

Analyse statistique des séries chronologiques

M1. MAS, M. BOUTAHAR, M. CARENZI

Chapitre 1. Décomposition d'une série chronologique

1.1. Introduction

Une série chronologique Z_t peut être modélisée en considérant les trois composantes suivantes:

- La tendance, notée m_t , elle traduit le fait que la valeur de la série croît ou décroît dans le temps t
- La saisonnalité notée S_t , elle permet de décrire le phénomène cyclique de la série
- Terme d'erreur, notée B_t , il modélise la partie aléatoire de la série.

Selon la nature de la composante cyclique, on peut utiliser l'un des deux modèles suivants:

- Modèle additif: $Z_t = m_t + \mathbf{S}_t + B_t$
- Modèle multiplicatif: $Z_t = m_t * \mathbf{S}_t * B_t$

En général si l'amplitude du cycle varie dans le temps on utilise le modèle multiplicatif, phénomène que l'on peut observer sur le chronogramme de la série, cependant il existe plusieurs tests statistiques qui permettent de décider si le modèle est additif ou multiplicatif.

1.2. Test de Buys-Ballot

Nous allons présenter le test de Buys-Ballot qui est simple à programmer:

Soient n la taille de la série Z_t , et r la période, posons $n_s = n/r$; (on suppose que n est un multiple de r). On définit alors la matrice

$$Y[i,j] = Z[i + (j - 1) * r], 1 \leq i \leq r, 1 \leq j \leq n_s.$$

et les deux vecteurs des moyennes et des écart-types de l'année j suivants

$$\text{Moyenne}[j] = \frac{1}{r} \sum_{i=1}^r Y[i, j],$$

$$\text{Sigma}[j] = \sqrt{\frac{1}{r} \sum_{i=1}^r (Y[i, j] - \text{Moyenne}[j])^2}, 1 \leq j \leq n_s.$$

On effectue la régression des écart-types sur les moyennes

$$\text{Sigma}[j] = a_1 + a_2 \text{Moyenne}[j] + \epsilon_j.$$

Si le coefficient a_2 est significatif (p-value inférieur à 0.05) alors on décide que le modèle est multiplicatif sinon il est additif.

Exemple 1. On considère la série des ventes suivante:

Trimestre / Année	2015	2016	2017
T_1	1248	891	1138
T_2	1392	1065	1456
T_3	1057	1118	1224
T_4	3159	2934	3090

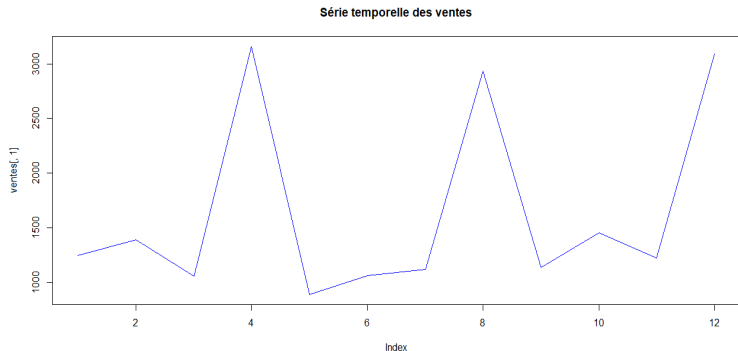


Figure 1: Chronogramme des ventes

L'application du test de Buys-Ballot nous donne une $p\text{-value}=0.7864$, par conséquent le modèle est additif.

1.3. Identification des trois composantes

Une fois le test de Buys-Ballot effectué, si le modèle est additif on prends $X_t = Z_t$ alors que s'il est multiplicatif on prendra $X_t = \log(Z_t)$, avec cette transformation logarithmique le modèle multiplicatif deviendra un modèle additif. Dans la suite nous allons donner la démarche de modélisation par un modèle additif.

1.3.1. Identification de la tendance

On se limite dans ce cours à identifier une tendance linéaire de la forme $m_t = b_1 + b_2 t$, pour ceci on effectue la régression suivante:

$$X_t = b_1 + b_2 t + B_t, 1 \leq t \leq n,$$

Si la p-value associée à b_2 est inférieure à 0.05 alors la tendance est significative, c'est à dire que la chronique croît (si $b_2 > 0$) et décroît (si $b_2 < 0$) avec le temps.

- Si la tendance est significative alors on prendra $m_t = \hat{b}_1 + \hat{b}_2 t$, où \hat{b}_1 et \hat{b}_2 sont les estimateurs de b_1 et b_2 respectivement.
- Si la tendance n'est pas significative alors on prendra $m_t = \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$.

Exemple 1 (suite). Si on considère la série des ventes, l'application de la régression nous donne une p-value = 0.4213 pour b_2 et donc la tendance n'est pas significative par conséquent $m_t = \bar{X} = 1647.6667$

1.3.2. Identification de la composante saisonnière

Dans ce cours nous allons présenter la méthode des moyennes saisonnières

Pour identifier la composante saisonnière on considère la série sans tendance $W_t = X_t - m_t$.

La composante saisonnière est supposée de la forme:

$$S_t = S_i \text{ si } (t \bmod r) = i, 1 \leq i \leq r, 1 \leq t \leq n \quad (1)$$

$(t \bmod r)$ est le reste de la division de t par r .

Pour que le modèle soit identifiable on impose la condition suivante sur les coefficients saisonniers S_1, \dots, S_r

$$S_1 + S_2 + \dots + S_r = 0. \quad (2)$$

On définit la matrice

$$Y[i, j] = W[i + (j - 1) * r], 1 \leq i \leq r, 1 \leq j \leq n_s.$$

le vecteur des coefficients saisonniers est donné par

$$S'[i] = \frac{1}{n_s} \sum_{j=1}^{n_s} Y[i, j];$$

Par la suite on normalise les coefficients saisonniers pour satisfaire la condition (2):

$$S = S' - \bar{S}', \text{ avec } \bar{S}' = \frac{1}{r} \sum_{i=1}^r S'[i].$$

Exemple 1 (suite). On considère la série des ventes, on obtient alors les coefficients saisonniers suivants:

S_1	-555.3333
S_2	-343.3333
S_3	-514.6667
S_4	1413.3333

Interprétation :

-La moyenne des ventes sur la période analysée 2015-2017 est de $m_t = 1647.6667$, à chaque trimestre les ventes vont augmenter ou diminuer de la valeur de S_i selon le signe de celui-ci, par exemple pour le trimestre 4, les ventes vont augmenter de 1413.3333 par rapport à la moyenne alors que pour le trimestre 1, les ventes vont chuter de -555.3333.

-La chronique désaisonnalisée est définie par:

$$X_t^s = X_t - \mathbf{S}_t, \text{ avec } \mathbf{S}_t = S_i \text{ si } (t \bmod r) = i, 1 \leq i \leq r.$$

-La chronique corrigée de la tendance est donnée par:

$$X_t^d = X_t - m_t.$$

-La série résiduelle (partie aléatoire ou stochastique) est donnée par :

$$R_t = X_t - m_t - S_t.$$

La figure 2 montre les trois composantes m_t , S_t , R_t ainsi que la série brute X_t de la série des ventes.

Décomposition de la chronique Ventes

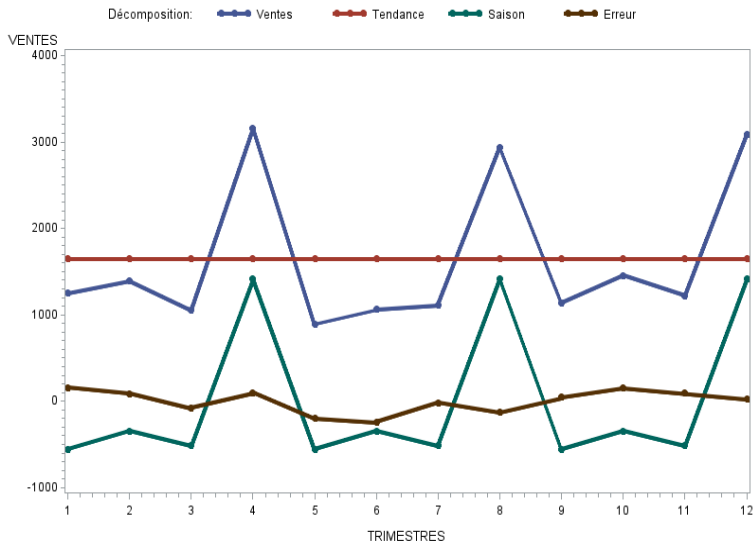


Figure 2: Les trois composantes de la chronique des ventes,

Chapitre 2. Modèles ARMA

2.1. Processus stationnaires au second ordre

Definition

On appelle **processus aléatoire du second ordre** à temps discret, toute suite réelle $Z_t : \Omega \rightarrow \mathbb{R}$ définie sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, indexée par $t \in \mathbb{N}$ ou \mathbb{Z} telle que:

$$E \left(|Z_t|^2 \right) < \infty \text{ pour tout } t.$$

Pour de tel processus, on peut définir les deux fonctions suivantes:

i) Moyenne: $\mu : \mathbb{Z} \longrightarrow \mathbb{R}$, $\mu(t) = E(Z_t)$, $\forall t \in \mathbb{Z}$.

ii) Auto-covariance: $K : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{R}$,

$K(t, s) = E((Z_t - \mu(t))(Z_s - \mu(s)))$, $\forall t, s \in \mathbb{Z}$.

Processus strictement stationnaire

Definition

Soit Z un processus; on dit que Z est strictement (ou fortement) stationnaire si pour toute suite finie d'instantes t_1, \dots, t_k élément de \mathbb{Z} et tout entier $r \in \mathbb{Z}$, les lois jointes de $(Z_{t_1}, \dots, Z_{t_k})$ et de $(Z_{r+t_1}, \dots, Z_{r+t_k})$ sont les mêmes (lois jointes invariantes par translation dans le temps).

Processus faiblement stationnaire

Definition

Un processus est stationnaire au second ordre (on dit aussi faiblement stationnaire) si moyenne et auto-covariance sont invariantes par translation dans le temps, i.e.:

$$\begin{cases} E(Z_t) = \mu, \forall t \in \mathbb{Z}, \\ \text{et} \\ E(Z_t - \mu)(Z_s - \mu) = E(Z_{t+r} - \mu)(Z_{s+r} - \mu), \quad \forall (t, s, r) \in \mathbb{Z}^3. \end{cases}$$

où μ est une constante et $K(t, s)$ ne dépend plus que de $t - s$, donc il existe une fonction $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$, telle que:
 $K(t, s) = \gamma(t - s)$.

Alors pour un processus faiblement stationnaire, on a

$$\left\{ \begin{array}{l} \text{var}(Z_t) = \gamma(0), \forall t \in \mathbb{Z}, \\ |\gamma(k)| \leq \gamma(0), \forall k \in \mathbb{Z}, \text{ (par l'inégalité de Cauchy-Schwarz)} \\ \gamma(k) = \gamma(-k), \forall k \in \mathbb{Z}, \\ \rho(j) = \frac{\gamma(j)}{\gamma(0)}, \forall j \in \mathbb{Z}, \text{ fonction d'auto-corrélation ou mémoire} \\ r(j) = \frac{\text{cov}(Z_{1-P_{[Z_2, \dots, Z_j]}]}(Z_1), Z_{j+1-P_{[Z_2, \dots, Z_j]}]}(Z_{j+1}))}{\text{var}(Z_{1-P_{[Z_2, \dots, Z_j]}]}(Z_1))} \\ \text{fonction d'auto-corrélation partielle.} \end{array} \right.$$

Definition

On dit que ε_t est un bruit blanc fort si $(\varepsilon_t)_{1 \leq t \leq n}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) et centrées.

Definition

On dit que ε_t est un bruit blanc faible si $(\varepsilon_t)_{1 \leq t \leq n}$ est une suite de variables aléatoires non corrélées et centrées.

Si ε_t est un bruit blanc faible alors $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$ est tel que $E(\varepsilon_t \varepsilon_s) = \sigma^2 \delta_t^s = \sigma^2$ si $t=s$, et 0 sinon.

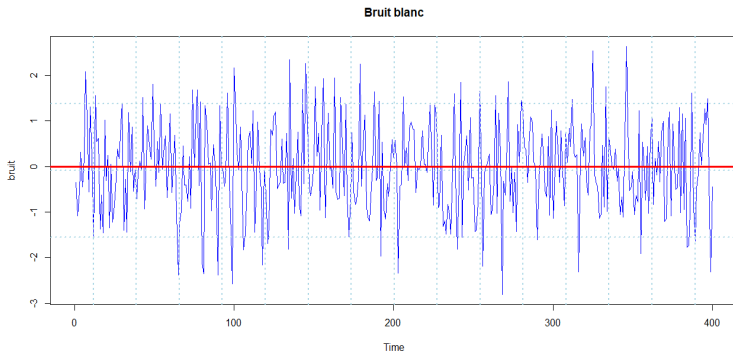


Figure 3: Trajectoire d'un bruit blanc

La figure 29 montre que la série oscille d'une manière stable

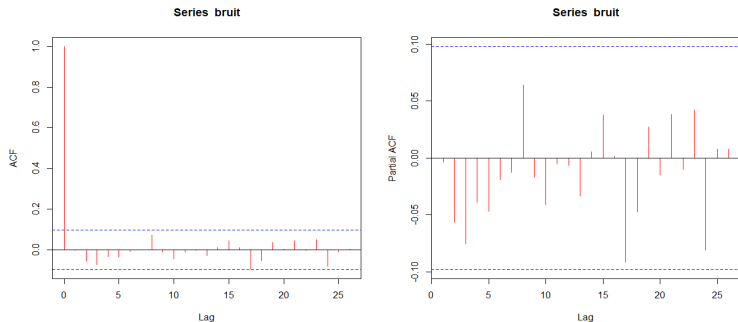


Figure 4: Les corrélations (ACF) et les corrélations partielles (PACF) du bruit blanc.

Les corrélations (ACF) et les corrélations partielles (PACF), représentées dans la figure 30 ne sont pas significatives.

2.2.Processus linéaires

Un bruit blanc est généralement noté $(\varepsilon_t)_t$.

Definition

- Un **processus linéaire** $(Z_t)_t$ est un processus qui peut s'écrire comme une combinaison linéaire d'un bruit blanc :

$$Z_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k}, \text{ avec } \psi_0 = 1 \text{ et } \psi_k \in \mathbb{R}. (3)$$

- Un **processus linéaire causal** $(Z_t)_t$ est un processus qui peut s'écrire comme une combinaison linéaire du présent et du passé d'un bruit blanc :

$$Z_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \text{ avec } \psi_0 = 1 \text{ et } \psi_k \in \mathbb{R}. (4)$$

2.3. Test KPSS de stationnarité

Soit $(\varepsilon_t)_t$ un bruit blanc fort de variance σ_ε^2 , et soit $(X_t)_t$ la marche aléatoire associée :

$$\begin{cases} X_0 = 0 \\ X_t = X_{t-1} + \varepsilon_t. \end{cases}$$

Soit $(B_t)_t$ un processus stationnaire centré indépendant de $(\varepsilon_t)_t$. On notera v' la transposée du vecteur v . Considérons le modèle

$$Z_t = \beta' D_t + X_t + B_t,$$

où D_t désigne

- soit la fonction constante égale à 1, et dans ce cas $\beta' = E(Z_t)$,
- soit le vecteur colonne $(1, t)'$ et dans ce cas $\beta' D_t$ représente une tendance linéaire.

Le test de stationnarité, proposé par Kwiatkowski-Phillips-Schmidt-Shin (KPSS) en 1992 permet de tester:

$$(\underline{H_0}) : \sigma_\varepsilon^2 = 0 \text{ (et donc } X_t = 0) \text{ contre } (\underline{H_1}) : \sigma_\varepsilon^2 \neq 0.$$

- i) Sous (H_0) , le processus $(Z_t)_t$ est stationnaire autour de $\beta' D_t$ car $Z_t - \beta' D_t = B_t$.
- ii) Sous (H_1) , le processus $(Z_t)_t$ n'est pas stationnaire autour de $\beta' D_t$
car $\text{var}(Z_t - \beta' D_t) = t \sigma_\varepsilon^2 + \sigma_B^2$.

2.4.Processus ARMA

Une classe importante de processus linéaires considérée largement dans la modélisation des séries chronologiques est celle des processus ARMA (AutoRegressive Moving Average).

Definition

On dit que $(Z_t)_t$ est un processus **autorégressif moyenne mobile d'ordres p et q** , noté **ARMA(p,q)** s'il existe un bruit blanc faible $(\varepsilon_t)_t$ de variance σ^2 et des nombres réels $a_1, \dots, a_p, b_1, \dots, b_q$ tels que

$$Z_t - \mu = \sum_{k=1}^p a_k (Z_{t-k} - \mu) + \varepsilon_t + \sum_{k=1}^q b_k \varepsilon_{t-k}. \quad (5)$$

Ainsi, pour un processus ARMA, la valeur actuelle du processus peut s'écrire comme la somme de deux composantes; la première est une combinaison linéaire du passé du processus lui-même, alors que la deuxième composante est une combinaison linéaire d'un bruit blanc. En particulier:

Definition

- $(Z_t)_t$ est dit **autorégressif d'ordre p** , noté **AR(p)**, si

$$\begin{aligned} Z_t - \mu &= a_1 (Z_{t-1} - \mu) + \cdots + a_p (Z_{t-p} - \mu) + \varepsilon_t \\ &= \sum_{k=1}^p a_k (Z_{t-k} - \mu) + \varepsilon_t. \end{aligned} \quad (6)$$

- $(Z_t)_t$ est dit **moyenne mobile d'ordre q** , noté **MA(q)**, si

$$\begin{aligned} Z_t &= \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q} \\ &= \mu + \sum_{k=0}^q b_k \varepsilon_{t-k}, \text{ où } b_0 = 1. \end{aligned}$$

Stationnarité d'un processus ARMA

Ecriture simplifiée

Definition

On appelle **opérateur de retard** l'application L qui fait associer à la valeur présente Z_t la valeur retardée Z_{t-1} , c'est à dire que

- $L(Z_t) = Z_{t-1}$,
- $L^2(Z_t) = L(L(Z_t)) = L(Z_{t-1}) = Z_{t-2}$.

L'équation (5) peut alors s'écrire sous une forme plus simple.

Definition

Soit $(Z_t)_t$ un processus ARMA(p,q) défini par l'équation (5).
Alors

$$A(L)(Z_t - \mu) = B(L)\varepsilon_t, \quad (7)$$

où $A(z) = 1 - a_1z - \dots - a_pz^p$ et $B(z) = 1 + b_1z + \dots + b_qz^q$
sont deux polynômes. On suppose généralement que A et B
n'ont pas de racines communes.

Propriété de stationnarité

Theorem

Soit $(Z_t)_t$ un processus ARMA(p,q) défini par l'équation (7). $(Z_t)_t$ est stationnaire si et seulement si le polynôme A n'admet pas de racine unitaire (de module égal à 1).

En particulier, un processus MA(q) est toujours stationnaire.

Proposition

Soit $(Z_t)_t$ un processus ARMA(p,q) stationnaire centré associé au bruit blanc $(\varepsilon_t)_t$ de variance σ_ε^2 . Alors $(Z_t)_t$ peut s'écrire comme un processus linéaire associé à $(\varepsilon_t)_t$;

$$Z_t = \sum_{k \in \mathbb{Z}} \beta_k \varepsilon_{t-k} \quad \text{avec } \beta_k \text{ tels que}$$

$$C(z) = \frac{B(z)}{A(z)} = \sum_{k \in \mathbb{Z}} \beta_k z^k .$$

Mémoire d'un processus ARMA

Processus MA

Soit $(Z_t)_t$ un processus MA(1) centré associé au bruit blanc $(\varepsilon_t)_t$ de variance σ^2 :

$$Z_t = \varepsilon_t + b_1 \varepsilon_{t-1}.$$

Alors

$$\begin{cases} \gamma(0) = \text{var}(Z_t) = \sigma^2(1 + b_1^2), \\ \gamma(-1) = \gamma(1) = b_1 \sigma^2, \\ \gamma(h) = 0 \text{ pour tout } |h| \geq 2. \end{cases}$$

Donc la fonction d'autocorrélation est donnée par :

$$\begin{cases} \rho(0) = 1, \\ \rho(-1) = \rho(1) = \frac{b_1}{1+b_1^2}, \\ \rho(h) = 0 \text{ pour tout } |h| \geq 2. \end{cases}$$

Ainsi la mémoire d'un processus MA(1) est finie et s'annule à partir de l'ordre $h = 2$.

Soit $(Z_t)_t$ un MA(q) centré. Alors,

$$\begin{aligned}\gamma(h) &= E(Z_{t+h} Z_t) = 0 \text{ si } |h| > q, \\ &= \sum_{l=0}^q \sum_{m=0}^q b_l b_m E(\varepsilon_{t+h-l} \varepsilon_{t-m}), \text{ si } |h| \leq q, \\ &= \sigma^2 \sum_{j=0}^{q-h} b_{h+j} b_j, \text{ si } |h| \leq q.\end{aligned}$$

Theorem

Soit $(Z_t)_t$ un $MA(q)$. Alors $\gamma(q) \neq 0$ et $\gamma(h) = 0, \forall |h| > q$.

Soit $(Z_t)_t$ un $MA(\infty)$ centré, c'est dire que

$$Z_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}.$$

Alors

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} b_j b_{j+|h|}, \quad (8)$$

$$\text{et } \rho(h) = \sum_{j=0}^{\infty} \frac{b_j b_{j+|h|}}{\sum_{k=0}^{\infty} b_k^2}. \quad (9)$$

Remarque : La fonction d'autocorrélation (ou la mémoire) d'un processus MA(∞) ne dépend pas de la variance σ^2 du bruit $(\varepsilon_t)_t$.

Processus AR

Soit $(Z_t)_t$ un AR(1) centré :

$$Z_t = a_1 Z_{t-1} + \varepsilon_t, \text{ avec } |a_1| < 1 \text{ et } a_1 \neq 0.$$

Première méthode pour calculer la mémoire du processus AR(1)

On a

$$\begin{aligned} Z_t &= a_1 Z_{t-1} + \varepsilon_t \\ &= a_1 (a_1 Z_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &\vdots \\ &= \sum_{j=0}^{\infty} a_1^j \varepsilon_{t-j}. \end{aligned} \tag{10}$$

La représentation (10) peut aussi se déduire à l'aide de l'opérateur retard. En effet, en posant $A(L) = I - a_1 L$, on obtient

$$Z_t = a_1 Z_{t-1} + \varepsilon_t \iff A(L)Z_t = \varepsilon_t.$$

Or, si $|a_1| < 1$, on a

$$(1 - a_1 L) (1 + a_1 L + a_1^2 L^2 + \dots) = 1,$$

donc l'opérateur $(1 - a_1 L)$ est inversible d'inverse

$$(1 - a_1 L)^{-1} = 1 + \sum_{j=1}^{\infty} a_1^j L^j.$$

Par conséquent,

$$A(L)Z_t = \varepsilon_t \iff Z_t = (1 - a_1 L)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} a_1^j \varepsilon_{t-j}.$$

Donc un $AR(1)$ est un processus $MA(\infty)$ et sa fonction d'autocorrélation peut s'obtenir de l'équation (9) avec $b_j = a_1^j$.

Seconde méthode pour calculer la mémoire du processus AR(1)

De l'égalité (10) on déduit que pour tout $h > 0$, $E(\varepsilon_{t+h} Z_t) = 0$. Donc

$$\begin{aligned}\gamma(h) &= E(Z_{t+h} Z_t) \\ &= E((a_1 Z_{t+h-1} + \varepsilon_{t+h}) Z_t) \\ &= a_1 E(Z_{t+h-1} Z_t) \\ &= a_1 \gamma(h-1) \\ &\vdots \\ &= a_1^h \gamma(0).\end{aligned}\tag{11}$$

Donc la fonction d'autocorrélation est donnée par :

$$\begin{aligned}\rho(h) &= a_1^h \\ &= \{\text{signe}(a_1)\}^h e^{h \ln(|a_1|)},\end{aligned}\quad (12)$$

où

$$\text{signe}(a_1) = \begin{cases} 1 & \text{si } a_1 > 0 \\ -1 & \text{si } a_1 < 0. \end{cases}$$

Ainsi la mémoire d'un processus AR(1) est infinie mais tend vers 0 avec une vitesse exponentielle.

Soit $(Z_t)_t$ un AR(p) centré.

Première méthode pour calculer la mémoire du processus AR(p)

On pose $A(z) = 1 - a_1z - \dots - a_pz^p$, alors

$$A(L)Z_t = \varepsilon_t.$$

On peut calculer sa mémoire en écrivant $A^{-1}(z)$ sous la forme

$A^{-1}(z) = \sum_{k=0}^{\infty} \psi_k z^k$. On a alors

$$Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (13)$$

et on applique l'Equation (8).

Seconde méthode pour calculer la mémoire du processus AR(p)

La deuxième méthode consiste à calculer la fonction d'autocovariance récursivement. En effet, de l'équation (13) on déduit que pour tout $k \geq p$, on a $E(\varepsilon_{t+k} Z_t) = 0$ et donc

$$\begin{aligned}\gamma(k) &= E(Z_{t+k} Z_t) \\ &= E((a_1 Z_{t+k-1} + \dots + a_p Z_{t+k-p} + \varepsilon_{t+k})Z_t) \\ &= a_1 \gamma(k-1) + \dots + a_p \gamma(k-p).\end{aligned}\quad (14)$$

En comparant les Equations (6) et (14), on constate que la fonction d'autocovariance vérifie la même équation de récursion que le modèle AR(p) dans laquelle on supprime le bruit.

Processus ARMA

Si $(Z_t)_t$ est un ARMA(p, q), on peut montrer que la fonction d'autocovariance vérifie

$$\gamma(k) = a_1\gamma(k-1) + \dots + a_p\gamma(k-p) + \sigma^2 \sum_{k \leq j \leq q} b_j a_{j-k}$$

si $0 \leq k < \max(p, q+1)$ et

$$\gamma(k) = a_1\gamma(k-1) + \dots + a_p\gamma(k-p)$$

si $k \geq \max(p, q+1)$.

Représentation canonique d'un processus ARMA

Definition

Soit $(Z_t)_t$ le processus ARMA défini par l'équation (7). On dit que $(Z_t)_t$ est

- **causal** si les racines de $A(z)$ sont toutes strictement à l'extérieur du disque unité.
- **inversible** si les racines de $B(z)$ sont toutes strictement à l'extérieur du disque unité.

Theorem

Soit $(Z_t)_t$ le processus ARMA défini par l'Equation (7), tel que A et B n'ont pas de racines de module égal à 1. Alors il existe deux polynômes $\tilde{A}(z)$ et $\tilde{B}(z)$ n'ayant que des racines strictement à l'extérieur du disque unité et un bruit blanc faible $(\eta_t)_t$ tels que:

$$\tilde{A}(L)(Z_t - \mu) = \tilde{B}(L)\eta_t.$$

Cette représentation est unique et dite canonique.

On note $|z|$ le module du complexe z , et \bar{z} son conjugué. Alors on peut réécrire les polynômes A et B de la façon suivante

$$A(z) = \prod_{j=1}^p (1 - \alpha_j z) \quad \text{et} \quad B(z) = \prod_{j=1}^q (1 - \beta_j z) \quad \text{avec}$$

$$|\alpha_j| < 1, \quad \text{si } 1 \leq j \leq r; \quad |\beta_j| < 1, \quad \text{si } 1 \leq j \leq s;$$

$$|\alpha_j| > 1, \quad \text{si } r + 1 \leq j \leq p; \quad |\beta_j| > 1, \quad \text{si } s + 1 \leq j \leq q.$$

Il suffit ensuite de poser

$$\tilde{A}(z) = \prod_{j=1}^r (1 - \alpha_j z) \prod_{j=r+1}^p \left(1 - \frac{1}{\bar{\alpha}_j} z\right);$$

$$\tilde{B}(z) = \prod_{j=1}^s (1 - \beta_j z) \prod_{j=s+1}^q \left(1 - \frac{1}{\bar{\beta}_j} z\right);$$

$$\text{et } \sigma_{\eta}^2 = \sigma_{\varepsilon}^2 \frac{\prod_{j=s+1}^q |\beta_j|^2}{\prod_{j=r+1}^p |\alpha_j|^2}.$$

2.5. Modélisation d'une série chronologique par un processus ARMA

Les processus ARMA sont destinés à la modélisation des séries chronologiques stationnaires, c'est à dire des séries dont la trajectoire fluctue autour d'une constante avec une variance stable dans le temps.

La modélisation ARMA se fait en cinq étapes :

- 1 Identification des degrés (p, q)
- 2 Estimation des paramètres a_i, b_j .
- 3 Réduction du modèle (significativité des paramètres)
- 4 Validation du modèle (tests de diagnostic des résidus)
- 5 Sélection du modèle (choix du modèle optimal parmi plusieurs proposés)

Identification du processus ARMA

Plusieurs méthodes ont été élaborées pour déterminer les ordres p et q du processus ARMA(p, q) capable de bien décrire la dynamique de la série chronologique.

Identification d'un MA(q) Soit $(Z_t)_t$ un MA(q) centré.
Alors la corrélation vérifie

$$\rho(h) = 0 \text{ si } |h| > q,$$

Identification d'un AR(p) Soit $(Z_t)_t$ un AR(p) centré.
Alors la corrélation partielle vérifie

$$r(h) = 0 \text{ si } |h| > p,$$

Identification d'un ARMA(p,q)

Méthode 1. Corrélation étendue

Proposition

Soit $(Z_t)_t$ un processus ARMA(p,q). Alors, asymptotiquement,

$$\begin{cases} r_j(k) = 0 & \text{si } j - q > k - p \geq 0 \\ r_j(k) = c(k - p, j - q) & \text{si } 0 \leq j - q \leq k - p, \end{cases}$$

où $c(k - p, j - q)$ est soit une constante non nulle, soit une variable aléatoire continue à support dans $] - 1, 1[$.

Table 1: Fonction d'autocorrélation étendue d'un ARMA(1,1)

		MA					
AR		0	1	2	3	4	5
0		x	x	x	x	x	x
1		x	0	0	0	0	0
2		x	x	0	0	0	0
3		x	x	x	0	0	0
4		x	x	x	x	0	0
5		x	x	x	x	x	0

Méthode 2. Critères d'information

Definition

Soit L la vraisemblance du modèle. On définit

Critère d'Akaike (AIC)

$$AIC(p, q) = -2 \ln(L) + 2(p + q); \quad (15)$$

Critère Bayésien (BIC) ou (SBC)

$$BIC(p, q) = -2 \ln(L) + 2(p + q) \ln(n); \quad (16)$$

On choisit alors le couple (p^*, q^*) qui minimise l'un des deux critères.

Estimation d'un ARMA(p,q)

Soit $(Z_t)_t$ un processus ARMA(p,q) centré

$$Z_t = \sum_{k=1}^p a_k Z_{t-k} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \text{ avec } (\varepsilon_t) \sim BB(0, \sigma^2). \quad (17)$$

où $\theta = (a_1, \dots, a_p, b_1, \dots, b_q)'$ est le vecteur des paramètres inconnus.

Les paramètres θ et σ^2 peuvent être estimés soit par la méthode des moindres carrés soit par la méthode du maximum de vraisemblance.

Réduction du modèle

Notons $\beta = (\beta_1, \dots, \beta_{p+q})$, avec $\beta_i = a_i$ pour $1 \leq i \leq p$ et $\beta_i = b_{i-p}$ pour $p+1 \leq i \leq p+q$. Pour tester

$$(\underline{H}_0) : \beta_i = 0 \quad \text{contre} \quad (\underline{H}_1) : \beta_i \neq 0 \quad ,$$

Soit $\hat{\sigma}_i^2$ un estimateur de $\text{var}(\hat{\beta}_i)$ alors on a sous (\underline{H}_0)

$$\frac{\hat{\beta}_i}{\hat{\sigma}_i} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Pour un risque d'erreur α , si $\beta_i = 0$ et si n est assez grand, $\frac{\hat{\beta}_i}{\hat{\sigma}_i}$ doit appartenir à l'intervalle $[-u_{1-\alpha/2}, u_{1-\alpha/2}]$, où $u_{1-\alpha/2}$ désigne le quantile d'ordre $1 - \alpha/2$, c'est à dire que $P(Z \leq u_{1-\alpha/2}) = 1 - \alpha/2$, et Z suit une loi gaussienne standard.

Tests de validation et de sélection des modèles

Tests de diagnostic des résidus.

Le modèle ajusté à la série chronologique sera jugé valide (on dit aussi adéquat) si la suite des résidus $(\hat{\varepsilon}_t)_t$ s'approche d'un bruit blanc.

Donc l'analyse des résidus permettra de valider le modèle ou non.

Plusieurs tests d'adéquation existent.

Il est indispensable que les résidus franchissent aussi le test d'autocorrélation.

Si les résidus sont corrélés la modélisation proposée ne sera pas validée.

Test du bruit blanc.

On désire effectuer le test

$$\left\{ \begin{array}{l} (\underline{H_0}) : \varepsilon_t \text{ est un bruit blanc} \\ (\underline{H_1}) : \varepsilon_t \text{ n'est pas un bruit blanc} \end{array} \right.$$

Deux statistiques sont souvent considérées.

- **Statistique de Box-Pierce (1970):**

$$Q_{BP}(m) = n \sum_{j=1}^m \hat{\rho}_{\varepsilon}^2(j),$$

- **Statistique Test de Ljung-Box (1978):**

$$Q_{LB}(m) = n(n+2) \sum_{j=1}^m \hat{\rho}_{\varepsilon}^2(j) / (n-m).$$

On rappelle que $\hat{\rho}_\varepsilon(j)$ est l'autocorrélation empirique de ε_t donnée par

$$\hat{\rho}_\varepsilon(j) = \frac{\hat{\gamma}_\varepsilon(j)}{\hat{\gamma}_\varepsilon(0)}, \hat{\gamma}_\varepsilon(j) = \frac{1}{n} \sum_{t=1}^{n-j} (\varepsilon_{t+j} - \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon}), \bar{\varepsilon} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t.$$

Sous (H_0), les deux statistiques $Q_{BP}(m)$ et $Q_{LB}(m)$ suivent asymptotiquement une loi $\chi^2(m)$. D'après Box et Pierce, il convient de choisir $m = n/3$.

Il existe d'autres tests de validation spécifiques aux modèles ARMA(p,q)

Tests d'autocorrélation du Portemanteau

On désire effectuer le test

$$\begin{cases} (\underline{H_0}) : Z_t \text{ est un ARMA}(p,q) \\ (\underline{H_1}) : Z_t \text{ n'est pas un ARMA}(p,q) \end{cases}$$

On utilise les statistiques de tests mais $Q_{BP}(m)$ et $Q_{LB}(m)$ en remplaçant u_t par Z_t . Sous $(\underline{H_0})$, les deux statistiques $Q_{BP}(m)$ et $Q_{LB}(m)$ suivent asymptotiquement une loi $\chi^2(m - p - q)$.

Une fois le modèle validé, on peut ensuite réaliser des tests de normalité.

Si la normalité des résidus est rejetée, cela ne remet pas en cause le modèle estimé.

En effet, grâce à la loi forte des grands nombres et le théorème central limite, la consistance ainsi que la normalité asymptotique des estimateurs restent valables si la taille n de l'échantillon est suffisamment grande. En outre, la normalité des résidus permettra d'ajouter facilement des intervalles de confiance autour des prévisions.

Sélection du modèle

Pour modéliser des données, plusieurs modèles peuvent être utilisés: des modèles de régression, des modèles

ARMA(p,q),....

Pour sélectionner un modèle optimal parmi plusieurs proposés on peut adopter l'une des deux stratégies:

Stratégie 1.

On sélectionne le modèle qui minimise l'un des critères d'information, par exemple le critère AIC.

Si les modèles sont emboîtés, on peut réaliser un test de contraintes du rapport de vraisemblance.

Stratégie 2.

On utilise des critères pour mesurer la qualité d'ajustement des modèles, c'est-à dire la proximité des données estimées par le modèle $\hat{z}_1, \dots, \hat{z}_n$ avec les données réelles z_1, \dots, z_n . On choisit le modèle qui minimise des critères suivants :

- Biais "Mean Error" (ME)

$$ME = \frac{1}{n} \sum_{k=1}^n z_k - \hat{z}_k .$$

- Critère quadratique "Root Mean Square Error" (RMSE)

$$RMSE = \sqrt{MSE}, \quad MSE = \frac{1}{n} \sum_{k=1}^n (z_k - \hat{z}_k)^2 .$$

- Biais absolu "Mean Absolute Error" (MAE)

$$MAE = \frac{1}{n} \sum_{k=1}^n |z_k - \hat{z}_k|.$$

- Critère du pourcentage d'erreur moyenne "Mean Percentage Error" (MPE)

$$MPE = \frac{100}{n} \sum_{k=1}^n \frac{z_k - \hat{z}_k}{z_k}.$$

- Critère du pourcentage d'erreur moyenne absolue
"Mean Absolute Percentage Error" (MAPE)

$$MAPE = \frac{100}{n} \sum_{k=1}^n \frac{|z_k - \hat{z}_k|}{|z_k|}.$$

- Critère du pourcentage d'erreur moyenne absolue
"Mean Absolute Scaled Error" (MASE)

$$MASE = \frac{1}{n} \sum_{k=1}^n \frac{|z_k - \hat{z}_k|}{S_r},$$

avec $S_r = \frac{1}{m-r} \sum_{k=r+1}^m |z_k - z_{k-r}|,$

où r =fréquence(z) dépend de la fréquence des données.

Les critères qui mesurent la qualité d'ajustement ont tendance à diminuer avec la complexité du modèle, et favorisent donc les modèles les plus complexes. Pour utiliser ces critères de façon pertinente, il convient de plutôt de les utiliser pour calculer le pouvoir prédictif des modèles sur un échantillon témoin. On procède donc de la façon suivante :

1) On divise la série en deux sous-séries: la première contient entre 80% et 95% premières observations $S_1 = (z_1, \dots, z_m)$, (série d'apprentissage); la deuxième contient entre 20% et 5% dernières observations $S_2 = (z_{m+1}, \dots, z_n)$ (série témoin).

2) Pour chaque modèle proposé, on calcule les prévisions $\hat{z}_{m+1/m}, \dots, \hat{z}_{m+N/m}$ des variables aléatoires Z_{m+1}, \dots, Z_{m+N} et on calcule les critères précédents en remplaçant $\hat{z}_1, \dots, \hat{z}_n$ par $\hat{z}_{m+1/m}, \dots, \hat{z}_{n/m}$ et z_1, \dots, z_n par z_{m+1}, \dots, z_n .

3) On sélectionne le modèle qui minimise l'un de ces critères. Il faut noter que les critères relatifs tels que MPE ou MAPE peuvent donner des valeurs infinies. Hyndman et al. conseillent d'utiliser plutôt le critère MASE. Mais il faut remarquer que les critères MASE et MAE, étant égaux à une constante près, vont donner la même classification des modèles sur une série donnée.