivant :

Première phase: On commence par tester la racine unitaire à partir du modèle le plus général, à savoir le troisième modèle incluant une constante et une tendance. On teste alors la présence d'une racine unitaire dans le processus en testant la nullité du paramètre ϕ de la variable endogène retardée, à l'aide d'une statistique de Student $t_{\hat{\phi}}$, on compare la réalisation de cette statistique avec la valeur critique VC^3 tabulés par Dickey & Fuller ou McKinnon. Dans la mesure où les valeurs critiques sont négatives, La règle de décision est la suivante:

- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est supérieure à la valeur critique VC^3 on accepte l'hypothèse nulle H_0 de non stationnarité et on rejette l'hypothèse alternative H_1 de stationnarité.
- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est inférieure à

la valeur critique VC^3 on rejette l'hypothèse nulle H_0 de non stationnarité et on accepte l'hypothèse alternative H_1 de stationnarité

On passe à vérifier par un test approprié que le modèle 3 retenu est le bon modèle c-à-d. on cherche à vérifier que si la spécification du modèle 3 était une spécification compatible avec les données. La vérification que le modèle 3 est le modèle approprié se passe comme suit :

1. Cas où $H_0 : \phi = 0$ est rejetée: rejeter l'hypothèse de racine unitaire

Dans ce cas on teste la nullité du coefficient β de la tendance par un simple test de student ca-d. tester l'hypothèse nulle $H_0^2 : \beta = 0$ (l'absence de la tendance) contre l'hypothèse alternative $H_1^2 : \beta \neq 0$ (la présence de la tendance).

- Si l'on accepte l'hypothèse nulle H_0^2 , cela signifie que la présence de la tendance est rejeter, donc le modèle 3 n'est pas le modèle approprié pour tester la présence de la racine unitaire. C'est pour cela on doit refaire le test de la racine unitaire à partir du modèle 2 qui contient qu'une constante, c-à-d. aller à la deuxième phase
- Si l'on rejette l'hypothèse nulle H_0^2 , cela signifie que le modèle 3 est le bon modèle pour tester la racine unitaire, puisque la présence de la tendance n'est pas rejeter. Dans ce cas on conclut que la racine unitaire est rejetée, la série est de type TS, du fait de la présence de la tendance.

2. Cas où $H_0 : \phi = 0$ est acceptée: accepter l'hypothèse de racine unitaire

Dans ce cas on teste ainsi la nullité de la tendance, conditionnellement à la présence de racine unitaire, par un test de Fisher de l'hypothèse jointe $\phi = 0$ et $\beta = 0$, c-a-d. tester l'hypothèse nulle $H_0^3 : (c, \beta, \phi) = (c; 0, 0)$ contre l'hypothèse alternative H_1^3 . la statistique de test notée F_{Cal}^3 se calcule suivant la relation suivante:

$$F_{Cal}^3 = \frac{(SCR_{3,c} - SCR_3) / 2}{SCR_3 / (T - 3)}$$

Tel que:

- SCR_3 : représente la somme carré des résidus du modèle 3 non contraint, c-à-d. le modèle $\Delta Y_t = \phi Y_{t-1} + c + \beta t + \mu_t$
- $SCR_{3,c}$: représente la somme carré des résidus du modèle 3 contraint sous H_0^3 , c-à-d. le modèle $\Delta Y_t = c + \mu_t$

- Si $F_{Cal}^3 \leq F_{Tab}^{a\%}$ on accepte l'hypothèse nulle H_0^3 , le coefficient β de la tendance est nul, donc le modèle 3 n'est pas le bon pour tester la présence de la racine unitaire. C'est pour cela on doit effectuer à nouveau le test de la racine unitaire à partir du modèle 2c-à-d. aller à la deuxième phase

- Si $F_{Cal}^3 > F_{Tab}^{a\%}$ on rejette l'hypothèse nulle H_0^3 , le coefficient de la tendance n'est pas nul, cela signifie que le modèle 3 est le modèle approprié pour tester la racine unitaire, et la série est intégrée d'ordre 1.

Remarque: Uniquement lorsque le coefficient β de la tendance est nul, on passe à la deuxième phase et refaire à nouveau le test de non stationnarité à partir du modèle 2 qui comporte uniquement une constante.

Deuxième phase (étape): On teste présence de la racine unitaire (la non stationnarité) dans le processus à partir du modèle 2 incluant uniquement une constante. On teste alors la nullité du paramètre ϕ à l'aide d'une statistique de Student $t_{\hat{\phi}}$, on compare la réalisation de cette statistique avec la valeur critique VC^2 tabulés par Dickey & Fuller ou McKinnon. La règle de décision est la suivante:

- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est supérieure à la valeur critique VC^2 on accepte l'hypothèse nulle H_0 de non stationnarité et on rejette l'hypothèse alternative H_1 de stationnarité.
- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est inférieure à la valeur critique VC^2 on rejette l'hypothèse nulle H_0 de non stationnarité et on accepte l'hypothèse alternative H_1 de stationnarité.

On passe à vérifier par un test approprié que le modèle 2 retenu est le bon modèle c-à-d. on cherche à vérifier que si la spécification du modèle 2 était une spécification compatible avec les données. La vérification que le modèle 2 est le modèle approprié se passe comme suit :

1. Cas où $H_0 : \phi = 0$ est rejetée: rejeter l'hypothèse de racine unitaire.

Dans ce cas on teste la nullité de la constante c par un simple test de student c-à-d. tester l'hypothèse nulle $H_0^2 : c = 0$ contre l'hypothèse alternative $H_1^2 : c \neq 0$.

- Si l'on accepte l'hypothèse nulle H_0^2 , cela signifie que la présence de la constante est rejeter, donc le modèle 2 n'est pas le modèle approprié pour tester la présence de la racine unitaire. C'est pour cela on doit refaire le test de la racine unitaire à partir du modèle 1 qui ne contient ni constante ni tendance, c-à-d. aller à la troisième phase
- Si l'on rejette l'hypothèse nulle H_0^2 , cela signifie que le modèle 2 est le bon modèle et le modèle approprié pour tester la racine unitaire, puisque la présence de la constante n'est pas rejeter. Dans ce cas on conclut que la racine unitaire est rejetée, la série est stationnaire intégrée d'ordre zéro,

$$I(0) + c.$$

2. Cas où $H_0 : \phi = 0$ est acceptée: accepter l'hypothèse de racine unitaire

Dans ce cas on teste ainsi la nullité de la constante c , conditionnellement à la présence de racine unitaire, par un test de Fisher de l'hypothèse jointe $\phi = 0$ et $c = 0$, c-à-d. tester l'hypothèse nulle $H_0^2 : (c; \phi) = (0, 0)$ contre l'hypothèse alternative H_1^2 . La statistique de test notée F_{Cal}^2 se calcule suivant la relation suivante:

$$F_{Cal}^2 = \frac{(SCR_{2,c} - SCR_2) / 2}{SCR_2 / (T - 2)}$$

Tel que:

- SCR_2 : représente la somme carré des résidus du modèle 2 non contraint, c-à-d. le modèle $\Delta Y_t = \phi Y_{t-1} + c + \mu_t$
- $SCR_{2,c}$: représente la somme carré des résidus du modèle 2 contraint sous H_0^2 , c-à-d. le modèle $\Delta Y_t = c + \mu_t$

- Si $F_{Cal}^2 \leq F_{Tab}^{\alpha\%}$ on accepte l'hypothèse nulle H_0^2 , la constante c est nul, donc le modèle 2 n'est pas le bon pour tester la présence de la racine unitaire. C'est pour cela on doit effectuer à nouveau le test de la racine unitaire à partir du modèle 1c-à-d. aller à la troisième phase
- Si $F_{Cal}^2 > F_{Tab}^{\alpha\%}$ on rejette l'hypothèse nulle H_0^2 , la constante c n'est pas nul, cela signifie que le modèle 2 est le modèle approprié pour tester la racine unitaire, et la série est intégrée du premier ordre, $I(1) + c$.

Troisième phase: On test la non stationnarité (la présence de racine unitaire) à partir du modèle 1 sans tendance et sans constante. On teste alors la nullité du paramètre ϕ à l'aide d'une statistique de Student $t_{\Phi\Phi}$, on compare la réalisation de cette statistique avec la valeur critique VC^1 tabulés par Dickey & Fuller ou McKinnon.

- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est supérieure à la valeur critique VC^1 on accepte l'hypothèse nulle H_0 de non stationnarité. Dans ce cas le processus est non stationnaire de type DS, la série est intégrée d'ordre 1 $Y_t \rightarrow I(1)$ qui correspond à une pure marche aléatoire $Y_t = Y_{t-1} + \mu_t$.
- Si la valeur calculée de la t-statistique associée à ϕ c-à-d. $t_{\hat{\phi}}$ est inférieure à la valeur critique VC^1 on rejette l'hypothèse nulle H_0 . Dans ce cas le processus est stationnaire, la série est intégrée d'ordre 0, $Y_t \rightarrow I(0)$, $Y_t = \rho Y_{t-1} + \mu_t$ avec $|\rho| < 1$.

Le déroulement de la stratégie du test est reporté sur la figure suivante:

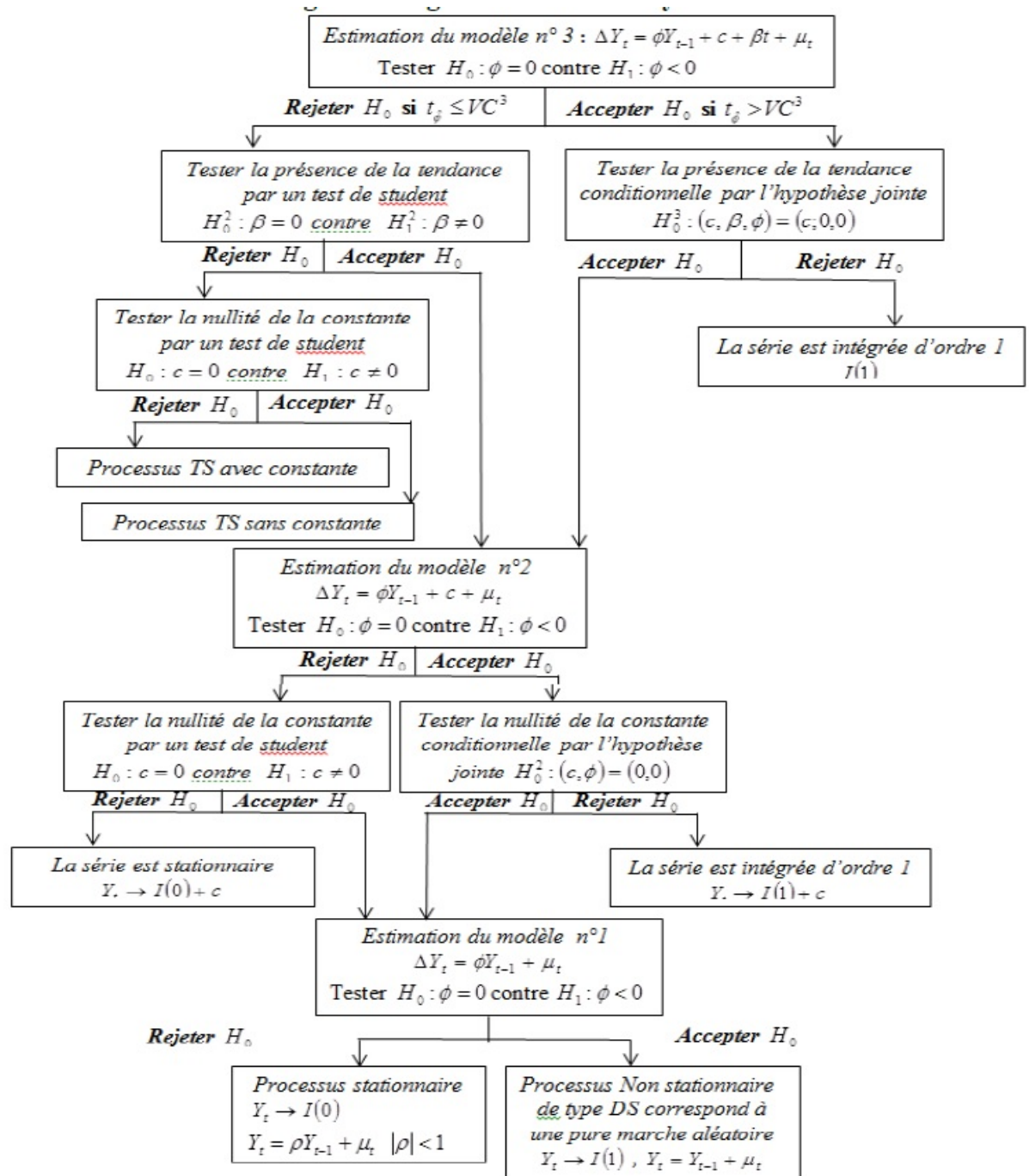


Figure 1: Stratégie du test de Dickey-Fuller

6.11.3 Le test de Dickey-Fuller augmenté et sa stratégie

Il arrive parfois que les résidus ε_t du modèle de Dickey - Fuller soient autocorrélés, donc il est nécessaire de tenir compte de l'éventuelle autocorrélation de ces résidus lors de la construction des tests de racine unitaire. Parmi les approches qui prennent en considération cette éventuelle autocorrélation, on cite l'approche développée et proposée en 1988 par Pierre Perron et Peter Phillips et celle proposée par Dickey et Fuller en 1979 nommée test de Dickey - Fuller Augmenté noté ADF. Ce dernier permet de tester la stationnarité de la série prenant en compte l'autocorrélation des perturbations ε_t . Donc Le test de Dickey-Fuller Augmenté est une version améliorée du test de Dickey-Fuller simple, par l'introduction dans les modèles du test des valeurs retardées de la série destinées à corriger une éventuelle autocorrélation du terme d'erreur. Le test ADF est fondé sur l'estimation des trois modèles suivants:

1. $\Delta Y_t = \phi Y_{t-1} + \sum_{j=1}^p \rho_j \Delta Y_{t-j} + \mu_t$ (04)
2. $\Delta Y_t = \phi Y_{t-1} + \sum_{j=1}^p \rho_j \Delta Y_{t-j} + c + \mu_t$ (05)
3. $\Delta Y_t = \phi Y_{t-1} + \sum_{j=1}^p \rho_j \Delta Y_{t-j} + c + \beta t + \mu_t \dots\dots\dots$ (06)

La stratégie du test de Dickey Fuller Augmenté est strictement identique à celle du test de Dickey Fuller Simple. On applique la stratégie séquentielle du test de Dickey Fuller Simple aux modèles (04), (05) et (06) en commençant par le modèle (06) bien sûr. L'application de ce test nécessite en première phase le choix du nombre de retard optimal.

Le choix du nombre de retards optimal

Le choix du nombre de retard optimal est très important lors de l'application du test de Dickey - Fuller Augmenté, car l'inclusion d'un nombre insuffisant de retards peut affecter le niveau du test, tandis que l'introduction d'un nombre trop-élevé de retards réduit le nombre de degrés de liberté et donc la puissance du test, c-à-d. qu'il y a une racine unitaires alors que notre processus est stationnaire.

Pour déterminer le nombre optimal de retard noté P , nous procédons à la minimisation des deux critères d'informations, le critère d'Akaike et le critère de Schwarz.

1. k_{\max} est fonction de la taille d'échantillon n . Pour implémenter le test de Dickey et Fuller augmenté, une règle empirique utile pour déterminer k_{\max} , suggérée par Schwert (1989), est donnée par

$$k_{\max} = \left\lceil 12 \left(\frac{n}{100} \right)^{1/4} \right\rceil$$

où $[x]$ désigne la partie entière de x . Ce choix permet à k_{\max} de croître avec la taille de l'échantillon.

Remarque La règle précédente nous renseigne sur l'ensemble des valeurs possibles de l'ordre k . C'est-à-dire, une fois la valeur k_{\max} calculée, on doit chercher la valeur de k dans l'ensemble $\{0, 1, 2, \dots, k, \dots, k_{\max}\}$.

2. La règle du choix de k est basée sur les critères d'information

Cette règle consiste à choisir k de façon à minimiser une fonction objective qui est de la forme :

$$IC_k = \log \hat{\sigma}_k^2 + (k + nr + 1) \frac{C_n}{n - k_{\max}}$$

avec $\hat{\sigma}_k^2$ la variance des résidus et (nr) le nombre de régresseurs. Les critères d'information les plus fréquemment utilisés sont le critère d'information de Akaike [1969] (AIC) qui fixe C_n à 2, le critère d'information de Schwarz [1978] (BIC) qui fixe C_n à $\log(n - k_{\max})$ et le critère d'information de Hannan et Quinn [1979] (HQ) qui fixe C_n à $2n \log\left(\frac{\log n}{n}\right)$ où $b > 2$ est une constante.

Exercices

Exercice 1:

On considère le modèle suivant :

$Y_t = \beta_0 + \beta_1 t + a_t, \forall t = 1 \dots T$, avec $\{a_t\}$ est une séquence de variables aléatoires vérifiant : $E(a_t) = 0$ et $var(a_t) = \sigma^2$.

Le processus Y_t est-il stationnaire ?

Exercice 2 : On considère le processus suivant :

$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$ avec ε_t est un BB de loi $N(0, \sigma_\varepsilon^2)$.

- 1- Déterminer l'expression de $E(y_t)$ en fonction de φ_0 et φ_1 .
- 2- Déterminer l'expression de $\gamma_0, \gamma_1, \gamma_2$ et $\gamma_k \forall k \in \mathbb{N}^*$
- 3- En déduire les expressions de ρ_1, ρ_2 et $\rho_k \forall k \in \mathbb{N}^*$
- 4- Reprendre les questions 1, 2 et 3 pour un processus AR(2) admettant comme équation $y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$

Exercice 3 : On considère le processus suivant :

$y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t$ avec ε_t est un BB de loi $N(0, \sigma_\varepsilon^2)$.

- 1- Calculer $E(y_t)$
- 2- Déterminer l'expression de $\gamma_0 \gamma_1$ et $\gamma_k \forall k > 1$
- 3- En déduire les expressions de ρ_1, ρ_2 et $\rho_k \forall k > 1$
- 4- Reprendre les questions 1, 2 et 3 pour un processus MA(2)

Exercice 4 : On considère le processus suivant :

$y_t = 2 + \varepsilon_t - 0.8\varepsilon_{t-1} - 0.7y_{t-2}$, avec ε_t suit $N(0, \sigma_\varepsilon^2)$

- 1- Préciser la nature de ce processus.
- 2- Tracer le corrélogramme correspondant à ce processus pour $h = 1, 2, 3, 4, 5$.

Exercice 5 : On considère le processus suivant :

$y_t = 0.9y_{t-1} - 0.7y_{t-2} + \varepsilon_t$, avec ε_t suit $N(0, \sigma_\varepsilon^2)$

1- Préciser la nature de ce processus.

2- Tracer le corrélogramme correspondant à ce processus pour $h = 1, 2, 3, 4, 5$.

Exercice 6 : On considère le processus suivant :

$y_t = 2 + 0.5y_{t-1} + \varepsilon_t$, avec ε_t suit $N(0, \sigma_\varepsilon^2)$

1- Préciser la nature de ce processus.

2- Calculer $E(y_t)$.

3- Tracer le corrélogramme correspondant à ce processus pour $h = 1, 2, 3, 4$.

Exercice 7 : Etudier les conditions de stationnarité et d'inversibilité des processus suivants :

1) $y_t = 0.9y_{t-1} + \varepsilon_t$

2) $y_t = \varepsilon_t - 0.8\varepsilon_{t-1}$

3) $y_t = \varepsilon_t - 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$

4) $y_t = 0.8y_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-1}$

Exercice 8 :

Soit $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ avec $|\theta| < 1$.

1) Calculer le coefficient d'autocorrélation d'ordre 1.

2) Etudier la corrélation entre X_t et X_{t-1} .

Exercice 9 :

Soit : $X_t = \frac{1}{6}X_{t-1} + \frac{1}{6}X_{t-2} + \varepsilon_t$.

1) Le processus $\{X_t\}$ est-il stationnaire ?

2) Calculer le coefficient d'autocorrélation ρ_1 .

3) Ecrire les équations de Yule-Walker et calculer les coefficients d'autocorrélation d'ordre 2 et 3.

Exercice 10 : Soit (Z_t) un bruit blanc. Pour chacune des équations suivantes, existe-t-il un processus stationnaire (X_t) solution?

1. $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$

2. $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$;

3. $X_t + 0.6X_{t-2} = Z_t + 1.2Z_{t-1}$;

4. $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$.

Solution de l'exercice 10.

1. L'équation $1 + 0.2z - 0.48z^2 = 0$ a pour solutions $-5/4$ et $5/3$, qui ne sont pas de module 1 donc il existe un processus stationnaire solution ;

2. L'équation $1 + 1.9z + 0.88z^2 = 0$ a pour solutions $-10/11$ et $-5/4$, qui ne sont pas de module 1 donc il existe un processus stationnaire solution;

3. L'équation $1 + 0.6z = 0$ a pour solution $-5/3$ qui n'est pas de module 1 donc il existe une solution stationnaire;

4. L'équation $1 + 1.8z + 0.81z^2 = 0$ a pour solution $-10/9$ (racine double) qui n'est pas de module 1 donc il existe une solution stationnaire.

Exercice 11 Soit (ε_t) un bruit blanc (faible) et (X_t) un processus stationnaire au second ordre, vérifiant la relation de récurrence

$$X_t = -0.4X_{t-1} + 0.12X_{t-2} + \varepsilon_t.$$

- (1) de quel modèle ARMA s'agit-il ?
- (2) montrer que (ε_t) est l'innovation de (X_t) , et que la fonction d'autocovariance de (X_t) , notée $\rho(\cdot)$, vérifie une relation de récurrence de la forme

$$\rho(h) = a\rho(h-1) + b\rho(h-2), \text{ pour } h \geq 2,$$

où a et b sont des constantes à préciser.

- (3) donnez la variance de (ε_t) pour que la variance de (X_t) soit unitaire.

- (4) montrer que

$$\rho(h) = \frac{2}{11}[0.2]^h + \frac{9}{11}[-0.6]^h \text{ pour } h \geq 0.$$

Solution de l'exercice 11

- (1) C'est un AR(2), ou ARMA(2,0).

- (2) Dans un processus $AR(p)$, le bruit est le processus d'innovation dès lors que la racines du polynôme retard sont à l'extérieur du disque unité. Or ici le polynôme est

$$\Phi(L) = (1 + 0.4L - .12L^2) = (1 + 0.6L)(1 - 0.2L)$$

dont les racines sont $-1/0.6$ et $1/0.2$, qui sont plus grandes que 1 (en valeur absolue).

Pour les autocorrélations, soit $h \geq 2$. Notons que le processus est centré. Aussi, en notant γ la fonction d'autocovariance, $\gamma(h) = E(X_t X_{t-h})$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Maintenant, si on multiplie l'équation de récurrence par X_{t-h} ,

$$X_t X_{t-h} = -0.4X_{t-1}X_{t-h} + 0.12X_{t-2}X_{t-h} + \varepsilon_t X_{t-h}$$

or comme (ε_t) est le processus d'innovation, $E(\varepsilon_t X_{t-h}) = 0$. Et donc, en prenant l'espérance des termes de l'équation précédente, on a

$$\gamma(h) = -0.4\gamma(h-1) + 0.12\gamma(h-2), \text{ pour } h \geq 2.$$

soit, en divisant par $\gamma(0)$,

$$\rho(h) = -0.4\rho(h-1) + 0.12\rho(h-2), \text{ pour } h \geq 2.$$

(3) On s'était limité au cas $h \geq 2$ dans la question précédente. Regardons $h = 1$, en multipliant la relation de récurrence par X_{t-1} ,

$$X_t X_{t-1} = -0.4 X_{t-1} X_{t-1} + 0.12 X_{t-2} X_{t-1} + \varepsilon_t X_{t-1}.$$

Si on prend l'espérance, le terme de droite va disparaître (comme auparavant) et on obtient

$$\gamma(1) = -0.4\gamma(0) + 0.12\gamma(1)$$

de telle sorte que

$$\rho(1) = \frac{-0.4}{1 - 0.12} = \frac{-5}{11}.$$

Maintenant, pour calculer la variance de X_t , notons que

$$E(X_t^2) = E([-0.4X_{t-1} + 0.12X_{t-2} + \varepsilon_t]^2)$$

On va alors développer le terme de droite. On va passer les détails, mais on se souvient que (ε_t) est le processus d'innovation, et donc il est orthogonal au passé de X_t . Aussi,

$$E(X_t^2) = [0.4^2 + 0.12^2] E(X_t^2) - 0.4 \cdot 0.12 E(X_t X_{t-1}) + 0 + E(\varepsilon_t^2)$$

avec $E(X_t X_{t-1}) = \gamma(1) = \rho(1)E(X_t^2)$, avec $\rho(1)$ que l'on vient de calculer. Aussi, la variance de X_t vérifie

$$\left[1 - [0.4^2 + 0.12^2] - 0.4 \cdot 0.12 \cdot \frac{-5}{11} \right] \text{var}(X_t) = \text{var}(\varepsilon_t).$$

soit (pour utiliser une formule qu'on peut retrouver dans des cours sur les AR(2))

$$\text{var}(X_t) = \frac{1 - 0.12}{1 + 0.12} \frac{1}{(1 - 0.12)^2 - 0.4^2} \text{var}(\varepsilon_t).$$

Peu importe la valeur numérique, je donne les points pour ceux qui ont noté qu'on pouvait effectivement extraire la variance du processus dans le cas d'un AR(2).

(4) On sait que la solution générale de la suite définie par récurrence,

$$\rho(h) = -0.4\rho(h-1) + 0.12\rho(h-2), \text{ pour } h \geq 2$$

est de la forme

$$\rho(h) = Ar_1^h + Br_2^h, \text{ pour } h \geq 0$$

où r_1 et r_2 sont les racines distinctes du polynôme caractéristique. Or on a calculé les racines du-dit polynôme dans la question (2). Aussi,

$$\rho(h) = A[0.2]^h + B[-0.6]^h, \text{ pour } h \geq 0$$

On obtient les valeurs de A et B avec les premières valeurs, $\rho(0) = 1 = A + B$ et $\rho(1) = -5/11 = 0.2A - 0.6B$. On a ainsi un système (linéaire) à résoudre.... Je passe ici les détails, mais on peut vérifier rapidement que la solution proposée est la seule qui marche.

Exercice 12 On considère l'équation

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

où (Z_t) est un bruit blanc de moyenne nulle et de variance σ^2 et ϕ et θ sont des réels.

1. À quelles conditions sur ϕ et θ existe-t-il un processus (X_t) stationnaire solution de l'équation ci-dessus?
2. À quelles conditions cette solution est-elle causale? inversible? Dans la suite on supposera que ces conditions sont vérifiées.
4. Calculer la fonction d'autocorrélation de (X_t) .

Solution de l'exercice 12.

1. Une solution stationnaire existe (elle est alors unique, et c'est un processus linéaire) lorsque le polynôme $\Phi(z) = P_{\alpha\phi}(z) = 1 - \phi z$ n'a pas de racine de module 1, c'est-à-dire lorsque $|\phi| \neq 1$ puisque Φ a une unique racine $z_1 = 1/\phi$
2. La solution stationnaire est causale lorsque Φ n'a pas de racine de module ≤ 1 , c'est-à-dire lorsque $|z_1| > 1$, c'est-à-dire $|\phi| < 1$. La solution stationnaire est inversible lorsque $\theta(z) = 1 + \theta z$ n'a pas de racine de module ≤ 1 , c'est-à-dire lorsque l'unique racine $-1/\theta$ de θ est de module > 1 , c'est-à-dire lorsque $|\theta| < 1$;

Exercice 13. On considère le processus (X_t) solution de

$$(1 - B + B^2/4) X_t = (1 + B)Z_t,$$

où (Z_t) est un bruit blanc de variance σ^2 .

1. Montrer que l'on peut écrire X_t sous la forme $X_t = \sum_{k \geq 0} \psi_k Z_{t-k}$.
2. Calculer les coefficients $(\psi_k)_{k \geq 0}$.
3. Calculer la fonction d'autocovariance de (X_t) .

Solution de l'exercice 13

1. On a $P_\varphi(z) = 1 - z + z^2/4 = (1 - z/2)^2$ qui ne s'annule qu'en $z = 2$ (racine double). Comme cette racine est de module > 2 , il en découle que l'équation ARMA ci-dessus admet une unique solution stationnaire, qui est causale ;

2. Les $(\psi_k)_{k \geq 0}$ s'obtiennent en développant en série de puissance de z autour du cercle unité la fraction rationnelle de l'ARMA : (ici $|z/2| < 1$ si $|z| = 1$) $\frac{P_\theta(z)}{P_\varphi(z)} = \frac{1+z}{(1-z/2)^2} = (1+z) \sum_{k \geq 0} (k+1)2^{-k} z^k = 1 + \sum_{k \geq 1} \underbrace{\left((k+1)2^{-k} + k2^{-(k-1)} \right)}_{(3k+1)2^{-k}} z^k$

d'où $\psi = \mathbf{1}_{k=0} + (3k+1)2^{-k} \mathbf{1}_{k>0}$. Note : $(1-u)^{-2} = \sum_{k \geq 0} (k+1)u^k$ si $|u| < 1$;

Exercice 14 Soit $Z \sim \text{BB}(0, \sigma^2)$.

1. Préciser les polynômes Φ et Θ de l'équation $\text{ARMA}(1, 2)$

$$X_t - 3X_{t-1} = Z_t - \frac{10}{3}Z_{t-1} + Z_{t-2}$$

et montrer qu'il existe une unique solution stationnaire;

2. Cette solution est-elle causale? Est-ce que Φ s'annule sur le disque unité?

3. Calculer explicitement cette solution en fonction de Z ;

4. Cette solution est-elle inversible?

Solution de l'exercice 14.

1. On a $\Phi(z) = 1 - 3z$ et $\Theta(z) = 1 - (10/3)z + z^2$. Le polynôme Φ a une seule racine, égale à $1/3$. Il ne s'annule donc pas sur le cercle unité, et donc, d'après le cours, l'équation $\text{ARMA}(1, 2)$ admet une unique solution stationnaire, qui est un processus linéaire filtre de Z .

2. Le polynôme Φ s'annule sur le disque unité. Comme l'absence d'annulation sur le disque unité du polynôme Φ est une condition suffisante mais pas forcément nécessaire de causalité, on ne peut pas en déduire que la solution n'est pas causale.

La solution s'obtient en développant en série de puissances de z autour du cercle unité la fraction rationnelle de l'équation. Or on remarque que $1/3$ est également racine de Θ , et pour tout $z \in C$ avec $|z| = 1$,

$$\frac{\Theta(z)}{\Phi(z)} = \frac{(1 - 3z)(1 - \frac{1}{3}z)}{1 - 3z} = 1 - \frac{1}{3}z.$$

Le dénominateur de cette fraction rationnelle irréductible ne s'annule pas sur le disque unité, et donc la solution est causale! Ainsi il faut toujours simplifier les racines communes à Φ et Θ car la solution ne dépend que de la fraction rationnelle irréductible.

3. La fraction rationnelle irréductible indique que la solution est donnée pour tout $t \in Z$ par $X_t = Z_t - (1/3)Z_{t-1}$. Ce processus est manifestement un filtre causal de Z .