

HAX818X - Séries Temporelles

M1 Statistique, Science des Données

Elodie Brunel

Université de Montpellier – Faculté des Sciences

2024-2025



Modalités du Contrôle des Connaissance (MCC)

▶ **session 1 :**

50% Contrôle Continu : compte-rendu de deux TP (TP2 et TP5) à remettre sur Moodle à la fin de la séance de TP.

50% Contrôle écrit Terminal de 2H pendant la période des examens de Mai.

▶ **session 2 :**

Contrôle Ecrit de 2H pendant la période des examens de Juin.

La règle du MAX s'applique entre les sessions 1 et 2.

Objectifs du cours : introduire les méthodes statistiques classiques pour l'analyse et la prévision des **séries temporelles**.

Plan du cours :

- ▶ **Partie I. Démarche et premières définitions**
- ▶ **Partie II. Processus ARMA et leur estimation**
- ▶ **Partie III. Séries temporelles non-stationnaires**
- ▶ **Partie IV. Lissage exponentiel**

Partie 1 : Démarche et premières définitions

Définition

Une *série temporelle* ou *série chronologique* est une suite d'observations d'un phénomène physique au cours du temps.

Exemples¹ :

- ▶ consommation d'électricité,
- ▶ population d'un pays,
- ▶ trafic Internet,
- ▶ ventes de téléphones mobiles,
- ▶ hauteurs des crues du Nil,
- ▶ concentration en dioxyde de carbone de l'atmosphère,
- ▶ taux de glucose dans le sang.

1. et de nombreux exemples : <http://www.seriestemporelles.com/>

Exemple 1 : population en France²



FIGURE – population de la France en millions d'habitants (1846 à 1951)

Exemple 2 : accidents de la route

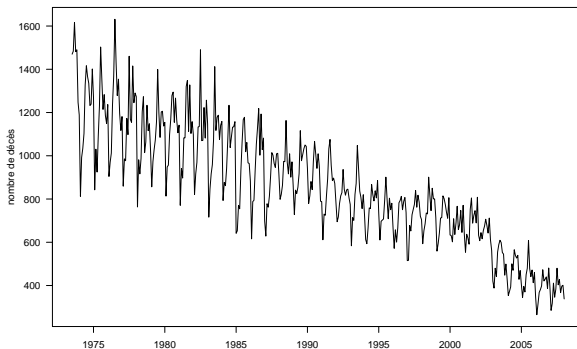


FIGURE – nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

Exemple 3 : lac Huron

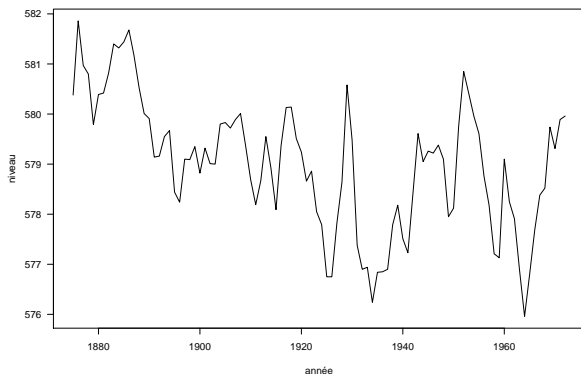


FIGURE – mesures annuelles en pieds du niveau du lac Huron de 1875 à 1972.

Exemple 4 : ventes de champagne

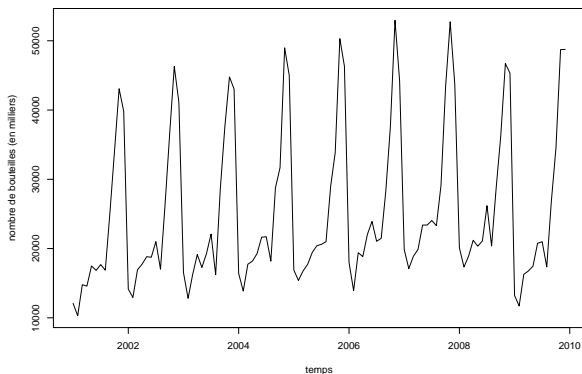


FIGURE – ventes mensuelles de champagne (2001 à 2010)

On propose de modéliser les observations comme une réalisation d'un processus stochastique $(X_t)_{t \in T}$.

Définition

Un *processus stochastique* est une famille de variables aléatoires $(X_t)_{t \in T}$ définies sur un espace de probabilisé (Ω, \mathcal{A}, P) .

Lorsque $T = \mathbb{Z}$, on dira que (X_t) est un processus à temps discret.

Lorsque $T = \mathbb{R}$, on dira que (X_t) est un processus à temps continu.

$$\begin{aligned} X : \Omega \times T &\rightarrow E = \mathbb{R} \\ (\omega, t) &\mapsto X_t(\omega) \end{aligned}$$

Remarque :

- ▶ pour chaque $t \in T$, $\omega \mapsto X_t(\omega)$ est une variable aléatoire.
- ▶ pour chaque $\omega \in \Omega$, $t \mapsto X_t(\omega)$ est une fonction que l'on appelle trajectoire du processus (X_t) .

Dans la suite on se restreindra à l'étude des processus à temps discret.

On se place dans l'ensemble \mathbb{L}^2 des variables aléatoires réelles admettant un moment d'ordre 2 $\mathbb{E}(X_t^2) < +\infty$. L'ensemble \mathbb{L}^2 est un espace vectoriel normé pour la norme définie par :

$$\|X\|^2 = \mathbb{E}(X^2) \text{ associée au produit scalaire } \langle X, Y \rangle = \mathbb{E}(XY)$$

De plus l'espace \mathbb{L}^2 est complet : c'est un espace de Hilbert.

Définition

Soit X et Y deux v.a. de \mathbb{L}^2 , X et Y sont dites orthogonales si $\mathbb{E}(XY) = 0$.

Dans toute la suite, les v.a. X_t sont dans \mathbb{L}^2 . On parle alors de processus du second ordre .

Définition

Le processus $(X_t)_{t \in \mathbb{Z}}$ est un *processus du second ordre* si $E(X_t^2) < \infty$.

On note sa moyenne $\mu(t)$, pour $t \in \mathbb{Z}$:

$$\mu(t) = \mathbb{E}(X_t),$$

et sa fonction de covariance pour $s, t \in \mathbb{Z}$:

$$\gamma_X(s, t) = \text{Cov}(X_s, X_t) = \mathbb{E}[(X_s - E(X_s))(X_t - E(X_t))].$$

Définition

Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit *stationnaire* ou *faiblement stationnaire* ou *stationnaire au second ordre* si

(i) $E[X_t^2] < \infty$

(ii) $E[X_t] = \mu, \forall t \in \mathbb{Z}$ (ne dépend pas du temps t)

(iii) $\text{Cov}(X_t, X_{t-h}) = \mathbb{E}[(X_t - E(X_t))(X_{t-h} - E(X_{t-h}))] = \gamma(h), \forall t, h \in \mathbb{Z}$
(ne dépend que de h).

En particulier, on a la propriété d'*homoscédasticité* :

$$\gamma(0) = \text{Var}(X_t) = \text{cste}, \forall t \in \mathbb{Z}.$$

La fonction $\gamma(h)$ est appelée la *fonction d'autocovariance*.

Définition

La fonction d'*autocorrélation* d'un processus stationnaire au second ordre est définie par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \forall h \in \mathbb{Z}.$$

Définition

Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit *strictement stationnaire* si les lois jointes de $(X_{t_1}, \dots, X_{t_k})$ et de $(X_{t_1+h}, \dots, X_{t_k+h})$ sont identiques pour tout entier positif k et pour tous $t_1, \dots, t_k, h \in \mathbb{Z}$.

Remarque : Un processus strictement stationnaire du second ordre est faiblement stationnaire.

Exemple important : Processus gaussien.

Définition

On dit qu'un processus à valeurs réelles $(X_t)_{t \in T}$ est *gaussien* si pour tout ensemble fini d'indices $I = \{t_1, \dots, t_n\}$, $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.

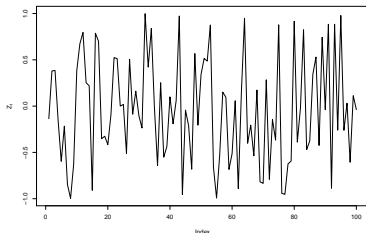


FIGURE – Simulation d'une trajectoire d'un processus stationnaire

Un processus gaussien stationnaire au second ordre est stationnaire au sens strict car la loi d'un processus gaussien est entièrement caractérisée par sa moyenne et sa covariance.

Exemple : Bruit Blanc.

Définition

Un processus stationnaire $(Z_t)_{t \in \mathbb{Z}}$ de moyenne μ et d'autocovariance γ_Z est :

- ▶ *bruit blanc (faible)* si $\text{Cov}(Z_s; Z_t) = 0$ pour tous $s \neq t$;
- ▶ *bruit blanc fort* lorsque les variables $(Z_t)_{t \in \mathbb{Z}}$ sont indépendantes ;

On a dans les deux cas , $\gamma_Z(h) = \sigma^2 \mathbb{1}_{(h=0)}$.

Quelques exemples

Exemple 1 : Soit (Z_t) une suite de variables aléatoires i.i.d. d'espérance nulle et de variance finie $\sigma^2 < +\infty$. On pose

$$X_t = Z_t + \theta Z_{t-1}, \theta \in \mathbb{R}$$

Calculer la fonction d'autocovariance de (X_t) . En déduire que (X_t) est (faiblement) stationnaire.

Quelques exemples

Exemple 2 : Soit (Y_t) un processus stationnaire. On définit

$$X_t = \begin{cases} Y_t, & \text{si } t \text{ est pair} \\ Y_t + 1, & \text{si } t \text{ est impair} \end{cases}$$

Calculer la fonction d'autocovariance de (X_t) . (X_t) est-il stationnaire ?

Quelques exemples

Exemple 3 : Soit $X_t = Z_1 + Z_2 + \dots + Z_t$ où les Z_i sont des variables aléatoires i.i.d. d'espérance nulle et de variance $\sigma^2 < +\infty$. Calculer la fonction d'autocovariance de (X_t) . (X_t) est-il stationnaire ?

Quelques exemples

Exemple 4 : Soit (X_t) une suite de variables aléatoires indépendantes telles que :

$X_t \sim \text{Exp}(1)$ lorsque t est pair

$X_t \sim \mathcal{N}(1, 1)$, lorsque t est impair.

Montrer que (X_t) est stationnaire mais pas strictement stationnaire.

Décomposition d'une série temporelle

Les séries temporelles que l'on veut modéliser ne sont pas en général des réalisations de processus stationnaires. Mais il est possible de s'y ramener par décomposition de la série temporelle en plusieurs composantes.

On peut décrire comme suit ces composantes :

- ▶ **Tendance.** La tendance ou trend m_t modélise l'évolution à long terme de la série.
- ▶ **Composante saisonnière.** La composante saisonnière s_t représente un comportement qui se répète avec une certaine périodicité (toutes les douze périodes pour des séries mensuelles, toutes les sept périodes pour des séries quotidiennes, toutes les quatre périodes pour des séries trimestrielles...).
- ▶ **Composante irrégulière.** C'est une composante aléatoire e_t c'est-à-dire un processus stochastique.

Décomposition d'une série temporelle

La tendance m_t et la saisonnalité s_t sont des composantes **déterministes**.

- ▶ **Tendance.** La tendance ou trend m_t modélise l'évolution à long terme de la série. Par exemple :
 - tendance linéaire : $m_t = a + bt$
 - tendance quadratique $m_t = a + bt + ct^2$
 - tendance logarithmique $m_t = \log(t)$.
- ▶ **Composante saisonnière.** La composante saisonnière s_t représente un comportement qui se répète avec une certaine périodicité connue. Elle peut être modélisée par une fonction périodique. Par exemple, pour une composante saisonnière de période k :

$$s_t = a_0 + \sum_{j=1}^d (a_j \cos(\omega_j t) + b_j \sin(\omega_j t))$$

où $\omega_j = 2\pi j/k$, pour $j = 1, \dots, d$.

Décomposition d'une série temporelle

Seule la composante irrégulière e_t est **aléatoire**. On peut aussi la désigner comme **l'erreur ou le bruit**.

- ▶ **Composante irrégulière**. C'est une composante aléatoire e_t c'est-à-dire un processus stochastique, idéalement une fois que la partie déterministe a été modélisée ou "éliminée", on espère que la composante e_t est un processus stationnaire.

Décomposition d'une série temporelle

Définition

Un modèle de décomposition *additif* est défini par : *multiplicative* ou *combiner les deux aspects*.

$$X_t = m_t + s_t + e_t, \text{ où } \mathbb{E}(e_t) = 0$$

Un modèle de décomposition *multiplicatif* est défini par :

$$X_t = m_t \times s_t \times e_t, \text{ où } \mathbb{E}(e_t) = 1$$

ou encore une *"combinaison"* des deux définie par :

$$X_t = (m_t + s_t) \times e_t, \text{ où } \mathbb{E}(e_t) = 1$$

Exemple de décomposition

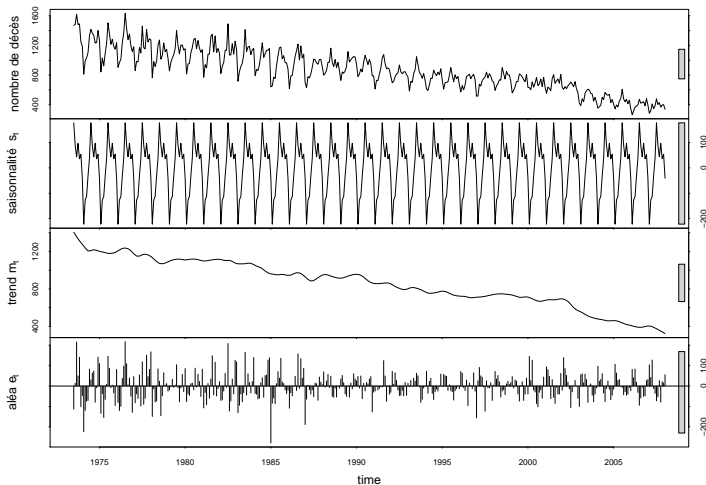


FIGURE – Décomposition du nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

Modélisation de la partie déterministe

a) Modélisation de la tendance par Moyenne mobile :

La méthode des **moyennes mobiles** permet d'extraire la tendance d'une série temporelle. C'est une méthode de lissage car elle agit comme un filtre qui élimine le "bruit".

La méthode des moyennes mobiles consiste à calculer une moyenne glissante des valeurs de la série que l'on souhaite "lisser".

La série des **moyennes mobiles** \hat{m}_t d'ordre $2q + 1$, pour $q \in \mathbb{N}^*$ se calcule ainsi :

$$\hat{m}_t = \frac{1}{2q + 1} \sum_{\ell=-q}^q x_{t+\ell}$$

Remarque : on "perd" $2q$ observations quand on fait un lissage moyenne mobile d'ordre $2q + 1$.

Remarque : Plus q est grand plus le lissage est important.

Exemple de lissage par moyenne mobile de la tendance

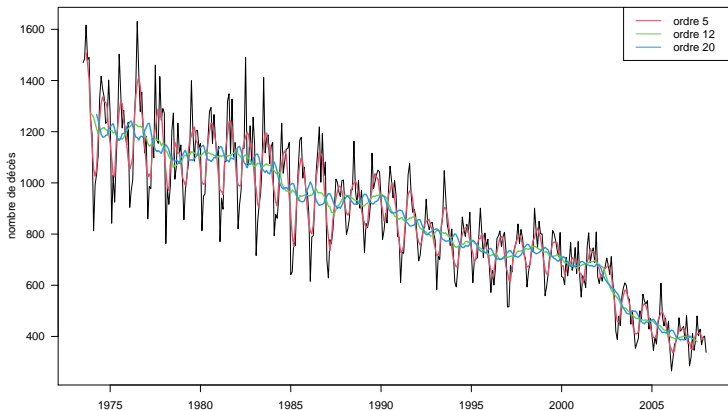


FIGURE – Lissage par moyenne mobile du nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

Exemple de lissage par moyenne mobile de la tendance

Lorsque la série présente une composante saisonnière de période K , un lissage par moyenne mobile d'ordre égal à la période K permet de "gommer" la saisonnalité.

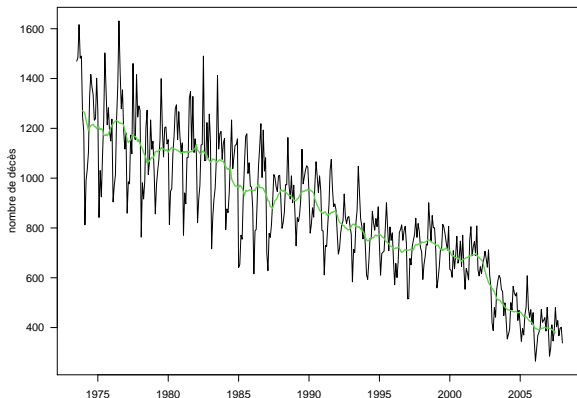


FIGURE – Lissage par moyenne mobile d'ordre 12 du nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

a) **Modélisation de la saisonnalité par moyenne mobile** : Une fois que la tendance m_t est estimée par moyenne mobile, on peut estimer la composante saisonnière s_t à partir de la série $X_t - \hat{m}_t$ (dans le modèle additif) :

$$\hat{s}_k = \frac{1}{d+1} \sum_{j=0}^d (X_{k+jK} - \hat{m}_{k+jK}), \text{ pour } k = 1, \dots, K.$$

où K est la période ($K = 12$ pour une série mensuelle, $K = 4$ pour une série trimestrielle, etc) et $d + 1$ est le nombre de périodes disponibles pour le calcul de la moyenne.

Estimation des coefficients saisonniers

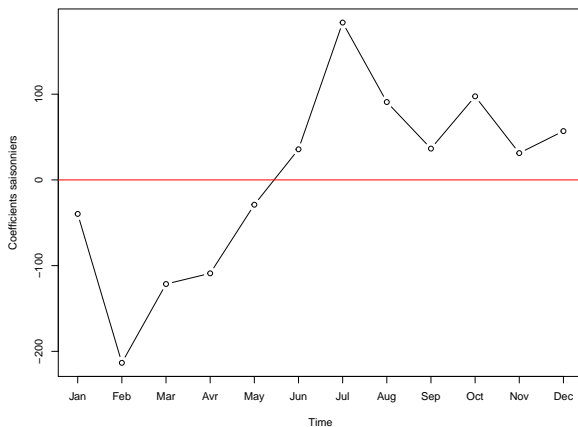


FIGURE – Coefficients \hat{s}_k , pour $k = 1, \dots, 12$ de la série du nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

Estimation de la partie déterministe de la série

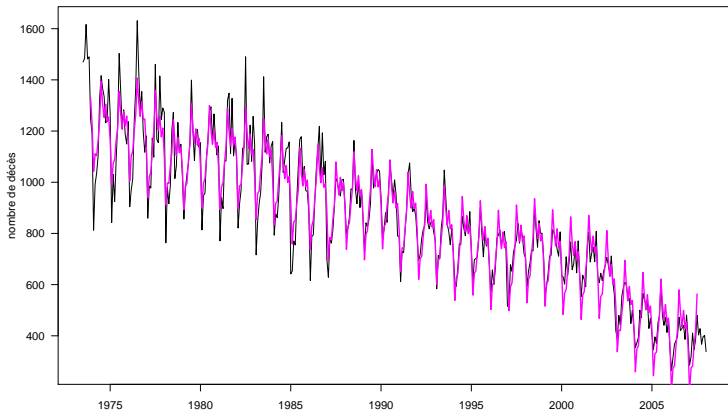


FIGURE – Reconstruction des composantes déterministes (magenta) m_t et s_t dans un modèle additif du nombre mensuel de morts et blessés graves par accident de la route en France (1974 à 2008)

Etapes en pratique de la méthode

- ▶ Représenter la série temporelle et rechercher graphiquement la présence d'une tendance et d'une composante saisonnière ;
- ▶ Modéliser la tendance et la composante saisonnière par la méthode des moyennes mobiles. Différentes fonctions implémentées dans R permettent de le faire (`decompose()`, `filter()`, `stl()`,)
- ▶ Inspecter les résidus et proposer une modélisation par un processus stationnaire (cf. partie II)

Nous allons voir dans de ce qui suit qu'il est aussi possible d'utiliser d'autres techniques pour "stationnariser" la série.

Définition

Un *opérateur* est une application définie sur l'ensemble des processus.

- ▶ L'*opérateur retard* B est l'application qui a toute v.a. X_t associe la variable retardée X_{t-1} , pour tout $t \in \mathbb{Z}$, par :

$$BX_t = X_{t-1}$$

Il décale le processus d'une unité de temps vers le passé.

- ▶ L'*opérateur différence* Δ est défini, pour tout $t \in \mathbb{Z}$, par :

$$\Delta X_t = X_t - X_{t-1} = (1 - B)X_t$$

- ▶ Pour tout entier $k \geq 1$, l'*opérateur différence saisonnier*, noté Δ_k est défini, pour tout $t \in \mathbb{Z}$, par :

$$\Delta_k X_t = X_t - X_{t-k} = (1 - B^k)X_t$$

On note abusivement "1" l'opérateur identité.

Propriété

L'opérateur retard B satisfait pour tout $i, j \in \mathbb{Z}$:

$$\text{i) } B^j X_t = \underbrace{B \circ B \cdots \circ B}_{j \text{ fois}} X_t = X_{t-j},$$

$$\text{ii) } B^0(X_t) = X_t$$

$$\text{iii) } B^j c = c = \text{cste}$$

$$\text{iv) } B^i(B^j X_t) = B^{i+j}(X_t)$$

$$\text{v) } B^{-j} X_t = X_{t+j}$$

$$\text{vi) } (B^i + B^j)X_t = B^i X_t + B^j X_t$$

vi)

$$(1 - aB)^{-1} X_t = \sum_{j=0}^{+\infty} a^j B^j X_t \text{ si } |a| < 1$$

Propriété

L'opérateur différence Δ (ou Δ_1) élimine une tendance linéaire.

Exemple : Pour un processus $X_t = a + bt + e_t$, on a :

$$\Delta X_t = b + e_t - e_{t-1}.$$

Propriété

Les puissances de l'opérateur différence Δ^d élimine une tendance polynomiale de degré d .

Exemple : Pour un processus $X_t = a + bt + ct^2 + e_t$, on a :

$$\Delta^2 X_t = \Delta^2(a + bt + ct^2 + e_t) = \dots = 2c + (e_t - 2e_{t-1} + e_{t-2})$$

Propriété

L'opérateur différence saisonnier $\Delta_k = (1 - B^k)$ élimine une composante saisonnière de période k .

Exemple : Pour un processus $X_t = m_t + s_t + e_t$, on a :

$$\Delta_k X_t = m_t - m_{t-k} + e_t - e_{t-k}$$

La série X_t (haut) est celle du nombre mensuel de morts et blessés graves par accidents, et différence saisonnière $\Delta_{12}X_t$ (milieu), puis différentiation supplémentaire de degré 1 pour éliminer la tendance (bas).

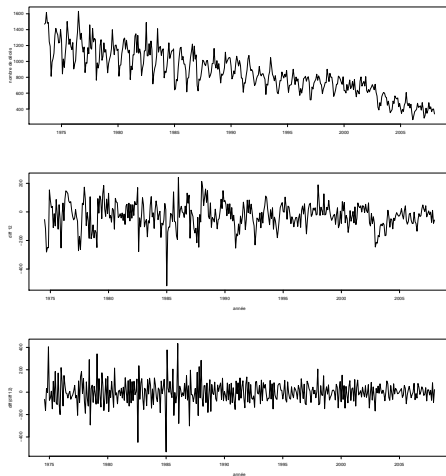


FIGURE – Elimination de la saisonnalité et de la tendance par différences

Partie 2 : Processus ARMA et leur estimation

Partie 2 : Processus ARMA et leur estimation
1 - Filtrage Linéaire

Filtrage Linéaire

Définition

On dit que le processus $(Y_t)_{t \in \mathbb{Z}}$ est obtenu par *filtrage linéaire* à partir d'un processus $(X_t)_{t \in \mathbb{Z}}$ par :

$$Y_t = \sum_{k \in \mathbb{Z}} \psi_k X_{t-k}$$

où (ψ_k) est une suite de nombre réels.

- ▶ si il y a un nombre fini de ψ_k non nuls alors la somme est bien définie
- ▶ sinon Y_t est une limite dans un sens à préciser.

C'est l'objet du théorème qui suit.

Filtrage Linéaire

Théorème

Soit $(\psi_k)_{k \in \mathbb{Z}}$ une suite absolument sommable, i.e. $\sum_{t \in \mathbb{Z}} |\psi_k| < +\infty$ et soit (X_t) un processus tel que $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] < +\infty$. Alors, pour tout $t \in \mathbb{Z}$, la suite définie par :

$$Y_{n,t} = \sum_{k=-n}^n \psi_k X_{t-k}$$

converge presque sûrement, lorsque $n \rightarrow +\infty$ vers une limite notée Y_t et $Y_t = \sum_{k \in \mathbb{Z}} \psi_k X_{t-k}$.

De plus $\mathbb{E}[|Y_t|] < +\infty$ (i.e. Y_t est intégrable) et $\lim_{n \rightarrow +\infty} \mathbb{E}[|Y_{n,t} - Y_t|] = 0$ (i.e. convergence dans $\mathbb{L}^1(\Omega, \mathcal{A}, P)$).

Enfin, si $\mathbb{E}[|X_t|^2] < +\infty$ alors $\mathbb{E}[|Y_t|^2] < +\infty$ et on a : $\lim_{n \rightarrow +\infty} \mathbb{E}[|Y_{n,t} - Y_t|^2] = 0$.

Filtrage Linéaire d'un processus stationnaire du 2nd ordre

Théorème

Soit $(\psi_k)_{k \in \mathbb{Z}}$ une suite réelle telle que $\sum_{k \in \mathbb{Z}} |\psi_k| < +\infty$ et soit (X_t) un processus *stationnaire au second ordre* de moyenne $\mu_X = \mathbb{E}[X_t]$ et de fonction d'autocovariance $\gamma_X(h) = \text{Cov}(X_{t+h}, X_t)$ alors le processus $Y_t = \sum_{k \in \mathbb{Z}} \psi_k X_{t-k}$ est stationnaire au second ordre de moyenne :

$$\mu_Y = \mu_X \sum_{k=-\infty}^{+\infty} \psi_k$$

et de fonction d'autocovariance :

$$\gamma_Y(h) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \psi_j \psi_k \gamma_X(h + k - j).$$

Ce résultat garantit qu'un processus (Y_t) obtenu par filtrage linéaire d'un processus (X_t) stationnaire au second ordre est lui même stationnaire au second ordre à condition que $\sum_{k \in \mathbb{Z}} |\psi_k| < +\infty$.

Processus linéaire

Définition

On dit que (X_t) est un *processus linéaire* s'il existe un bruit blanc (Z_t) centré, de variance σ^2 – on note $Z_t \sim BB(0, \sigma^2)$ – une suite de nombre réels $(\psi_k)_{k \in \mathbb{Z}}$ absolument sommable et un nombre réel μ tels que :

$$X_t = \mu + \sum_{k=-\infty}^{+\infty} \psi_k Z_{t-k}.$$

D'après le théorème de filtrage linéaire des processus stationnaire du 2nd ordre, on a :

Propriété

Un processus linéaire est stationnaire au second ordre de moyenne μ et de fonction d'autocovariance :

$$\gamma_X(h) = \sigma^2 \sum_{k=-\infty}^{+\infty} \psi_k \psi_{k+h}$$

Partie 2 : Processus ARMA et leur estimation

2 - Processus à moyenne ajustée $MA(q)$ et autorégressifs $AR(p)$

Processus à Moyenne Ajustée $MA(q)$

Définition

Soit q un entier non nul. On dit que (X_t) est un processus à moyenne ajustée (ou moyenne mobile) d'ordre q , noté $MA(q)$ si :

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

avec $Z_t \sim BB(0, \sigma^2)$ et $\theta_1, \dots, \theta_q$ des nombres réels et $\theta_q \neq 0$).

En utilisant la propriété des processus linéaires et en notant

$X_t = \sum_{k=0}^q \theta_k Z_{t-k}$ avec $\theta_0 = 1$, on peut en déduire que :

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \theta_k \theta_{k+h} & \text{si } 0 \leq h \leq q \\ \sigma^2 \sum_{k=0}^{q+h} \theta_k \theta_{k-h} & \text{si } -q \leq h \leq 0 \\ 0 & \text{sinon.} \end{cases}$$

à faire en exercice.

Processus Autorégressifs AR(p)

Définition

Soit p un entier non nul. On dit que (X_t) est un processus autorégressif d'ordre p , noté AR(p), si (X_t) est un processus stationnaire au second ordre solution de l'équation :

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

avec $Z_t \sim BB(0, \sigma^2)$ et ϕ_1, \dots, ϕ_p des nombres réels (et $\phi_p \neq 0$).

Remarque : La question de l'existence et l'unicité d'une solution stationnaire au second ordre de cette équation de récurrence se pose. Elle n'existera pas toujours.

Etude d'un processus AR(1)

Cas où $p = 1$

Théorème

Soient $Z_t \sim BB(0, \sigma^2)$ et ϕ_1 un nombre réel. Alors l'équation de récurrence

$$X_t = \phi_1 X_{t-1} + Z_t$$

- ▶ n'admet pas de solution stationnaire si $|\phi_1| = 1$,
- ▶ admet une solution stationnaire si $|\phi_1| \neq 1$ donnée par :

$$X_t = \sum_{k \geq 0} \phi_1^k Z_{t-k} \text{ si } |\phi_1| < 1$$

$$X_t = - \sum_{k \geq 1} \phi_1^{-k} Z_{t+k} \text{ si } |\phi_1| > 1$$

Etude d'un processus $AR(1)$

preuve : Pour démontrer le théorème de l'existence d'un $AR(1)$: On peut commencer par réécrire l'équation à l'aide d'un polynôme de l'opérateur retard B :

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + Z_t \\ \iff X_t - \phi_1 X_{t-1} &= Z_t \\ \iff \phi(B)(X_t) &= Z_t \end{aligned}$$

où $\phi(z) = 1 - \phi_1 z$ et B est l'opérateur retard tel que $B(X_t) = X_{t-1}$.

Etude d'un processus $AR(1)$

Exercice 1 : Vérifier que la fonction d'autovariance d'un processus $AR(1)$ stationnaire est donnée par :

$$\gamma_X(h) = \sigma^2 \phi_1^{|h|} / (1 - \phi_1^2) \text{ si } |\phi_1| < 1$$

et

$$\gamma_X(h) = \sigma^2 \phi_1^{-|h|} / (\phi_1^2 - 1) \text{ si } |\phi_1| > 1$$

à faire en exercice.

Exemples de trajectoires d'un processus $AR(1)$

Trajectoires de longueur 500 d'un processus $AR(1)$ gaussien.

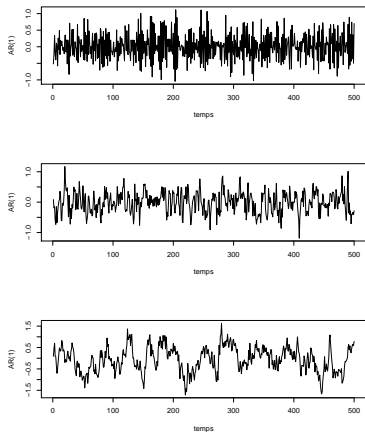


FIGURE – $\phi_1 = -0.7$ (en haut), $\phi_1 = 0.5$ (au milieu) et $\phi_1 = 0.9$ (en bas)

Partie 2 : Processus ARMA et leur estimation

3 - Propriétés des autocorrélations des processus à moyenne ajustée $MA(q)$ et autorégressifs $AR(p)$

Propriétés des Autocorrélations (simple et partielle)

On étudie les caractéristiques des fonctions d'autocorrélations simples et partielles propres aux processus AR et MA.

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

On dit que le processus $AR(p)$ est **causal** s'il existe une suite $(\psi_k)_k$ absolument sommable telle que :

$$X_t = \sum_{k \geq 0} \psi_k Z_{t-k}$$

C'est-à-dire que X_t s'exprime en fonction de Z_t et de son passé. Nous avons vu que c'est le cas d'un processus AR(1) lorsque $|\phi_1| < 1$. Quitte à faire un changement de bruit blanc, il est toujours possible de se ramener à un processus causal. Nous verrons plus loin à quelle condition sur le polynôme $\phi(B)$ dans le cas général d'un AR(p), on obtient un processus causal.

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

On suppose que le processus (X_t) est un AR(p) causal et centré :

Sa fonction d'autocorrélation est donnée par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\mathbb{E}[X_t X_{t-h}]}{\mathbb{E}[X_t^2]}$$

Le processus AR(p) est défini par :

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + Z_t, \quad t \in \mathbb{Z}$$

On calcule sa **variance** , en multipliant l'équation par X_t et en prenant l'espérance :

$$\gamma(0) = \mathbb{E}[X_t^2] = \sum_{k=1}^p \phi_k \mathbb{E}[X_t X_{t-k}] + \mathbb{E}[X_t Z_t] = \sum_{k=1}^p \phi_k \gamma(k) + \sigma^2$$

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

Puis, comme $\gamma(k) = \gamma(0)\rho(k)$

$$\gamma(0) = \gamma(0) \sum_{k=1}^p \phi_k \rho(k) + \sigma^2$$

et enfin,

$$\gamma(0) = \frac{\sigma^2}{1 - \sum_{k=1}^p \phi_k \rho(k)}.$$

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

De manière analogue, on calcule les **autocovariances** pour $h \geq 1$:

$$\mathbb{E}[X_t X_{t-h}] = \sum_{k=1}^p \phi_k \mathbb{E}[X_{t-h} X_{t-k}] + \underbrace{\mathbb{E}[X_{t-h} Z_t]}_{=0}$$

qui s'écrit (en divisant par $\gamma(0)$)

$$\rho(h) - \sum_{k=1}^p \phi_k \rho(h-k) = 0, \quad h \geq 1$$

On reconnaît une équation de récurrence linéaire homogène d'ordre p dont le polynôme caractéristique est $z^h - \sum_{k=1}^p \phi_k z^{h-k}$.

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

- pour un AR(1), l'équation devient simplement :

$$\rho(h) - \phi_1 \rho(h-1) = 0, h \geq 1$$

ce qui donne $\rho(h) = \phi_1^h$. Comme le processus est causal ($|\phi_1| < 1$), la fonction d'autocorrélation décroît exponentiellement vers 0 lorsque h tend vers l'infini.

- De façon générale, par résolution de l'équation de récurrence, on peut montrer que si les racines r_1, \dots, r_p du polynôme $\phi(x) = 1 - \sum_{k=1}^p \phi_k x^k$ sont distinctes et de module supérieur à 1 alors la solution est de la forme :

$$\rho(h) = c_1 \left(\frac{1}{r_1}\right)^h + c_2 \left(\frac{1}{r_2}\right)^h + \dots + c_p \left(\frac{1}{r_p}\right)^h$$

où c_1, c_2, \dots, c_p sont des constantes réelles.

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

Pour tout entier $h \geq 1$,

$$\rho(h) = c_1 \left(\frac{1}{r_1}\right)^h + c_2 \left(\frac{1}{r_2}\right)^h + \cdots + c_p \left(\frac{1}{r_p}\right)^h$$

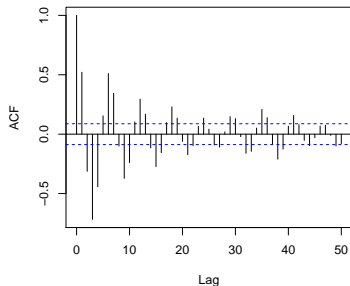
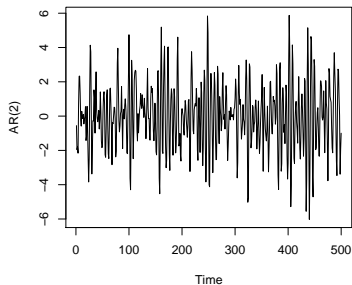
Si les racines r_1, r_2, \dots, r_p du polynôme ϕ sont toutes de module strictement supérieur à 1, la fonction d'autocorrélation décroît

- soit de manière exponentielle (si les racines sont réelles),
- soit selon des cycles amortis (si les racines sont complexes).

Propriétés des Autocorrélations d'un processus AR(p)

Autocorrélations d'un processus AR(p)

Exemple : $X_t - 0.9X_{t-1} + 0.8X_{t-2} = Z_t$ avec Z_t i.i.d $\mathcal{N}(0, \sigma^2)$.



Propriétés des Autocorrélations d'un processus MA(q)

Autocorrélations d'un processus MA(q)

Rappel : La fonction d'autocovariance d'un processus $MA(q)$ est donnée par :

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

et donc $\gamma(0) = \sigma^2(1 + \sum_{k=1}^q \theta_k^2)$, et

$$\rho(h) = \begin{cases} \frac{\sum_{k=0}^{q-|h|} \theta_k \theta_{k+|h|}}{1 + \sum_{k=1}^q \theta_k^2}, & \text{si } |h| \leq q \\ 0 & \text{si } |h| > q \end{cases}$$

La fonction d'autocorrélation d'un $MA(q)$ est nulle au-delà de l'ordre q .

Propriétés des Autocorrélations d'un processus MA(q)

Autocorrélations d'un processus MA(q)

Exemple : $X_t = Z_t - 0.8Z_{t-1}$ avec Z_t i.i.d $\mathcal{N}(0, \sigma^2)$.

Calculer les autocorrélations de X_t : à faire en exercice.

Partie 2 : Processus ARMA et leur estimation

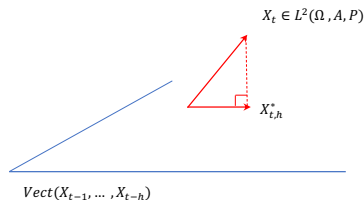
4 - Définition et Propriétés des autocorrélations partielles des processus *MA(q)* et *AR(p)*

Autocorrélations partielles

Définition

Soit un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ du second ordre. On considère la projection orthogonale de X_t sur le sous-espace vectoriel engendré par X_{t-1}, \dots, X_{t-h} notée $X_{t,h}^*$ de L^2 muni du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$. On a alors :

$$X_{t,h}^* = a_{1,h}X_{t-1} + \dots + a_{h,h}X_{t-h} \quad \text{et} \quad (X_t - X_{t,h}^*) \perp X_{t,h}^*$$



On définit l'**autocorrélation partielle** d'ordre h notée $r(h)$, comme le coefficient $a_{h,h}$ dans la régression de X_t sur X_{t-1}, \dots, X_{t-h} , pour $h \neq 0$.

Lien entre les autocorrélations et autocorrélations partielles

- En utilisant que $X_{t,1}^* = a_{1,1}X_{t-1}$ et $X_t - X_{t,1}^*$ est orthogonal à X_{t-1} ,

$$\langle X_t - X_{t,1}^*; X_{t-1} \rangle = 0 \iff \gamma(1) - a_{1,1}\gamma(0) = 0$$

et donc en divisant par $\gamma(0)$, on trouve $a_{1,1} = \rho(1) = r(1)$.

- De la même manière on calcule $r(2) = a_{2,2}$ en écrivant :

$$\langle X_t - X_{t,2}^*; X_{t-1} \rangle = 0 \iff \gamma(1) - a_{1,2}\gamma(0) - a_{2,2}\gamma(1) = 0$$

$$\langle X_t - X_{t,2}^*; X_{t-2} \rangle = 0 \iff \gamma(2) - a_{1,2}\gamma(1) - a_{2,2}\gamma(0) = 0$$

qui conduit au système :

$$\begin{cases} \rho(1) = a_{1,2} + a_{2,2}\rho(1) = 0 \\ \rho(2) = a_{1,2}\rho(1) + a_{2,2} = 0 \end{cases} \implies a_{2,2} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}.$$

Lien entre les autocorrélations et autocorrélations partielles

De façon générale, en utilisant que $X_{t,h}^* = a_{1,h}X_{t-1} + \dots + a_{h,h}X_{t-h}$ et $X_t - X_{t,h}^*$ est orthogonal à $X_{t-1}, X_{t-2}, \dots, X_{t-h}$, on obtient le système d'équations :

$$\begin{cases} \rho(1) = a_{1,h} & + a_{2,h}\rho(1) & + \dots & + a_{h,h}\rho(h-1) = 0 \\ \rho(2) = a_{1,h}\rho(1) & + a_{2,h} & + \dots & + a_{h,h}\rho(h-2) = 0 \\ \vdots & \vdots & & \\ \rho(h) = a_{1,h}\rho(h-1) & + a_{2,h}\rho(h-2) & + \dots & + a_{h,h} = 0 \end{cases}$$

$$\Leftrightarrow \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \vdots & \vdots & \dots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} a_{1,h} \\ a_{2,h} \\ \vdots \\ a_{h,h} \end{pmatrix}$$

Lien entre les autocorrélations et autocorrélations partielles

Ainsi on voit que connaissant les autocorrélations, on peut connaître les autocorrélations partielles et vice-versa.

En pratique, on obtient les autocorrélations partielles empiriques en résolvant le système d'équations (appelées équations de Yule-Walker) une fois que l'on a remplacé les autocorrélations par les autocorrélations empiriques.

Autocorrélations partielles d'un AR(p)

Soit le processus autorégressif $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$ En supposant que (X_t) est causal on a $\mathbb{E}[Z_t X_{t-h}] = 0$ pour $h \geq 1$ ou encore :

$$\mathbb{E}\left[\left(X_t - \sum_{k=1}^p \phi_k X_{t-k}\right) X_{t-h}\right] = 0, \text{ pour } h \geq 1$$

on en déduit que $(X_t - \sum_{k=1}^p \phi_k X_{t-k}) \perp X_{t-h}$, pour tout $h \geq 1$ donc en particulier

$$\left(X_t - \sum_{k=1}^p \phi_k X_{t-k}\right) \perp \text{Vect}(X_{t-1}, \dots, X_{t-p}),$$

et de plus

$$\sum_{k=1}^p \phi_k X_{t-k} \in \text{Vect}(X_{t-1}, \dots, X_{t-p}),$$

ce qui caractérise le projeté orthogonal et on en déduit que les coefficients :

$$a_{k,p} = \begin{cases} \phi_k & \text{si } 1 \leq k \leq p \\ 0 & \text{si } k > p \end{cases}$$

Autocorrélations partielles d'un $AR(p)$ et d'un $MA(q)$

- La fonction d'autocorrélation partielle d'un $AR(p)$ est nulle pour $|h| > p$ et vaut ϕ_p pour $|h| = p$.
- On peut montrer que les autocorrélations partielles d'un $MA(q)$ décroissent soit de manière exponentielle (si les racines du polynôme θ sont réelles) soit selon des cycles amortis (si les racines sont complexes).
(admis)

Utilisation des autocorrélogrammes simples et partiels

En résumé, le comportement des autocorrélogrammes empiriques simples et partiels nous guide pour identifier l'ordre d'un processus AR, MA (puis ARMA).

	$AR(p)$	$MA(q)$
autocorrélations	décroissance exponentielle ou par cycles amortis	$\rho(h) = 0$ si $ h > q$
autocorrélations partielles	$r(h) = 0$ si $ h > p$	décroissance exponentielle ou par cycles amortis

Partie 2 : Processus ARMA et leur estimation
5 - Estimation de la moyenne et des autocorrélations

Estimation de la moyenne et de l'autocovariance d'un processus stationnaire

Si (X_t) est un processus stationnaire alors sa moyenne μ_X et son autocovariance γ_X (ou son autocorrélation ρ_X) permettent de le caractériser. Nous allons voir comment les estimer à partir des observations.

Estimation de μ

Soit $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire de moyenne μ et d'autocovariance γ . On construit un estimateur de μ à partir d'observations X_1, \dots, X_n .

Un estimateur **sans biais** naturel de la moyenne μ d'un processus stationnaire (X_t) est la moyenne empirique :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

Estimation de μ

Propriété

Si (X_t) est un processus stationnaire de moyenne μ et de fonction d'autocovariance γ , alors, lorsque n tend vers l'infini, on a, si $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$

$$n\mathbb{E}[(\bar{X}_n - \mu)^2] \rightarrow \sum_{h \in \mathbb{Z}} \gamma(h).$$

et

$$\text{Var}(\bar{X}_n) = \mathbb{E}[(\bar{X}_n - \mu)^2] \rightarrow 0$$

Estimation de μ

Théorème

Soit (X_t) un processus linéaire tel que

$$X_t = \mu + \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k}$$

(avec $Z_t \sim BB(0, \sigma^2)$ un bruit blanc fort et $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$).

Si $\sum_{k \in \mathbb{Z}} \psi_k \neq 0$, alors

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu)$$

où $\nu = \sum_{h \in \mathbb{Z}} \gamma(h) = \sigma^2 (\sum_{h \in \mathbb{Z}} \psi_k)^2$.

Estimation de γ et de ρ

Un estimateur de l'autovariance $\gamma(h)$ est donné par l'autocovariance empirique :

$$\hat{\gamma}(h) = \begin{cases} \frac{1}{n} \sum_{k=1}^{n-h} (X_{k+h} - \bar{X}_n)(X_k - \bar{X}_n) & \text{si } 0 \leq h < n \\ \hat{\gamma}(-h) & \text{si } -n < h \leq 0 \end{cases}$$

et on définit l'autocorrélation empirique pour tout h tel que $|h| < n$ par :

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

Estimation de γ et de ρ

Propriété

Pour tout $n \geq 1$, la matrice des autocovariances empiriques

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{pmatrix}$$

est définie positive.

Normalité asymptotique des autocorrélations empiriques

Théorème (admis)

Soit (X_t) un processus linéaire tel que

$$X_t = \mu + \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k}$$

(avec $Z_t \sim BB(0, \sigma^2)$ un bruit blanc fort et $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$) et si $\mathbb{E}[Z_t^4] < \infty$, alors pour tout $h > 0$,

$$\sqrt{n} \left(\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} - \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, W)$$

avec la matrice de variance-covariance de taille $h \times h$, $W = (w_{i,j})$ donnée par la formule de Bartlett :

$$w_{i,j} = \sum_{k \neq 1} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)) (\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k))$$

Partie 2 : Processus ARMA et leur estimation
6 - Processus $ARMA(p, q)$

Généralisation : Processus ARMA(p, q)

Le théorème suivant donne une condition nécessaire et suffisante d'existence d'une solution stationnaire à l'équation définissant un processus ARMA(p, q).

Théorème (Existence des processus ARMA(p, q))

Soit Z_t un bruit blanc $BB(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que les polynômes $\phi(z)$ et $\theta(z)$ n'ont pas de racines communes. Alors l'équation :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

admet une solution stationnaire au second ordre si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| = 1$. Cette solution est unique et s'écrit :

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}$$

où ψ_k est la suite des coefficients du développement en série de Laurent de $\theta(z)/\phi(z)$ au voisinage du cercle unité.

Généralisation : Processus ARMA(p, q) stationnaire

Idée de la preuve :

L'existence du processus stationnaire ARMA(p, q) repose sur l'inversibilité du polynôme retard $\phi(B)$.

On a le lemme suivant :

Lemma

Le polynôme $\phi(B)$ est inversible si et seulement si les racines du polynôme $\phi(z)$ sont de module différent de 1.

Admis dans le cas général.

Pour le processus AR(1) d'équation $X_t = \phi_1 X_{t-1} + Z_t$, nous avons vu que le polynôme $\phi(z) = 1 - \phi_1 z$ admet pour racine $1/\phi_1$ et le processus AR(1) est stationnaire si et seulement si $|\phi_1| \neq 1$.

Processus ARMA(p, q) causal

Théorème (Existence d'un processus "causal")

Soit Z_t un bruit blanc $BB(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que les polynômes $\phi(z)$ et $\theta(z)$ n'ont pas de racines communes. Alors l'équation :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

admet une solution stationnaire au second ordre **causale** si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| \leq 1$. Cette solution est unique et s'écrit :

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$$

où ψ_k est la suite des coefficients du développement en série de Laurent de $\theta(z)/\phi(z) = \sum_{k=0}^{+\infty} \psi_k z^k$ qui converge sur le disque unité.

On dit que cette solution est **causale** c'est-à-dire que la solution (X_t) s'exprime uniquement en fonction du passé du processus (Z_t).

Cas du Processus AR(p) causal

Si les racines du polynôme autorégressif $\phi(Z)$ sont toutes de module strictement supérieur à 1 alors on a :

$$X_t = \sum_{k=0}^{+\infty} \psi_k Z_{t-k}$$

et dans ce cas on a $\sum_{k=0}^{\infty} |\psi_k| < \infty$, et $\psi_0 = 1$.

Et X_{t-1}, X_{t-2}, \dots s'expriment linéairement en fonction de Z_{t-1}, Z_{t-2}, \dots et sont donc non corrélées avec Z_t :

$$\mathbb{E}[Z_t X_{t-k}] = 0, \text{ pour } k \geq 1$$

Dans ce cas, le bruit blanc (Z_t) est appelé l'innovation du processus X_t .

- Pour AR(1) : $X_t = \sum_{k=1}^{\infty} \phi_1^k Z_{t-k}$

Cas du Processus AR(p) causal

- On a $\psi_0 = 1$. En effet avec $X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$

$$\mathbb{E}[Z_t X_t] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[Z_t Z_{t-k}] = \psi_0 \sigma^2$$

et d'autre part, avec l'équation autorégressive :

$$\mathbb{E}[Z_t X_t] = \mathbb{E}[Z_t(\phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t)] = \sigma^2.$$

Cas du Processus $MA(q)$

La définition d'un processus $MA(q)$ est explicite et ne pose donc pas de problème : le processus X_t tel que

$$X_t = \sum_{k=0}^q \theta_k Z_{t-k}$$

est parfaitement défini et est automatiquement stationnaire.

Par définition, un processus $MA(q)$ est toujours causal puisqu'il s'exprime en fonction du passé de (Z_t) .

Processus ARMA(p, q) inversible

Théorème (Existence d'un processus "inversible")

Soit Z_t un bruit blanc $BB(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que les polynômes $\phi(z)$ et $\theta(z)$ n'ont pas de racines communes. Soit (X_t) vérifiant

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

alors il existe une suite (π_k) absolument sommable telle que

$$Z_t = \sum_{k=0}^{\infty} \pi_k X_{t-k}$$

si et seulement si le polynôme $\theta(z) \neq 0$ pour $|z| \leq 1$. (π_k) est la suite des coefficients du développement en série de Laurent de

$\phi(z)/\theta(z) = \sum_{k=0}^{+\infty} \pi_k z^k$ qui converge sur le disque unité.

On dit que le processus (X_t) est **inversible**.

Processus ARMA(p, q) inversible

Idee de la preuve :

L'inversibilité du processus ARMA(p, q) repose sur l'inversibilité du polynôme retard $\theta(B)$. C'est la notion duale de la causalité. De manière analogue à l'AR(p), on peut montrer que :

- si les racines de θ sont toutes de module strictement supérieur à 1 alors

$$\theta^{-1}(B)X_t = \sum_{k=0}^{+\infty} \pi_k X_{t-k} = Z_t$$

Le processus MA(q) est alors stationnaire, causal et inversible.

- Pour le processus MA(1) d'équation $X_t = Z_t - \theta Z_{t-1}$, et $|\theta| < 1$ on a :

$$Z_t = (1 - \theta B)^{-1} X_t = \sum_{k=0}^{\infty} \theta^k B^k X_t = X_t + \sum_{k=1}^{\infty} \theta^k X_{t-k}$$

Le processus MA(1) admet alors une écriture autorégressive "infinie" .

En résumé :

Théorème (Causalité et inversibilité d'un processus ARMA)

Soit Z_t un bruit blanc $BB(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que les polynômes $\phi(z)$ et $\theta(z)$ n'ont pas de racines communes. On considère l'équation ARMA(p, q) définissant le processus (X_t) :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

en suppose que le polynôme ϕ n'a pas de racines de module 1, (assure l'existence d'une solution stationnaire unique), alors :

- ▶ la solution (X_t) est **causale** si ϕ n'a pas de racine de module ≤ 1 ;
- ▶ la solution (X_t) est **inversible** si θ n'a pas de racine de module ≤ 1 ;

Dans toute la suite, on considère que les polynômes ϕ et θ n'ont pas de racines communes, et que leurs racines sont toutes de module strictement supérieur à 1. (Quitte à se ramener à ce cas en changeant de bruit blanc)

Dans ce cas on dit qu'on a une **représentation canonique** du processus ARMA(p, q) :

- Le processus est **stationnaire**
- Le processus est **causal**
- Le processus est **inversible**
- le processus de bruit blanc est **l'innovation** c'est-à-dire que Z_t est non corrélé au passé X_{t-1}, X_{t-2}, \dots

Exercice 2 : 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. $X_t + 0.01X_{t-2} = Z_t$;

5. $X_t - 0.5X_{t-1} + X_{t-2} = Z_t$;

Exercice 3 : Soit l'équation ARMA(1, 2) :

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

1. Montrer qu'il existe une unique solution stationnaire ;
2. Cette solution est-elle causale ?
3. Calculer explicitement cette solution en fonction de Z ; en déduire que la solution est causale.
4. Cette solution est-elle inversible ?

Partie 2 : Processus ARMA et leur estimation

7 - La méthode de Box-Jenkins

La méthode de Box et Jenkins pour modéliser une série temporelle

La méthode de Box-Jenkins (1970) est une méthode de modélisation pour les séries temporelles qui s'appuie sur les processus ARMA. C'est sans doute la méthode la plus utilisée. Elle repose sur trois grandes étapes :

1. Identification
2. estimation
3. validation

Etape 1 : Identification

La première étape consiste à identifier le modèle $ARMA(p, q)$ qui pourrait engendrer la série.

- ▶ On commence par transformer la série afin de la rendre stationnaire en appliquant des différentiations successives $\Delta = 1 - B$, $\Delta^d = (1 - B)^d$, différentiations saisonnières $\Delta_k = (1 - B^k)$, ou transformations (logarithmiques, ...)
- ▶ On identifie un ou plusieurs modèles $ARMA(p, q)$ de la série transformée à l'aide des corrélogrammes simples et partiels et du corrélogramme partiel :
On détermine des valeurs p_{max} et q_{max} : on considèrera comme modèles candidats l'ensemble des modèles $ARMA(p, q)$ avec $0 \leq p \leq p_{max}$ et $0 \leq q \leq q_{max}$.

Etape 2 : Estimation et sélection de modèle

La deuxième étape consiste à estimer les coefficients du/des modèle(s) $ARMA(p, q)$ pour les valeurs de p et q retenues. Les paramètres d'un processus ARMA peuvent être estimés à l'aide de la fonction d'autocorrélation car on dispose d'estimateurs convergents pour la fonction d'autocorrélation.

Méthode 1 :

- ▶ On estime d'abord les coefficients ϕ_1, \dots, ϕ_p de la partie autorégressive à l'aide des équations de Yule-Walker (que nous avons vues pour un AR(p) pur et qui se généralise pour un ARMA(p,q).
- ▶ puis une fois les ϕ_1, \dots, ϕ_p estimés, on utilise la relation qui lie les coefficients $\theta_1, \dots, \theta_q$ à la fonction d'autocorrélation.
- ▶ enfin on estime la variance σ^2 .
Cette méthode n'est pas celle qui est implémentée dans la fonction `arima` de R.

Etape 2 : Estimation et sélection de modèle

Méthode 2 :

- ▶ On peut aussi utiliser la représentation auto-régressive infinie (processus inversible) et estimer les coefficients ϕ_k et θ_j en minimisant la somme des carrés :

$$\min_{\phi, \theta} \sum_{t=1}^N (X_t - \sum_{i=1}^{\infty} \pi_i^{\phi, \theta} X_{t-i})^2$$

C'est la méthode CSS de la fonction `arima` de R (voir TP). Cette méthode ne fait pas d'hypothèse supplémentaire sur le processus (Z_t) .

- ▶ On obtient ensuite directement un estimateur de la variance σ^2 .

Etape 2 : Estimation et sélection de modèle

Méthode 3 :

- ▶ Si le processus Z_t est un **bruit blanc gaussien** alors on peut écrire une vraisemblance et obtenir les estimateurs ϕ_k , θ_j et σ^2 par maximisation de cette vraisemblance.
- ▶ Cette maximisation utilise des algorithmes numériques itératifs et nécessite de se donner des valeurs initiales des paramètres. C'est la méthode ML de la fonction `arima` de R (voir TP). On peut choisir d'initialiser l'algorithme avec les valeurs obtenues par la méthode CSS : c'est la méthode CSS-ML (voir TP).

Etape 2 : Estimation et sélection de modèle

Une fois que tous les modèles "candidats" ont été estimés, il faut choisir lequel retenir (quelle valeurs de p et q ?). On peut utiliser des critères de comparaison :

▶ **Critère d'Akaike (AIC) :**

sans vraisemblance :

$$AIC(p, q) = N \log(\hat{\sigma}^2) + 2K$$

avec vraisemblance :

$$AIC(p, q) = -2 \log(L) + 2K$$

▶ **Critère de Schwarz (BIC) :**

On remplace dans les formules de l'AIC la pénalité $2K$ par $K \times \log(N)$ où K est le nombre de paramètres estimés dans le modèle. Le critère BIC pénalise davantage que AIC dès que $\log(N) > 2$.

Etape 2 : Estimation et sélection de modèle

Les critères AIC et BIC sont basés sur le principe de la pénalisation. En effet, on peut montrer que si on passe d'un modèle $ARMA(p, q)$ à un modèle $ARMA(p + 1, q)$ (ou $ARMA(p, q + 1)$), alors la variance estimée diminue (ou la log-vraisemblance augmente). On cherche un compromis entre une faible variance (ou une log-vraisemblance grande) et un faible nombre de paramètres à estimer : c'est le principe de **parcimonie** .

Etape 3 : Validation du modèle retenu

La troisième étape consiste à vérifier si le modèle estimé est cohérent avec l'hypothèse de bruit blanc. Pour cela les résidus obtenus à partir du modèle estimé sont utilisés pour vérifier s'ils se comportent comme des erreurs bruit blanc à l'aide d'un test dit **test du "portmanteau"**.

On note $\hat{\rho}_{\hat{Z}}(h)$ la fonction d'autocorrélation empirique des résidus estimés \hat{Z}_t . On considère la statistique de Box-Pierce Q_{BP} définie par :

$$Q_{BP} = \sum_{h=1}^H \hat{\rho}_{\hat{Z}}(h)^2$$

Sous l'hypothèse $\mathcal{H}_0 : Z_t \sim BB(0, \sigma^2)$ alors on peut montrer que si $H > p + q$

$$Q_{BP} \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \chi^2(H - p - q)$$

Partie 2 : Processus ARMA et leur estimation
8 - prévision d'un ARMA

Méthode de prévision d'un processus ARMA

Une fois le modèle choisi et les paramètres estimés, il est possible de faire de la prévision.

Utilisation de la forme Moyenne Mobile $MA(\infty)$

On dispose de X_1, \dots, X_T observations d'une série stationnaire et on souhaite prédire la valeur de X_{T+h} , pour $h > 0$. Partant de la forme causale $X_{t+h} = \sum_{j \geq 0} \psi_j Z_{t+h-j}$ (si elle existe), la prévision de X_{T+h} connaissant X_1, \dots, X_T notée $X_{T+h|T}$ à l'horizon $h > 0$ est donnée par la combinaison linéaire des valeurs passées du bruit blanc Z :

$$X_{T+h|T} = \sum_{j \geq h} \psi_j Z_{T+h-j}$$

Méthode de prévision d'un processus ARMA

Utilisation de la forme Moyenne Mobile $MA(\infty)$

L'erreur de prédiction (ou de prévision) à l'horizon h est donnée par :

$$X_{T+h} - X_{T+h|T} = \sum_{j=0}^{h-1} \psi_j Z_{T+h-j}$$

Remarque : L'erreur de prévision à l'horizon 1 est donc égale au bruit blanc Z_{T+1} .

Méthode de prévision d'un processus ARMA

Utilisation de la forme Moyenne Mobile $MA(\infty)$

La variance de l'erreur de prédiction est donnée par :

$$\mathbb{E}[(X_{T+h} - X_{T+h|T})^2] = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

Remarque : La variance de l'erreur de prédiction à l'horizon h d'un processus ARMA croît depuis la variance du bruit σ^2 (valeur prise pour $h = 1$) jusqu'à la variance du processus lui-même.

Méthode de prévision d'un processus ARMA

Intervalle de confiance pour la prévision

Si on suppose que le bruit Z_t (d'innovation) est **gaussien**, alors les X_t sont elles aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, dans le cas d'un bruit blanc fort gaussien, on peut obtenir des intervalles de confiance pour la prévision à l'horizon h :

$$\mathbb{P} \left(X_{T+h} \in \left[X_{T+h|T} - z_{1-\alpha/2} \sigma(h); X_{T+h|T} + z_{1-\alpha/2} \sigma(h) \right] \right) = 1 - \alpha$$

où $\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ est la variance de l'erreur de prédiction à l'horizon h et $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite.

Méthode de prévision d'un processus ARMA

Mise à jour Quand une nouvelle observation X_{T+1} est disponible, la prévision à l'horizon h de X_{T+h} peut être mise à jour de façon récursive. On dispose d'une observation supplémentaire X_{T+1} , donc on ajoute le terme d'indice $h - 1$, on a :

$$\begin{aligned} X_{T+h|T+1} &= \sum_{j \geq h-1} \psi_j Z_{T+h-j} \\ &= X_{T+h|T} + \psi_{h-1} Z_{T+1} \end{aligned}$$

et aussi $X_{T+1} - X_{T+1|T} = \psi_0 Z_{T+1} = Z_{T+1}$ ($\psi_0 = 1$) donc :

$$X_{T+h|T+1} = X_{T+h|T} + \psi_{h-1}(X_{T+1} - X_{T+1|T})$$

Autrement dit, on ajoute à la prévision précédente $X_{T+h|T}$ de X_{T+h} une correction proportionnelle à l'erreur ($X_{T+1} - X_{T+1|T}$) que l'on avait faite en prédisant, avant de l'avoir observée, la nouvelle donnée X_{T+1} .

Méthode de prévision d'un processus ARMA

Mise à jour en pratique

- ▶ une fois que l'on a choisi les ordres p et q , puis estimé (par max. de vraisemblance) les paramètres ϕ_1, \dots, ϕ_p et $\theta_1, \dots, \theta_q$.
- ▶ On estime alors la suite des coefficients $\psi_0, \psi_1, \dots, \psi_{h-1}$ (obtenus par le développement en série de la fraction des polynômes $\theta(z)/\phi(z)$).
- ▶ on initialise les h premières valeurs $\hat{X}_{1|0} = \dots = \hat{X}_{h|0} = 0$, puis avec la formule de mise à jour :

$$\hat{X}_{T+h|T+1} = \hat{X}_{T+h|T} + \hat{\psi}_{h-1}(X_{T+1} - \hat{X}_{T+1|T})$$

Méthode de prévision d'un processus ARMA

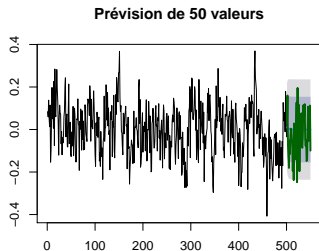
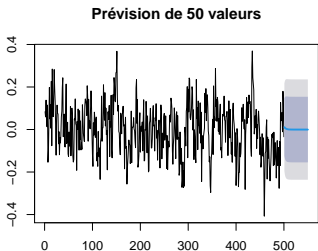
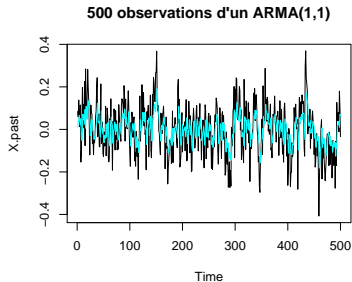
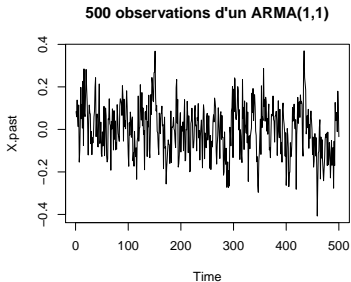
Intervalle de prévision en pratique Une fois que l'on a pu estimer $X_{T+h|T}$ et les ψ_k , l'intervalle de prévision de X_{T+h} de niveau $1 - \alpha$ devient alors

$$\left[\hat{X}_{T+h|T} \pm z_{1-\alpha/2} \hat{\sigma} \sqrt{\sum_{j=0}^{h-1} \hat{\psi}_j^2} \right]$$

Méthode de prévision d'un processus ARMA

Il est possible aussi de donner des prévisions d'un processus ARMA en exploitant la forme autorégressive. Cette méthode n'est pas abordée.

Exemple de prévision d'un processus ARMA(1,1)



Partie 3 : Séries temporelles non-stationnaires

Séries Intégrées

Les processus ARIMA et leur version saisonnière SARIMA sont des processus **non-stationnaires** qui se ramènent après différenciation simple ou saisonnière à des processus ARMA.

Séries Intégrées : Modèles ARIMA

Définition

Une série est dite *intégrée d'ordre d* s'il faut la différencier d fois pour obtenir une série stationnaire. Ainsi :

Un processus $ARIMA(p, d, q)$ est un processus dont la différence d'ordre d est un $ARMA(p, q)$.

Exemple : La série (X_t) vérifiant

$$(1 - B)X_t = c + \frac{1 + \theta_1 B + \theta_2 B^2}{1 - \phi_1 B - \phi_3 B^3} Z_t$$

est un $ARIMA(3, 1, 2)$. La constante c s'appelle la *dérive (drift)* et (Z_t) est un bruit blanc.

Séries Intégrées : Modèles ARIMA

Exemple (suite) : La série (X_t) vérifiant

$$(1 - B)X_t = c + \frac{1 + \theta_1 B + \theta_2 B^2}{1 - \phi_1 B - \phi_3 B^3} Z_t$$

peut se réécrire :

$$(1 - (1 + \phi_1)B + \phi_1 B^2 - \phi_3 B^3 + \phi_3 B^4)X_t = b + (1 + \theta_1 B + \theta_2 B^2)Z_t$$

avec $b = c(1 - \phi_1 - \phi_3)$. Mais le polynôme d'autorégression admet 1 comme racine et la série n'est pas stationnaire. Alors on n'écrit pas que c'est un ARMA(4,2) mais un ARIMA(3,1,2).

Séries Intégrées : Modèles SARIMA

Considérons à présent la version saisonnière des séries intégrées.

Définition

Une série (X_t) est un *SARIMA* $(p, d, q)(P, D, Q)_s$ si elle vérifie l'équation :

$$(1 - B)^d(1 - B^s)^D X_t = c + \frac{\theta(B)\Theta_s(B)}{\phi(B)\Phi_s(B)} Z_t$$

où (Z_t) est un bruit blanc et

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \\ \Phi_s(B) &= 1 - \Phi_1 B^s - \dots - \Phi_P B^{sP} \\ \Theta_s(B) &= 1 + \Theta_1 B^s + \dots + \Theta_Q B^{sQ} \end{aligned}$$

Séries Intégrées : Modèles SARIMA

Exemple : La série (X_t) vérifiant

$$(1 - B)(1 - B^{12})X_t = c + \frac{1 + \theta_1 B}{1 - \Phi_1 B^{12}} Z_t$$

où $|\Phi_1| < 1$ est un $SARIMA(0, 1, 1)(1, 1, 0)_{12}$. On a $p = 0, d = 1, q = 1$ et

$P = 1, D = 1, Q = 0$ et $s = 12$.

En différenciant la série simplement une fois et saisonnièrement une fois avec saisonnalité 12, on obtient une série stationnaire.

Séries Intégrées : Modèles SARIMA

Exemple : La série (X_t) vérifiant

$$(1 - B^4)X_t = c + \frac{1 + \theta_1 B}{(1 - \phi_1 B)(1 - \Phi_1 B^4 - \Phi_2 B^8)} Z_t$$

est un $SARIMA(1, 0, 1)(2, 1, 0)_4$. On a $p = 1, d = 0, q = 1$ et

$P = 2, D = 1, Q = 0$ et $s = 4$.

La série différenciée saisonnièrement une fois de saisonnalité 4 est stationnaire, lorsque que les racines des polynômes $1 - \phi z$ et $1 - \Phi_1 z^4 - \Phi_2 z^8$ sont strictement supérieures à 1 en module.

Séries Intégrées : Modèles SARIMA

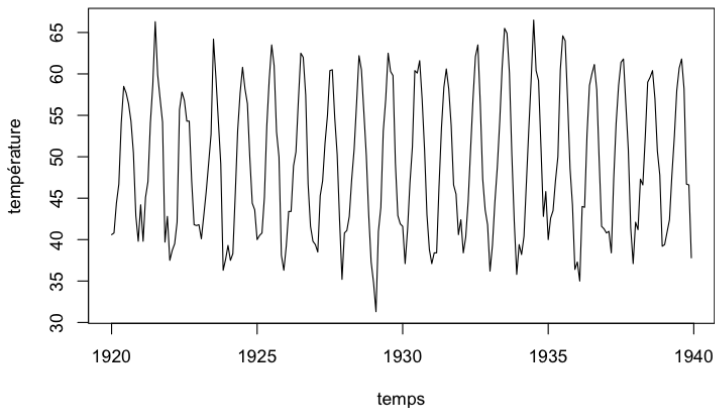
Exemple (suite) : Estimation du modèle $SARIMA(1,0,1)(2,1,0)_4$ avec R :

$$(1 - B^4)X_t = c + \frac{1 + \theta_1 B}{(1 - \phi_1 B)(1 - \Phi_1 B^4 - \Phi_2 B^8)} Z_t$$

```
Arima(Xt, order=c(1,0,1), seasonal=list(order=c(2,1,0),  
period=4),include.drift=TRUE)
```

Exemple complet d'estimation d'un SARIMA

Température de l'air moyenne mensuelle à Nottingham Castle en degrés Fahrenheit, de janvier 1920 à décembre 1939.



Partageons la série en deux sous-séries :

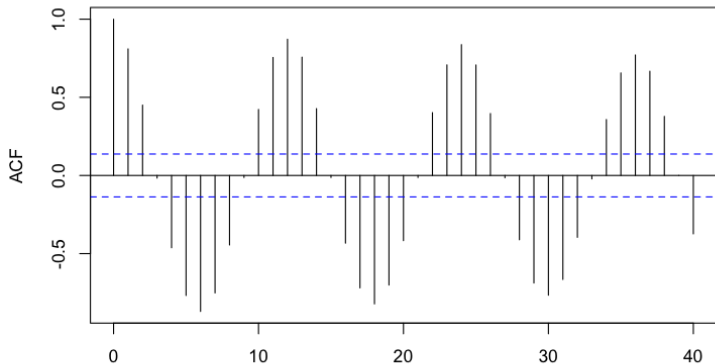
- ▶ Le début 1920 à 1936 servira à la modélisation, c'est la série d'apprentissage.
- ▶ on comparera prévisions et réalisations sur les années 1937 à 1939, c'est la série de validation.

```
data(nottem)
plot.ts(nottem,xlab='temps',ylab='température')

nott_a = window(nottem, end=c(1936,12))
nott_v = window(nottem, start=c(1937,1))
```

Modélisation

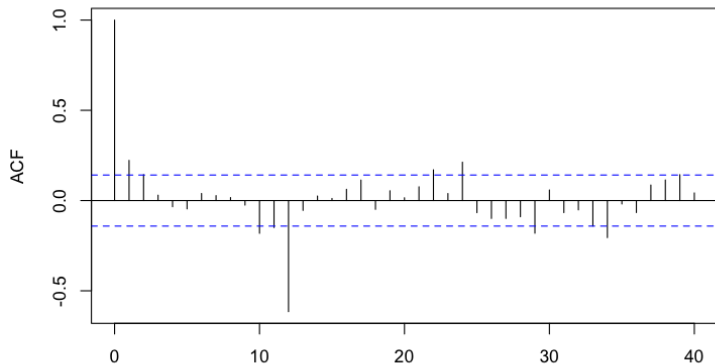
Autocorrélations de la série des températures X_t



On observe que l'ACF présente des pics aux retards multiples de 12 qui ne s'atténuent que très lentement : c'est le signe d'une **non-stationnarité saisonnière**.

Modélisation

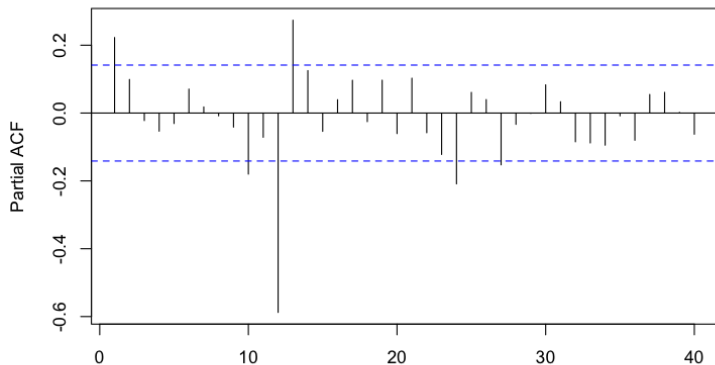
Autocorrélations de la série différenciée saisonnièrement $(1 - B^{12})X_t$:



Les autocorrélations de la série différenciée à l'ordre 12 présente un pic à l'ordre 12 et une valeur non nulle en 24 puis une forte atténuation : caractéristique d'une saisonnalité.

Modélisation

Autocorrélations partielles de la série différenciée $(1 - B^{12})X_t$:



Le PACF de la série différenciée à l'ordre 12 présente un pic à l'ordre 12 et à l'ordre 13, ainsi qu'à l'ordre 1.

Modélisation

Nous proposons de modéliser cette série par un $SARIMA(1, 0, 0)(2, 1, 0)_{12}^3$ soit :

$$(1 - B^{12})X_t = \frac{1}{(1 - \phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})} Z_t$$

avec cette modélisation, la partie autorégressive de la série différenciée est :

$$\begin{aligned} & (1 - \phi_1 B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24}) \\ = & 1 - \phi_1 B^{12} - \Phi_2 B^{24} - \phi_1 B + \phi_1 \Phi_1 B^{13} + \phi_1 \Phi_2 B^{25} \end{aligned}$$

ce qui est en accord avec les ordres identifiés par le PACF.

D'autres modèles pourraient aussi être proposés comme des candidats.

3. modèle proposé dans Anderson O.D. (1976). Time Series Analysis and Forecasting : The Box-Jenkins approach. Butterworths

Estimation avec R

$SARIMA(1, 0, 0)(2, 1, 0)_{12}$

```
fitm = arima(nott_a, order=c(1,0,0), list(order=c(2,1,0), period=12))
summary(fitm)

## Coefficients:
##          ar1      sar1      sar2
##          0.324  -0.8848  -0.3042
## s.e.      0.069   0.0708   0.0752
##
## sigma^2 estimated as 5.76:  log likelihood = -445.44,  aic = 898.88
```

Les écart-types permettent d'observer que les coefficients estimés sont significatifs.

Analyse des résidus

On teste la blancheur du bruit avec le test du portmanteau :

```
Box.test(fitm$residuals,lag=15)
Box.test(fitm$residuals,lag=20)

## Box-Pierce test
##
## data:  fitm$residuals
## X-squared = 13.709, df = 15, p-value = 0.5477

## Box-Pierce test
##
## data:  fitm$residuals
## X-squared = 16.861, df = 20, p-value = 0.662
```

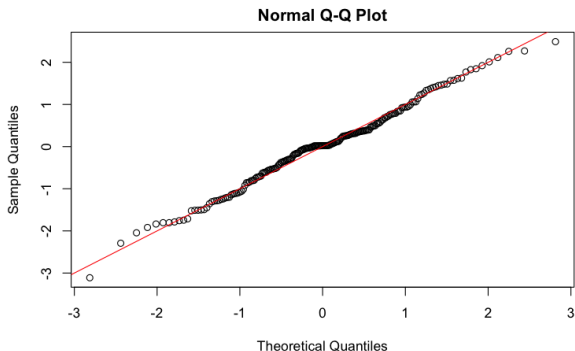
On ne rejette pas la nullité des autocorrélations jusqu'à l'ordre 15 ou 20.

Analyse des résidus

On teste la normalité des résidus :

```
shapiro.test(fitm$residuals)
qqnorm(fitm$residuals/sqrt(fitm$sigma2))
abline(a=0,b=1,col="red")
```

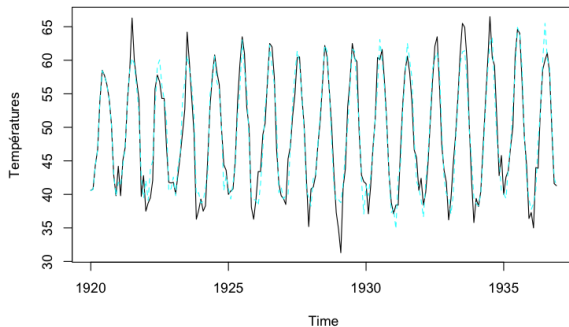
```
## Shapiro-Wilk normality test
##
## data:  fitm$residuals
## W = 0.99196, p-value = 0.3228
```



Valeurs "ajustées" sur l'échantillon d'apprentissage

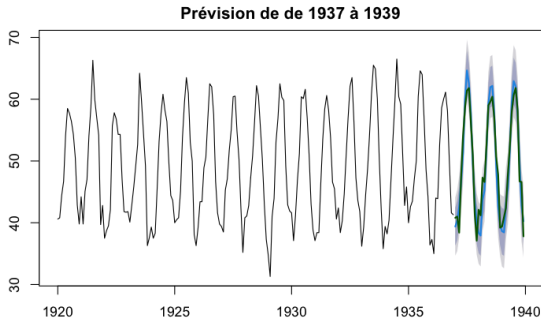
```
library(forecast)
pred.sarima = forecast(fitm, h=36)

plot(nott_a)
points(pred.sarima$fitted, col="cyan", lty=2, type="l")
```



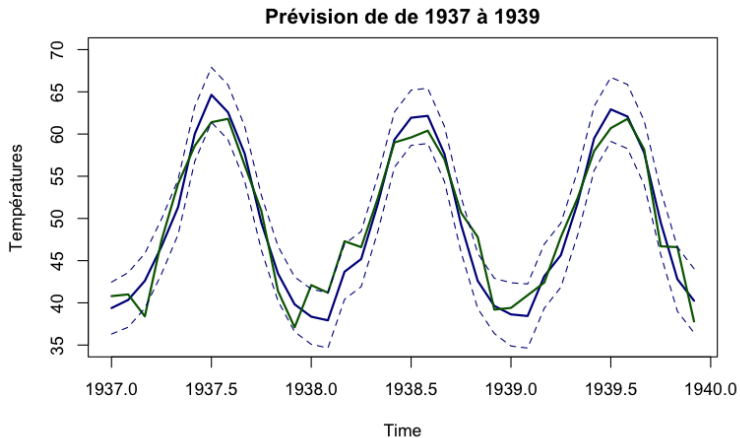
Prévisions de 1937 à 1939 et comparaison avec l'échantillon de validation

```
plot(pred.sarima,main="Prévision de de 1937 à 1939")  
points(nott_v,col='darkgreen',lwd=2,type='l')
```



Prévisions de 1937 à 1939 et comparaison avec l'échantillon de validation

Observons les prévisions de plus près, et leurs intervalles de confiance de niveau 80% :



Quelques exemples particuliers de séries

Nous allons examiner quelques cas particuliers et en tirer quelques recommandations pour le choix de la modélisation.

- ▶ la différenciation n'est pas toujours recommandée
- ▶ une élimination de la tendance ou de la saisonnalité lorsque celles-ci semblent être "déterministes" par une méthode d'estimation directe peut être parfois préférable.

Série stationnaire à une tendance déterministe près

Soit le modèle :

$$X_t = a + bt + Z_t \quad \text{avec } b \neq 0$$

et Z_t défini par :

$$(1 - 0.4B)(Z_t) = (1 + 0.8B)(e_t) \quad \text{et } e_t \text{ un bruit blanc.}$$

En fait la série X_t est stationnaire à une tendance "déterministe" près, puisque Z_t est un $ARMA(1, 1)$.

Série stationnaire à une tendance déterministe près

On peut donc procéder

- ▶ soit par différenciation :

$$\Delta X_t = b + Z_t - Z_{t-1}$$

qui permet d'éliminer la tendance affine et de se ramener à un processus stationnaire : on estime alors un $ARIMA(1, 1, 1)$.

- ▶ soit en ajustant la tendance par une régression linéaire (comme dans le TP1) et travailler ensuite sur la série $X_t - \hat{a} - \hat{b}t$ pour la modéliser par un $ARMA(1, 1)$.

Cette deuxième solution peut s'avérer parfois plus efficace.

Recommandation :

- ▶ Quand une tendance linéaire (ou polynômiale) n'apparaît pas clairement et que l'équation de la tendance semble changer au cours du temps, il est préférable de différencier la série.
- ▶ En revanche si la tendance semble être régie par la même équation déterministe au cours du temps, alors on peut essayer d'ajuster un modèle à tendance "déterministe".

Série doublement différenciée

Recommandation :

- ▶ Si une série différenciée présente encore une tendance, on différencie la série initiale une seconde fois : on travaille ainsi sur la série des accroissements des accroissements.

Série avec tendance et saisonnalité

Si l'on veut différencier une série présentant une saisonnalité et une tendance, il est préférable de commencer par différencier saisonnièrement la série et d'examiner la série obtenue avant de décider de différencier aussi à l'ordre 1.

Exemple : cas d'une série trimestrielle de périodicité 4. La différenciation saisonnière se factorise en :

$$1 - B^4 = (1 - B)(1 + B + B^2 + B^3)$$

Elle contient donc une différenciation simple : souvent, la seule différenciation saisonnière suffit pour éliminer à la fois la tendance et la saisonnalité.

Partie 4 : Lissage exponentiel

Prévision par lissage exponentiel

- ▶ Introduites par Holt en 1958, puis par Winters en 1960, les méthodes de lissage constituent des **techniques empiriques de prévision** qui prennent en compte les valeurs du passé d'une série temporelle.
- ▶ Elles présentent l'intérêt d'une implémentation récursive qui en font un outil efficace pour le traitement de gros volumes de données.
- ▶ Elles sont très largement utilisées.
- ▶ pas de garanties théoriques, pas d'intervalles de confiance, etc ...

Prévision par lissage exponentiel

Toutes ces méthodes consistent à ajuster à une série temporelle une **estimation locale** des valeurs futures :

- ▶ une constante pour le lissage exponentiel simple
- ▶ une droite pour le lissage exponentiel double ou de Holt
- ▶ des fonctions polynomiales ou périodiques pour les lissages plus généraux

On considèrera les trois modèles de décomposition suivants, pour $t \in \mathbb{Z}$:

$$X_t = m_t + e_t$$

$$X_t = m_t + s_t + e_t$$

$$X_t = m_t s_t + e_t$$

où m_t et s_t désignent les composantes déterministes de tendance ou de saisonnalité, et e_t une composante irrégulière aléatoire (qui peut être ou non un bruit blanc) ;

Lissage exponentiel simple

Le lissage exponentiel simple permet d'effectuer des prévisions pour des séries temporelles dont la tendance est constante et en absence de saisonnalité. Soit $X_t = a + e_t$, une telle série dont on a observé les T valeurs X_1, \dots, X_T . Pour $h \in \mathbb{N}^*$, on cherche à prédire la valeur X_{T+h} . On dit que l'on souhaite **prédire la série à l'horizon h** .

Définition

Etant donné un réel β tel que $0 < \beta < 1$, on définit la prévision $\hat{X}_{T+h|T}$ à l'horizon h comme solution du problème de minimisation :

$$\min_a \sum_{j=0}^{T-1} \beta^j (X_{T-j} - a)^2$$

C'est la constante qui ajuste le mieux au sens de moindres carrés pondérés la série des observations X_1, \dots, X_T .

Lissage exponentiel simple

Propriété

La solution du problème de minimisation est donné par :

$$\hat{a}_T = \frac{1 - \beta}{1 - \beta^T} \sum_{j=0}^{T-1} \beta^j X_{T-j}$$

En faisant l'approximation $\beta^T \simeq 0$, lorsque T est "assez grand", la **prévision par lissage exponentiel simple** de la série (X_t) à l'horizon h , $\hat{X}_{T+h|T}$ est donnée par :

$$\hat{X}_{T+h|T} = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}$$

Lissage exponentiel simple

Quelques remarques :

- ▶ La méthode de lissage exponentiel simple prend en compte toutes les valeurs passées de la série (X_t) avec de moins en moins d'importance pour les valeurs les plus éloignées de T .
- ▶ Si β ne dépend pas de h , les prévisions sont identiques pour tout h . On peut alors noter $\hat{X}_{T+h|T} := \hat{X}_T$.
- ▶ Si β est proche de 0, la prévision est fortement influencée par les valeurs de la série les plus récentes.
- ▶ Si β est proche de 1, les valeurs passées ont plus d'influence.
- ▶ En pratique on exclut les cas $\beta = 0$ et $\beta = 1$.

Lissage exponentiel simple : Formule de mise à jour

A partir de la définition, on obtient :

$$\begin{aligned}\hat{X}_{T+h|T} &= (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j} \\ &= (1 - \beta) X_T + (1 - \beta) \sum_{j=1}^{T-1} \beta^j X_{T-j} \\ &= (1 - \beta) X_T + \beta \hat{X}_{T+h|T-1} \\ &= \hat{X}_{T+h|T-1} + (1 - \beta)(X_T - \hat{X}_{T+h|T-1})\end{aligned}$$

Lissage exponentiel simple : Formule de mise à jour

$$\underbrace{\hat{X}_{T+h|T}}_{\text{prévision à la date } T} = \underbrace{\hat{X}_{T+h|T-1}}_{\text{prévision à la date } T-1} + (1 - \beta)(X_T - \hat{X}_{T+h|T-1})$$

- ▶ La formule de mise à jour permet de calculer directement une nouvelle prévision $\hat{X}_{T+h|T}$, à partir de la prévision $\hat{X}_{T+h|T-1}$ à la date $T - 1$, lorsqu'on récupère une nouvelle observation X_T de la série à la date T .
- ▶ Cette formule de récurrence nécessite une initialisation. En général on prend $\hat{X}_{1+h|1} = X_1$. Un autre choix consiste à prendre la moyenne de toutes les valeurs de la série. Ce choix importe peu en pratique car cette initialisation est rapidement "oubliée". Cet oubli est d'autant plus rapide que la constante de lissage est proche de 1.

Lissage exponentiel simple : choix de β

Le problème le plus important en pratique est le choix de la constante β dite de **lissage** .

- ▶ Ce choix peut être subjectif.
- ▶ Ce choix peut être dicté par les données **"data-driven"** . Par exemple, on peut chercher la valeur de β qui rend minimale la somme des carrés des erreurs de prévisions aux dates $1, \dots, T - h$:

$$\sum_{t=1}^{T-h} \left(X_{t+h} - \hat{X}_{t+h|t} \right)^2$$

ou on peut aussi ne prendre en compte que les erreurs de prévision obtenues sur la seconde moitié de la série (pour "oublier" l'initialisation) :

$$\sum_{t=\lceil (T-h)/2 \rceil}^{T-h} \left(X_{t+h} - \hat{X}_{t+h|t} \right)^2$$

Lissage exponentiel simple : exemples

Choix de $\beta = 0.2$

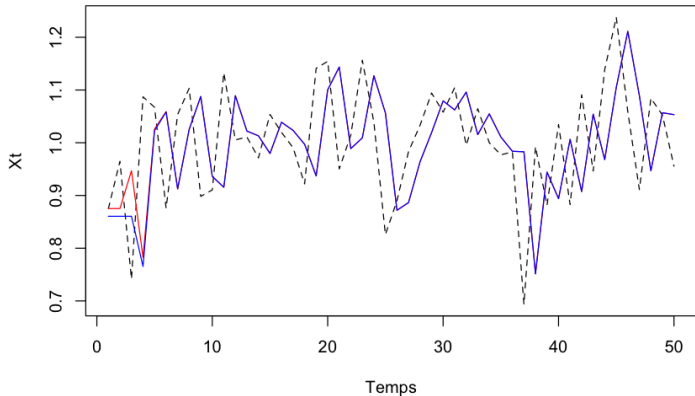


FIGURE – Influence de l'initialisation : série (X_t) (pointillés noirs), séries lissées avec initialisation par X_1 (rouge) ou par \bar{X} (bleu) avec $\beta = 0.2$

Lissage exponentiel simple : exemples

Choix de $\beta = 0.5$

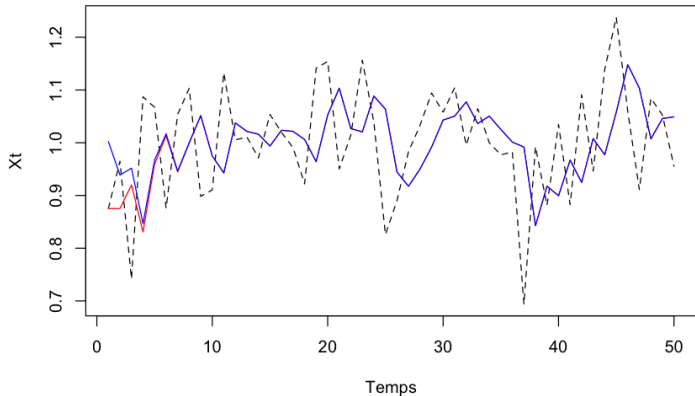


FIGURE – Influence de l'initialisation : série (X_t) (pointillés noirs), séries lissées avec initialisation par X_1 (rouge) ou par \bar{X} (bleu) avec $\beta = 0.5$

Lissage exponentiel simple : exemples

Choix de $\beta = 0.8$

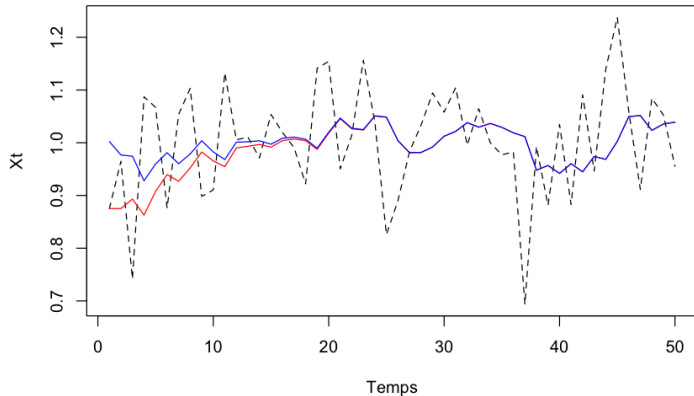


FIGURE – Influence de l'initialisation : série (X_t) (pointillés noirs), séries lissées avec initialisation par X_1 (rouge) ou par \bar{X} (bleu) avec $\beta = 0.8$

Lissage exponentiel simple : recommandations

- ▶ Quelle que soit la valeur de l'initialisation, la prévision est la même au bout d'un certain temps.
- ▶ En revanche, ce temps est d'autant plus court que β est proche de 0. Il faudra faire attention au choix de β d'autant plus si la série est courte.

Lissage exponentiel double

Le lissage exponentiel double généralise le lissage exponentiel simple au cas où la série présente une tendance **localement affine** : sur de courts intervalles de temps, elle évolue comme une droite dont l'équation peut changer légèrement au cours du temps.

On cherche une prévision à l'horizon h , $\hat{X}_{T+h|T}$ de la forme :

$$\hat{X}_{T+h|T} = \hat{a}_T + \hat{b}_T h$$

où le couple (\hat{a}_T, \hat{b}_T) minimise la somme des carrés pondérés :

$$\sum_{j=0}^{T-1} \beta^j (X_{T-j} - (a + bj))^2$$

Lissage exponentiel double

Dans ce problème de moindres carrés, la solution (\hat{a}_T, \hat{b}_T) s'obtient en annulant les dérivées partielles de la fonction $C(a, b)$:

$$C(a, b) = \sum_{j=0}^{T-1} \beta^j (X_{T-j} - (a + bj))^2$$

La solution approchée du problème des moindres carrés, lorsque T est assez grand est donnée par :

$$\begin{cases} \hat{a}_T &= 2S_1(T) - S_2(T) \\ \hat{b}_T &= \frac{1-\beta}{\beta}(S_1(T) - S_2(T)) \end{cases}$$

avec $S_1(T) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}$
et $S_2(T) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j S_1(T - j)$.

Lissage exponentiel double

On écrit les dérivées partielles de $C(a, b)$:

$$\begin{cases} \frac{\partial C}{\partial a}(a, b) = -2 \sum_{j=0}^{T-1} \beta^j (X_{T-j} - a - bj) \\ \frac{\partial C}{\partial b}(a, b) = -2 \sum_{j=0}^{T-1} j \beta^j (X_{T-j} - a - bj) \end{cases}$$

Le couple (\hat{a}_T, \hat{b}_T) vérifie :

$$\begin{cases} \sum_{j=0}^{T-1} \beta^j X_{T-j} - \hat{a}_T \sum_{j=0}^{T-1} \beta^j - \hat{b}_T \sum_{j=0}^{T-1} j \beta^j = 0 \\ \sum_{j=0}^{T-1} j \beta^j X_{T-j} - \hat{a}_T \sum_{j=0}^{T-1} j \beta^j - \hat{b}_T \sum_{j=0}^{T-1} j^2 \beta^j = 0 \end{cases}$$

On résout le problème "approché" où lorsque T est assez grand on a :

$$\sum_{j=0}^{T-1} \beta^j \simeq \frac{1}{1-\beta}; \quad \sum_{j=0}^{T-1} j \beta^j \simeq \frac{\beta}{(1-\beta)^2}; \quad \sum_{j=0}^{T-1} j^2 \beta^j \simeq \frac{\beta(1+\beta)}{(1-\beta)^3}$$

Lissage exponentiel double

Le système devient :

$$\left\{ \begin{array}{l} \sum_{j=0}^{T-1} \beta^j X_{T-j} - \hat{a}_T \times \frac{1}{1-\beta} - \hat{b}_T \times \frac{\beta}{(1-\beta)^2} = 0 \\ \sum_{j=0}^{T-1} j\beta^j X_{T-j} - \hat{a}_T \times \frac{\beta}{(1-\beta)^2} - \hat{b}_T \times \frac{\beta(1+\beta)}{(1-\beta)^3} = 0 \end{array} \right.$$

En remarquant que :

$$S_2(T) - (1 - \beta)S_1(T) = (1 - \beta)^2 \sum_{j=0}^{T-1} j\beta^j X_{T-j}$$

on obtient :

$$\left\{ \begin{array}{l} S_1(T) - \hat{a}_T - \hat{b}_T \frac{\beta}{1-\beta} = 0 \\ S_2(T) - (1 - \beta)S_1(T) - \hat{a}_T \beta - \hat{b}_T \frac{\beta(1+\beta)}{1-\beta} = 0 \end{array} \right.$$

Lissage exponentiel double

Et enfin,

$$\begin{cases} \hat{a}_T = S_1(T) - \hat{b}_T \frac{\beta}{1-\beta} \\ S_2(T) - (1-\beta)S_1(T) - \beta \left(S_1(T) - \hat{b}_T \frac{\beta}{1-\beta} \right) - \hat{b}_T \frac{\beta(1+\beta)}{1-\beta} = 0 \end{cases}$$

ou encore :

$$\begin{cases} \hat{a}_T = 2S_1(T) - S_2(T) \\ \hat{b}_T = \frac{1-\beta}{\beta} (S_1(T) - S_2(T)) \end{cases}$$

Lissage exponentiel double

Il ne reste plus qu'à vérifier que :

$$S_2(T) - (1 - \beta)S_1(T) = (1 - \beta)^2 \sum_{j=0}^{T-1} j\beta^j X_{T-j}$$

$$\begin{aligned} S_2(T) - (1 - \beta)S_1(T) &= (1 - \beta) \sum_{i=0}^{T-1} \beta^i S_1(T - i) - (1 - \beta)S_1(T) \\ &= (1 - \beta) \sum_{i=1}^{T-1} \beta^i S_1(T - i) \\ &= (1 - \beta)^2 \sum_{i=1}^{T-1} \beta^i \sum_{j=0}^{T-i-1} \beta^j X_{T-i-j} \\ &= (1 - \beta)^2 \sum_{i=1}^{T-1} \beta^i \sum_{k=i}^{T-1} \beta^{k-i} X_{T-k} \\ &= (1 - \beta)^2 \sum_{k=1}^{T-1} \sum_{i=k}^{T-1} \beta^k X_{T-k} = (1 - \beta)^2 \sum_{k=1}^{T-1} k\beta^k X_{T-k}. \end{aligned}$$

Lissage exponentiel double : Formule de mise à jour

Comme pour le lissage exponentiel simple, on obtient des formules de mises à jours lorsqu'une nouvelle observation est disponible à la date T :

$$\hat{X}_{T+h|T} = \hat{a}_T + \hat{b}_T h$$

avec :

$$\begin{cases} \hat{a}_T = \hat{a}_{T-1} + \hat{b}_{T-1} + (1 - \beta^2)(X_T - \hat{X}_{T|T-1}) \\ \hat{b}_T = \hat{b}_{T-1} + (1 - \beta)^2(X_T - \hat{X}_{T|T-1}) \end{cases}$$

Lissage exponentiel double : Initialisation et choix de β

Pour utiliser ces formules de mise à jour, il faut des valeurs initiales pour initialiser la récurrence des suites \hat{a}_T et \hat{b}_T .

On peut utiliser un critère similaire à celui introduit dans le cas du lissage exponentiel simple pour choisir la valeur du paramètre β .

Lissage exponentiel de Holt-Winters

- ▶ La méthode de lissage exponentiel double permet de traiter des séries avec une tendance linéaire mais sans saisonnalité.
- ▶ On peut définir des lissages exponentiels plus généraux sur le même principe qui vont permettre de traiter des séries avec une **saisonnalité**.
- ▶ Une méthode un peu différente, mais néanmoins assez proche du lissage exponentiel double a été introduite par Holt et Winters. Il existe deux versions de cette méthode une versions **non-saisonnière** et une version **saisonnière** .
- ▶ La différence entre la **méthode de Holt-Winters** et le lissage exponentiel double porte sur les formules de mise à jour.

Lissage exponentiel de Holt-Winters

Les formules de mise à jour du lissage exponentiel double peuvent se réécrire de la façon suivante :

$$\begin{cases} \hat{a}_T &= (1 - \beta^2)X_T + \beta^2(\hat{a}_{T-1} + \hat{b}_{T-1}) \\ \hat{b}_T &= \frac{(1-\beta)^2}{1-\beta^2}(\hat{a}_T - \hat{a}_{T-1}) + \left(1 - \frac{(1-\beta)^2}{1-\beta^2}\right) \hat{b}_{T-1} \end{cases}$$

(en remarquant que $\hat{X}_{T|T-1} = \hat{a}_{T-1} + \hat{b}_{T-1}$).

Lissage exponentiel de Holt-Winters

Remarque : cette réécriture permet de remarquer que :

- ▶ \hat{a}_T est le barycentre affecté des poids $1 - \beta^2$ et β^2 de la dernière valeur X_T observée et de sa prévision à l'horizon 1. La mise à jour corrige donc la prévision de la constante.
- ▶ De même, \hat{b}_T est le barycentre de la dernière pente prédite \hat{b}_{T-1} et de l'écart entre les deux dernières ordonnées à l'origine prédites $\hat{a}_T - \hat{a}_{T-1}$.

Lissage exponentiel de Holt-Winters

Le lissage exponentiel **non-saisonnier de Holt-Winters** est donné par :

$$\hat{X}_{T+h|T} = \hat{a}_T + h\hat{b}_T$$

où

$$\begin{cases} \hat{a}_T &= \alpha X_T + (1 - \alpha)(\hat{a}_{T-1} + \hat{b}_{T-1}) \\ \hat{b}_T &= \beta(\hat{a}_T - \hat{a}_{T-1}) + (1 - \beta)\hat{b}_{T-1} \end{cases}$$

et $\alpha \in [0, 1]$ et $\beta \in [0, 1]$.

Le lissage exponentiel double est donc un cas particulier du lissage de Holt-Winters avec un seul paramètre à régler.

L'avantage de ce lissage est d'avoir une plus grande flexibilité mais l'inconvénient est qu'il faut régler deux paramètres au lieu d'un seul !

Lissage exponentiel de Holt-Winters

Le lissage exponentiel **de Holt-Winters** se généralise pour une série saisonnière additive. La prévision est donnée par :

$$\hat{X}_{T+h|T} = \hat{a}_T + h\hat{b}_T + s_T$$

où s_T est une composante périodique de période P .

$$\begin{cases} \hat{a}_T &= \alpha(X_T - s_{T-P}) + (1 - \alpha)(\hat{a}_{T-1} + \hat{b}_{T-1}) \\ \hat{b}_T &= \beta(\hat{a}_T - \hat{a}_{T-1}) + (1 - \beta)\hat{b}_{T-1} \\ s_T &= \gamma(X_T - \hat{a}_T) + (1 - \gamma)s_{T-P} \end{cases}$$

avec α , β et γ des paramètres réels dans $[0, 1]$.

Lissage exponentiel de Holt-Winters

On a également une version **de Holt-Winters** pour une série **saisonnaire multiplicative**. La prévision est donnée par :

$$\hat{X}_{T+h|T} = (\hat{a}_T + h\hat{b}_T)s_T$$

où s_T est une composante périodique de période P .

$$\begin{cases} \hat{a}_T &= \alpha \frac{X_T}{s_{T-P}} + (1 - \alpha)(\hat{a}_{T-1} + \hat{b}_{T-1}) \\ \hat{b}_T &= \beta(\hat{a}_T - \hat{a}_{T-1}) + (1 - \beta)\hat{b}_{T-1} \\ s_T &= \gamma \frac{X_T}{\hat{a}_T} + (1 - \gamma)s_{T-P} \end{cases}$$

avec α , β et γ des paramètres réels dans $[0, 1]$.

Lissage exponentiel de Holt-Winters : exemple

avec $\alpha = 0.2$, $\beta = 0.2$ et $\gamma = 0.2$:

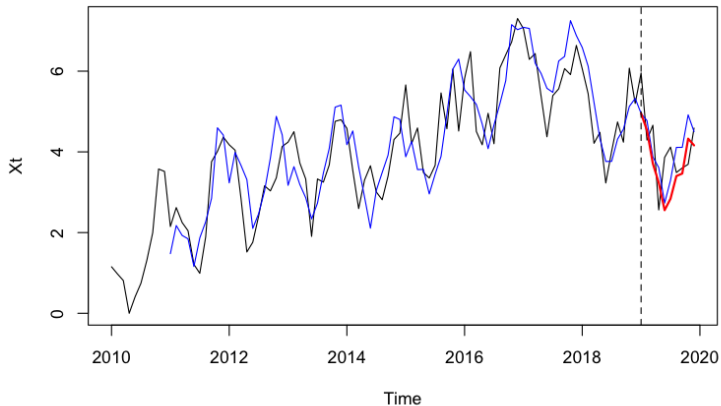


FIGURE – prévisions à l'horizon 1 (bleu) et prévisions à l'horizon h des 10 dernières valeurs (rouge)

Implémentation en R

Les méthodes de lissage exponentiel sont implémentés en R dans la fonction `HoltWinters` :

```
HoltWinters(x, alpha = NULL, beta = NULL, gamma = NULL,  
seasonal = c("additive", "multiplicative"),  
start.periods = 2, l.start = NULL, b.start = NULL,  
s.start = NULL,  
optim.start = c(alpha = 0.3, beta = 0.1, gamma = 0.1),  
optim.control = list())
```

Bibliographie

Y. Aragon, *Séries temporelles avec R*, EDP sciences, 2016.

P. Brockwell, R. Davis. *Time Series : Theory and Methods*
(bibliothèque de maths)

C. Lévy-Leduc. *Introduction à l'étude des séries temporelles*
(poly. de cours disponible sur Internet)

C. Lévy-Leduc, E. Moulines, F. Roueff. *Séries temporelles : théorie et méthodes*
(poly. de cours disponible sur Internet)