

Séries temporelles

Ahmed Kebaier

Université Paris-Saclay, Université d'Évry

ahmed.kebaier@univ-evry.fr

MAP565, 2022-23

Février 2023

▷ BIBLIOGRAPHIE

- Brockwell, P., Davis, R. : Introduction to Time Series and Forecasting. 1996, Springer.
- Brockwell, P., Davis, R. : Time Series : Theory and Methods. 2. Ed., 1991, Springer.
- Cryer, J.D., Chan K.S. : Time Series Analysis With Applications in R. 2. Ed., 2008, Springer.
- Gouriéroux, C., Monfort, A. : Séries Temporelles et Modèles Dynamiques. 2. Ed., 1995, Economica.
- Aragon, Y. : Séries temporelles avec R. 2016, Collection Pratique R, EDP sciences.
-

1 Introduction aux séries temporelles : définitions et exemples

1.1 Définitions

▷ Une série temporelle (ou chronologique) est une suite formée d'observations régulièrement espacées au cours du temps.

▷ La définition mathématique adéquate pour l'étude de telles suites consiste à les considérer comme une réalisation particulière d'une famille de variables aléatoires $X = \{X_t\}_{t \in \mathcal{I}}$ définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ où $\mathcal{I}$ est un ensemble de dates discrètes équiréparties (généralement $\mathcal{I} = \mathbb{N}$ ou $\mathbb{Z}$)

$$\begin{aligned}\mathcal{I} \times \Omega &\rightarrow \mathbb{R} \\ (t, \omega) &\mapsto X_t(\omega) = x_t\end{aligned}$$

On appelle trajectoire de X la courbe $t \mapsto x_t$ pour $t \in \mathcal{I}$.

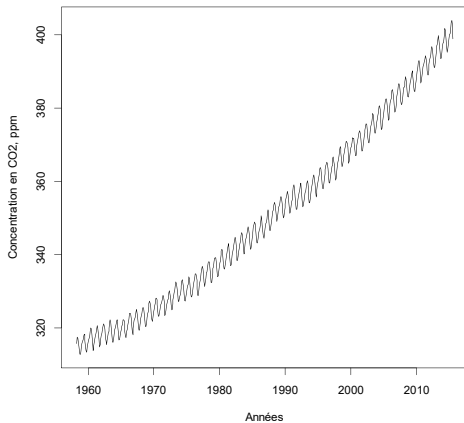
EXEMPLES :

Dans les sciences appliquées (climatologie, économie, finance, démographie,...), on est souvent amené à distinguer deux types de variables :

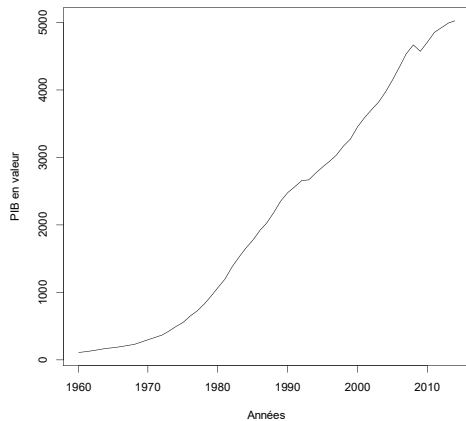
- les variables de niveau ou d'état : une observation représente un état à un moment donné (ex : prix, température,...),
- les variables de flux : une observation représente une "variation" survenue durant une période donnée (ex : consommation, rentabilité d'un actif,...).

La représentation graphique d'une série temporelle se fait sous forme d'un graphique où le temps est en abscisse et les valeurs de la variable en ordonnée. Les points sont reliés par des segments de droite.

**Evolution de la concentration en CO₂
à l'Observatoire de Mauna Loa (Hawaï)**

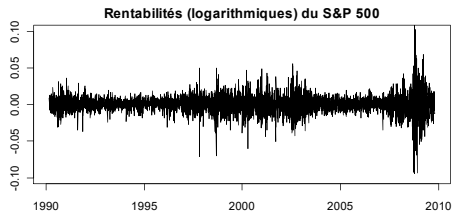
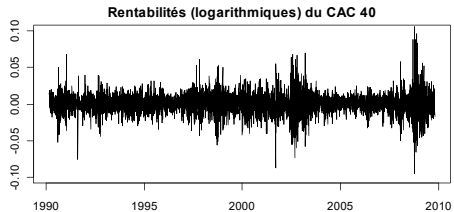


Evolution du PIB de la France, base 100 en 1959

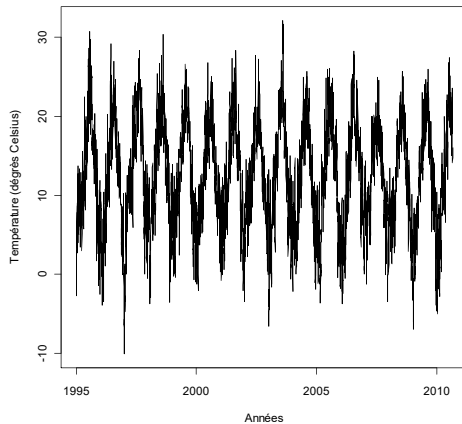


Variations du PIB en valeur de la France





Témpératures moyennes journalières à Paris



o Sous

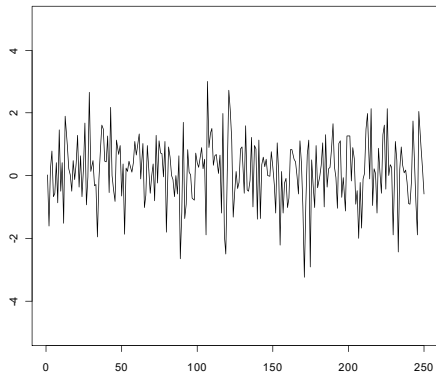
L'objet `ts` (time series) est l'objet série temporelle du langage R. Il y a cinq arguments principaux :

- 1 pour les données de la série : `data` ;
- 4 pour les dates : `start` (la date de départ), `end` (la date de fin), `frequency` (la fréquence d'échantillonnage ou le nombre d'observations par unité de temps), `deltat` (la période entre deux observations successives).

EXEMPLE :

```
> serietemp <- ts(data = rt(250, df = 10), start = 1, frequency = 1)
> plot(serietemp, ylim = c(-5,5), ylab = "", xlab = "", main =
"Réalisations de variables aléatoires iid de loi de Student à
10 degrés de liberté")
```

**Réalisations de variables aléatoires iid
de loi de Student à 10 degrés de liberté**



1.2 Problèmes statistiques

- ▷ **Description** : décomposer la série (tendance, saisonnalité, bruit) pour isoler les propriétés importantes de la série.
- ▷ **Prévision** : prévoir X_{T+h} à partir de l'observation de la série jusqu'à la date T (h = horizon de prévision).
- ▷ **Filtrage** : transformer la série de manière à éliminer certaines caractéristiques (ex : valeurs aberrantes).
- ▷ **Traitement des données manquantes** : estimer des valeurs perdues de la série.
- ▷ **Detection de rupture** : déterminer les instants où les paramètres de la série se modifient.
- ▷ **Causalité** : déterminer si une série à une influence instantanée ou décalée sur une autre.
- ▷ ...

1.3 Modélisation et modèle classique de décomposition

Important : Un *modèle* est toujours une image simplifiée de la réalité. Un *modèle* n'est pas meilleur qu'un autre tant que l'on n'a pas précisé un critère de comparaison.

Il est possible d'écrire une série temporelle comme étant une fonction déterministe du temps t et d'un aléa Y_t de manière à séparer ce qui relève du prévisible et de l'aléatoire

$$X_t = f(t, Y_t).$$

Cette représentation est cependant trop générale et présuppose que la fonction f est connue, ce qui n'est pas le cas en général.

Dans le cadre de ce cours, on supposera qu'il est possible de décomposer la série en trois termes :

$$X_t = m_t + S_t + Y_t$$

- m_t est une tendance déterministe,
- S_t est une saisonnalité déterministe,
- Y_t est une variable/perturbation aléatoire qui représente la composante ("erreur") non systématique, de moyenne nulle, et qui, soit possède une structure de corrélation non nulle et stable dans le temps (série stationnaire au sens faible, voir définition dans la section suivante), soit possède une structure non-stable mais qui peut être modélisé par un processus simple.

Il est souvent nécessaire d'effectuer une transformation sur les données initiales pour obtenir une série pour laquelle la décomposition "apparaît comme valide" (transformation par une fonction déterministe, différentiation,...).

La modélisation de la composante aléatoire est la partie la plus difficile, mais aussi la plus intéressante. Elle consiste à trouver une représentation du type suivant

$$Y_t = h(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_t, \dots, \varepsilon_{t-q})$$

où

- h est une fonction “simple” (généralement linéaire et/ou multiplicative)
- p et q ne sont pas trop “grands” pour que h ne soit pas trop difficile à estimer,
- (ε_t) est une série indépendante de (Y_t) ou non-corrélée à (Y_t) et qui ne présente pas d'autocorrélation dans le temps, i.e. les variables aléatoires ne sont pas corrélées entre elles (bruit blanc faible, voir définition dans la section suivante).

1.4 Processus stationnaires

Choisir un modèle de séries temporelles, c'est fixer des hypothèses sur la loi jointe ou simplement sur les moyennes et covariances (on parle de propriétés du second-ordre) de la suite de variables (X_t) .

Une série (X_t) est stationnaire si ses propriétés probabilistes sont les mêmes que celles de la série (X_{t+h}) , pour tout entier h .

DEFINITION :

(X_t) est stationnaire au sens strict (ou fort) si

$$(X_1, X_2, \dots, X_k) \text{ a même loi que } (X_{1+h}, X_{2+h}, \dots, X_{k+h})$$

pour tout h et tout $k \geq 1$ tels que $(h, k) \in \mathcal{I} \times \mathcal{I}$.

La stationnarité au sens strict se conserve par transformation déterministe.

PROPOSITION :

Si (X_t) est un processus stationnaire au sens strict alors pour toute fonction $f : \mathbb{R}^k \rightarrow \mathbb{R}$, le processus (Y_t) défini par

$$Y_t = f(X_t, \dots, X_{t-k})$$

est aussi stationnaire au sens strict.

La stationnarité au sens stricte impose que les lois de tous les X_t sont les mêmes. Ceci peut apparaître comme trop contraignant (en prévision notamment), on lui préfère alors la notion de stationnarité du second ordre (ou faible).

$$\text{i.e. } (X_t)_{t \in \mathbb{Z}} \in \mathcal{L}^2(\Omega)$$

DEFINITION :

Soit (X_t) telle que $\mathbb{E}[X_t^2] < \infty$ pour tout t . La fonction moyenne de (X_t) est définie par

$$\mu_X(t) = \mathbb{E}[X_t].$$

La fonction covariance de (X_t) est définie par

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s).$$

DEFINITION :

(X_t) est stationnaire au second-ordre si (ou faible)

(i) $t \mapsto \mu_X(t)$ est indépendante de t ,

(ii) $t \mapsto \gamma_X(t, t+h)$ est indépendante de t , pour tout h .

$$\underbrace{\gamma_X(t, t+h)}_{= \gamma_X(h)}$$

DEFINITION :

Soit (X_t) une série stationnaire (au second ordre). La fonction d'autocovariance de (X_t) est définie par

$$\gamma_X(h) = \text{Cov}(X_t, X_{t+h}), \quad h = 0, \pm 1, \pm 2, \dots$$

La fonction d'autocorrélation de (X_t) est définie par

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Cor}(X_t, X_{t+h}), \quad h = 0, \pm 1, \pm 2, \dots$$

PROPRIETES :

$$\gamma_X(0) = \text{Var}(X_t)$$

1. $\gamma_X(0) \geq 0$, $|\gamma_X(h)| \leq \gamma_X(0)$ pour tout h , $\gamma_X(h) = \gamma_X(-h)$, pour tout h
2. $\rho_X(0) = 1$, $|\rho_X(h)| \leq 1$ pour tout h et $\rho_X(h) = \pm 1$ ssi il existe une relation linéaire entre X_0 et X_h , $\rho_X(h) = \rho_X(-h)$, pour tout h .

3. Les fonctions d'auto-covariance et d'autocorrélation sont semi-définies positives dans le sens suivant : pour tout $n \in \mathbb{N}$, tous vecteurs $a = (a_1, \dots, a_n)$ de $\mathbb{R}^n$ et $(t_1, \dots, t_n)$ de $\mathcal{I}^n$,

$$\langle a, \gamma_X a \rangle = \sum_{i,j=1}^n a_i \gamma_X(t_i - t_j) a_j \geq 0 \quad \sum_{i,j=1}^n a_i \rho_X(t_i - t_j) a_j \geq 0. \quad (\gamma_X)_{i,j} = \gamma_X(t_i - t_j)$$

On peut en déduire, en particulier, que $\rho_X^2(1) \leq 1$ et que $(1 - \rho_X(2))(1 + \rho_X(2) - 2\rho_X^2(1)) \geq 0, \dots$ (étant donné une fonction γ semi-définie positive on peut construire un processus gaussien ayant pour auto-covariance cette fonction γ).

REMARQUES :

1. Si (X_t) est stationnaire au sens strict et si $\mathbb{E}[X_t^2] < \infty$, alors (X_t) est stationnaire au second ordre.
2. Si (X_t) est un processus Gaussien, i.e. si pour tout $n \in \mathbb{N}$ et tout $(t_1, \dots, t_n)$ de $\mathcal{I}^n$ $(X_{t_1}, \dots, X_{t_n})$ a une distribution Gaussienne, alors les notions de stationnarité au sens strict et au second ordre coïncident.

EXEMPLES DE PROCESSUS STATIONNAIRES :

DEFINITION D'UN BRUIT BLANC :

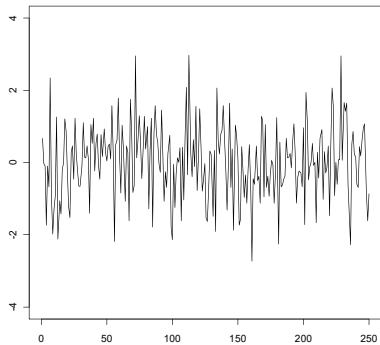
1. Un bruit blanc fort est une suite (ε_t) de variables iid (indépendantes et équadistribuées), centrées et de variance σ^2 finie. On note $(\varepsilon_t) \rightsquigarrow BBF(0, \sigma^2)$.
2. Un bruit blanc (faible) est une suite (ε_t) de variables centrées, de variances constantes et non corrélées :

$$\mathbb{E}[\varepsilon_t] = 0, \quad \text{Var}(\varepsilon_t) = \sigma^2, \quad \text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0, \text{ pour } t \neq t'.$$

On note $(\varepsilon_t) \rightsquigarrow BB(0, \sigma^2)$. Sa fonction d'autocovariance est donnée par

$$\gamma_\varepsilon(h) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

Bruit Blanc Gaussien



$$\varepsilon_t \sim \mathcal{N}(0, 1)$$

DEFINITION D'UN PROCESSUS MOYENNE MOBILE D'ORDRE 1 : (Moving average)

Un processus (X_t) est une moyenne mobile d'ordre 1 ($MA(1)$) si

$$X_t = m + \varepsilon_t + \theta \varepsilon_{t-1}, \quad (\varepsilon_t) \rightsquigarrow BB(0, \sigma^2).$$

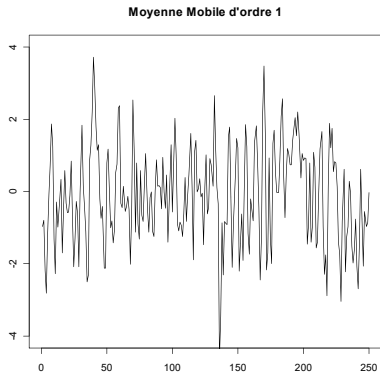
On a

$$\mathbb{E}[X_t] = m, \quad \text{Var}(X_t) = \sigma^2(1 + \theta^2),$$

et

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ \theta/(1 + \theta^2) & h = \pm 1 \\ 0 & |h| > 1 \end{cases}.$$

(à vérifier en exo)



$$m = 0, \theta = 1, \varepsilon_t \sim \mathcal{N}(0, 1)$$

DEFINITION D'UN PROCESSUS AUTOREGRESSIF D'ORDE 1 :

Un processus (X_t) est un autoregressif d'ordre 1 (AR(1)) si (O.V. dit)

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad (\varepsilon_t) \rightsquigarrow BB(0, \sigma^2), \quad |\varphi| < 1,$$

et si ε_t est non-corrélé avec les X_{t-i} pour $i > 0$.

On a

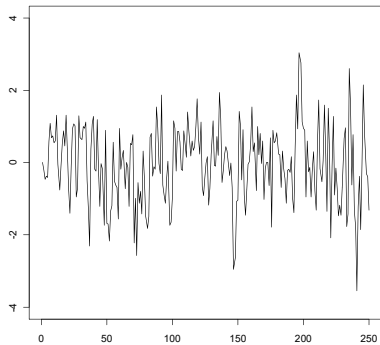
$$\mathbb{E}[X_t] = 0, \quad \mathbb{E}[X_t^2] = \sigma^2 / (1 - \varphi^2),$$

et

$$\rho_X(h) = \varphi^{|h|} \text{ pour tout } h.$$

(à vérifier en exo)

Autoregressif d'ordre 1



$$\varphi = 0,5, \varepsilon_t \sim \mathcal{N}(0,1)$$

EXEMPLES DE PROCESSUS NON-STATIONNAIRES :

DEFINITION D'UNE MARCHE ALEATOIRE :

Un processus (X_t) est une marche aléatoire (sans tendance) si

$$X_t = \sum_{i=1}^t \varepsilon_i, \quad (\varepsilon_t) \rightsquigarrow BBF(0, \sigma^2).$$

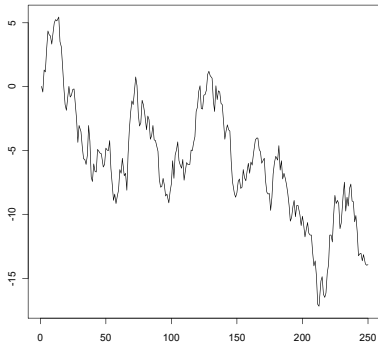
On a

$$\mathbb{E}[X_t] = 0 \quad \text{et} \quad \mathbb{E}[X_t^2] = t\sigma^2.$$

La variance dépend de t et donc X_t n'est pas stationnaire.

On parle de non-stationnarité aléatoire.

Marche aléatoire



$$W_t^H = \int_0^t (t-s)^{H-\frac{1}{2}} dW_s$$

$$t \mapsto W_t^H$$

est holderien d'ordre H

$$\varepsilon_t \sim \mathcal{N}(0, 1)$$

$$\gamma_{W_t^H} = f(t) \quad 0 < H < \frac{1}{2}$$

$$(H \geq \frac{1}{2} \quad W_t^H \simeq \text{m.b.s})$$

MODELE AVEC TENDANCE ET VARIANCE FONCTION DU TEMPS :

Un processus (X_t) est un processus avec tendance et variance fonction du temps si

$$X_t = m(t) + \kappa(t)\varepsilon_t, \quad (\varepsilon_t) \rightsquigarrow BBF(0, \sigma^2),$$

où m et κ sont deux fonctions déterministes dont l'une au moins n'est pas constante.

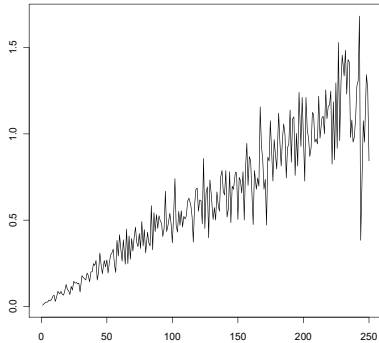
On a

$$\mathbb{E}[X_t] = m(t) \quad \text{et} \quad \mathbb{V}ar(X_t) = \sigma^2 \kappa^2(t)$$

et donc (X_t) n'est pas stationnaire.

On parle de non-stationarité déterministe.

Processus avec tendance et variance fonction du temps



$$m(t) = 0,005 \times t, \kappa(t) = 0,001 \times t, \varepsilon_t \sim \mathcal{N}(0,1)$$

2 Trend et saisonnalités

$$X_t = m_t + S_t + Y_t$$

Dans cette section, on présente des méthodes statistiques pour estimer les composantes déterministes m_t et S_t et en déduire une approximation de la composante aléatoire Y_t . On distingue deux grandes méthodes :

- les méthodes par régression qui consistent à ajuster un modèle déterministe à l'ensemble des données ou localement, (méthodes paramétriques)
- les méthodes par moyennes mobiles qui consistent à utiliser des filtres pour extraire chaque composante. (méthodes non-paramétriques)

2.1 Méthodes par régression

Ces méthodes consistent à effectuer une régression linéaire pour déterminer les deux composantes déterministes dans la décomposition

$$X_t = \underbrace{m_t} + \underbrace{S_t} + Y_t, \quad t = 1, \dots, T.$$

La méthode de Buys-Ballot généralisée consiste à spécifier les composantes déterministes comme une combinaison linéaire de fonctions connues du temps, puis à effectuer une régression linéaire ordinaire.

La méthode STL (Seasonal decomposition of Time series by Loess) consiste à spécifier les composantes déterministes comme une combinaison linéaire de fonctions simples et connues du temps, puis à effectuer localement des régressions linéaires ordinaires.

▷ La méthode de Buys-Ballot généralisée

Chacune des composantes déterministes est écrite comme combinaison linéaire de fonctions connues du temps :

$$m_t = \sum_{i=1}^k b_i m_t^{(i)} \quad S_t = \sum_{j=1}^l c_j S_t^{(j)}$$

où les b_i et c_j sont des constantes inconnues. On supposera que $m_t^{(1)} = 1$.

EXEMPLES :

1. Si l'on souhaite modéliser une tendance linéaire, on choisira $k = 2$ et $m_t^{(2)} = t$.
2. Si l'on souhaite modéliser une tendance quadratique, on choisira $k = 3$, $m_t^{(2)} = t$ et $m_t^{(3)} = t^2$.
3. Si l'on souhaite modéliser une saisonnalité d'ordre 4, on choisira $l = 4$ et $S_t^{(j)} = \mathbb{I}_{\{t=j \bmod 4\}}$.



Dans le cadre de cette section, on supposera que $(Y_t) \rightsquigarrow BB(0, \sigma^2)$. Les résultats se généralisent lorsque (Y_t) est un processus linéaire (stationnaire).

Pour que les coefficients soient identifiables (unicité de la décomposition) il est nécessaire d'imposer la contrainte $\sum_{j=1}^l c_j = 0$. On considérera donc plutôt le modèle suivant

$$X_t = \sum_{i=1}^k b_i m_t^{(i)} + \sum_{j=1}^{l-1} c_j \tilde{S}_t^{(j)} + Y_t$$

$$c_l = - \sum_{i=1}^{l-1} c_i$$

où $\tilde{S}_t^{(j)} = S_t^{(j)} - S_t^{(l)}$, pour $j = 1, \dots, l-1$.

La méthode des MCO consiste à choisir comme estimateur de b et c les valeurs qui rendent minimales le carré de la distance entre la série et sa partie déterministe

$$\sum_{t=1}^T \left(X_t - \underbrace{\sum_{i=1}^k b_i m_t^{(i)}}_{\text{deterministic}} - \underbrace{\sum_{j=1}^{l-1} c_j \tilde{S}_t^{(j)}}_{\text{stochastic}} \right)^2.$$

$$\dot{\tilde{m}} = \begin{pmatrix} n^{(0)} \\ \vdots \\ m^{(j)} \\ 1 \end{pmatrix}$$

Appelons :

- $X, m^{(1)}, \dots, m^{(k)}, \tilde{S}^{(1)}, \dots, \tilde{S}^{(l-1)}$ et Y les vecteurs de $\mathbb{R}^T$ dont les composantes sont $X_t, m_t^{(1)}, \dots, m_t^{(k)}, \tilde{S}_t^{(1)}, \dots, \tilde{S}_t^{(l-1)}$ et Y_t pour $t = 1, \dots, T$,
- m et $\tilde{S}$ les matrices dont les colonnes sont $m^{(1)}, \dots, m^{(k)}, \tilde{S}^{(1)}, \dots, \tilde{S}^{(l-1)}$,
- b et $\tilde{c}$ les vecteurs dont les composantes sont les b_i ($i = 1, \dots, k$) et les c_j ($j = 1, \dots, l-1$).

Les estimateurs des MCO $\hat{b}$ et $\hat{c}$ sont donnés par

$$\begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \underbrace{\begin{pmatrix} m'm & m'\tilde{S} \\ \tilde{S}'m & \tilde{S}'\tilde{S} \end{pmatrix}}^{-1} \underbrace{\begin{pmatrix} m'X \\ \tilde{S}'X \end{pmatrix}}.$$

PROPRIETES :

1. Les estimateurs sont sans biais

$$\mathbb{E} \left[\begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} \right] = \begin{pmatrix} b \\ c \end{pmatrix}$$

et de variance

$$\mathbb{V}ar \begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \sigma^2 \begin{pmatrix} m' m & m' \tilde{S} \\ \tilde{S}' m & \tilde{S}' \tilde{S} \end{pmatrix}^{-1}.$$

2. Un estimateur sans biais de σ^2 est donné par

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{Y}_t^2}{T - k - l + 1}$$

où $\hat{Y}_t = X_t - \sum_{i=1}^k \hat{b}_i m_t^{(i)} - \sum_{j=1}^{l-1} \hat{c}_j \tilde{S}_t^{(j)}.$

↑

à considérer.

Les avantages de la méthode de Buys-Ballot généralisée sont :

- il est possible de projeter m_t et $\tilde{S}_t$ au delà de T ;
- les estimateurs sont convergents même si (Y_t) est un processus stationnaire faible (mais les estimateurs ne sont plus optimaux).

Les inconvénients de la méthode de Buys-Ballot généralisée sont :

- la spécification manque de flexibilité (représentation linéaire et stable dans le temps) ;
- si il y a une erreur dans la spécification de m_t et $\tilde{S}_t$, il y aura des erreurs dans les estimations des coefficients et les projections ne seront pas pertinentes.

(propagation de l'erreur par les prédictions)

▷ La méthode STL (Seasonal decomposition of Time series by Loess)

Au lieu d'effectuer une régression linéaire sur l'ensemble des données, on effectue plusieurs regressions au voisinage des points pour lesquels on souhaite avoir des valeurs des composantes.

On utilise une fenêtre qui peut être ajustée par l'utilisateur ou de manière automatique. Les observations ont d'autant plus de poids qu'elles sont proches du point où l'on souhaite faire l'estimation. On effectue ensuite une régression linéaire généralisée avec prise en compte des poids.

Plus la fenêtre est grande, plus le nombre de données est grand, donc plus l'estimateur aura une petite variance mais un biais qui pourra être élevé, et réciproquement.

○ Sous

Pour effectuer un régression linéaire, on utilise la fonction

```
lm(formula, data, subset, weights, na.action,  
    method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,  
    singular.ok = TRUE, contrasts = NULL, offset, ...)
```

Pour effectuer une décomposition suivant la méthode STL, on utilise la fonction

```
stl(x, s.window, s.degree = 0, t.window = NULL, t.degree = 1,  
    l.window = nextodd(period), l.degree = t.degree,  
    s.jump = ceiling(s.window/10), t.jump = ceiling(t.window/10),  
    l.jump = ceiling(l.window/10), robust = FALSE, ...)
```

2.2 Méthodes par moyenne mobiles

Ces méthodes consistent à calculer des moyennes de quelques données de la série de départ autour de chaque date.

En choisissant bien la transformation, on peut :

- annuler/conservé une tendance constante, linéaire, polynômiale, géométrique,
- annuler/conservé les composantes saisonnières.

Elles doivent également annuler ou tout du moins réduire fortement le terme d'erreur.

▷ Définition des moyennes mobiles (MM) :

La moyenne mobile d'ordre $m_1 + m_2 + 1$ est définie par

$$MX_t = \theta_{-m_1}X_{t-m_1} + \theta_{-m_1+1}X_{t-m_1+1} + \dots + \theta_{m_2}X_{t+m_2}, \quad t \in \mathbb{Z},$$

où les m_j sont des entiers positifs et les θ_j des réels (dans la pratique, on ne calcule MX_t que pour $m_1 \leq t \leq T - m_2$, où T est le nombre d'observations disponibles).

PROPRIETES :

1. L'opérateur M est linéaire : pour deux nombres a, b et deux séries (X_t) et (Y_t)

$$M(aX_t + bY_t) = aMX_t + bMY_t.$$

2. La composée de deux moyennes mobiles est une moyenne mobile : la composée de M_1 et M_2 est la moyenne mobile M qui à X_t fait correspondre

$$MX_t = M_1(M_2X_t) = M_2(M_1X_t). \quad \leftarrow$$

Cette opération est commutative, associative.

▷ MM symétriques

Lorsque $m_1 = m_2 = m$, on parle de moyenne mobile centrée. Si de plus :

$$\theta_i = \theta_{-i}, \text{ pour } i = 1, \dots, m,$$

la moyenne mobile est dite symétrique.

On utilisera la notation symbolique pour une moyenne mobile symétrique :

$$M = \{[2m + 1]; [\theta_{-m}, \dots, \theta_0]\}$$

Si M_1 et M_2 sont centrées (resp. symétriques), $M_1 M_2$ est centrée (resp. symétrique).

En pratique, on construit des moyennes mobiles d'ordre élevé à partir de moyennes mobiles d'ordres inférieurs (donc plus simples).

▷ Série invariante par une moyenne mobile et noyau d'une moyenne mobile

DEFINITIONS :

1. On dit que (X_t) est invariante par M si $MX_t = X_t$, pour tout t .
2. Le noyau de M , $\text{Ker}M$, est l'ensemble des séries temporelles annulées par cette moyenne

$$\text{Ker}M = \{X_t : MX_t = 0\}.$$

▷ Résolution d'une équation de récurrence linéaire

Une équation de récurrence linéaire à coefficients constants (sans second membre) d'ordre p est une équation de la forme

$$\alpha_0 z_n + \alpha_1 z_{n+1} + \dots + \alpha_p z_{n+p} = 0$$

où $\alpha_0 \neq 0$ et $\alpha_p \neq 0$.

Soit P le polynôme associé

$$P(z) = \alpha_0 + \alpha_1 z + \dots + \alpha_p z^p.$$

Il admet p racines réelles ou complexes (conjuguées). On note μ_i les l racines et r_i leur multiplicité respective.

La solution générale de l'équation de récurrence est

$$z_n = \sum_{i=1}^l \left(\sum_{s=0}^{r_i-1} a_{is} n^s \right) \mu_i^n$$

où les p constantes a_{is} , $0 \leq s \leq r_i - 1$ et $1 \leq i \leq l$, sont déterminées par les valeurs initiales $z_0, \dots, z_{p-1}$.

Les suites $t^s \mu_i^t$, $0 \leq s \leq r_i - 1$ et $1 \leq i \leq l$, sont linéairement indépendantes et forment une base de l'espace des solutions de l'équation.

PROPOSITIONS :

- 1. M conserve les constantes ssi $\sum_{j=-m_1}^{m_2} \theta_j = 1$.
- 2. Une moyenne mobile symétrique conservant les constantes conserve les polynômes de degré 1.
- 3. Soit

$$P(x) = \theta_{-m_1} + \theta_{-m_1+1}x + \dots + \theta_{m_2}x^{m_1+m_2}.$$

Si 1 est racine de $P(x) - x^{m_1}$ de multiplicité $d+1$, alors M conserve les polynômes de degré inférieur ou égal à d .

Exo

✓ Vérifier 1/ et 2/

▷ Effet d'une moyenne mobile sur un bruit blanc faible

Soient (ε_t) un bruit blanc faible de variance σ_ε^2 et $M\varepsilon_t = \sum_{j=-m_1}^{m_2} \theta_j \varepsilon_{t+j}$. On a

$$\mathbb{E}[M\varepsilon_t] = 0, \text{ Var}(M\varepsilon_t) = \sigma_\varepsilon^2 \sum_{j=-m_1}^{m_2} \theta_j^2 =: \sigma_{M\varepsilon}^2,$$

$$\text{Cov}(M\varepsilon_t, M\varepsilon_{t'}) \neq 0, \text{ pour } t \neq t'.$$

La transformée d'un bruit blanc faible n'est pas un bruit blanc faible. Le rapport de réduction de la variance du bruit

$$\frac{\sigma_{M\varepsilon}^2}{\sigma_\varepsilon^2} = \sum_{j=-m_1}^{m_2} \theta_j^2$$

doit être choisi le plus petit possible tout en conservant certaines caractéristiques de la série initiale.

L'utilisation d'une moyenne mobile sur un bruit blanc conduit à l'effet de Slutsky-Yule, c'est-à-dire des oscillations régulières pouvant passer pour de la saisonnalité. L'effet est négligeable si le rapport de réduction est petit.

▷ Moyennes mobiles arithmétiques

PROPOSITION :

La MM symétrique qui conserve les constantes et a le plus grand pouvoir de réduction de la variance du bruit :

$$\min \sum_{j=-m}^m \theta_j^2 \text{ sous la contrainte que } \sum_{j=-m}^m \theta_j = 1$$

est la série arithmétique M définie par

$$\theta_j = \frac{1}{2m+1} \text{ pour tout } j = 1, \dots, m.$$

On utilisera la notation symbolique $M = \{[2m+1]\}$.

moyenne mobile arithmétique :
symétrique, conserve les constantes
et réduit optimalement la variance du bruit

PROPRIETES DES MM ARITHMETIQUES :

1. M conserve les polynômes de degré 1 mais pas ceux de degré 2.
2. M annule les saisonnalités de période $2m + 1$: si $S_t = S_{t+2m+1}$ pour tout t , alors on a

$$MS_t = \frac{1}{2m+1}(S_{-m} + S_{-m+1} + \dots + S_m)$$

est indépendant de t .

Par exemple : une saisonnalité due au jour de la semaine correspond à $m = 3$. Elle disparaît par application de la MM

$$MX_t = \frac{1}{7}(X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3}).$$

▷ Comment annuler une saisonnalité d'ordre pair? *tout en restant symétrique*

Les moyennes mobiles M_1 et M_2 définies par

$$M_1 X_t = \frac{1}{2m} (X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1})$$

$$M_2 X_t = \frac{1}{2m} (X_{t-m+1} + \dots + X_{t+m-1} + X_{t+m})$$

annulent les saisonnalités d'ordre $2m$.

Pour obtenir une moyenne mobile symétrique, on prend

$$MX_t = \frac{1}{2} (M_1 X_t + M_2 X_t) = \frac{1}{2m} (0,5X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1} + 0,5X_{t+m})$$

qui annule les saisonnalités d'ordre $2m$.

▷ Lissage d'une série par MM

Lorsque les MM sont utilisées pour estimer la tendance déterministe, il peut être intéressant de réduire les aspérités (pics et creux) de la tendance.

On montre que MX_t est d'autant plus lisse que la distribution des θ_i l'est. On examine donc l'évolution des θ_i en fonction de i . On utilise parfois le critère suivant (à minimiser) :

$$Q = \sum_i (\Delta^3 \theta_i)^2 = \sum_i (\theta_i - 3\theta_{i-1} + 3\theta_{i-2} - \theta_{i-3})^2$$

qui est nul si les coefficients se trouvent sur une parabole.

EXEMPLE : Les moyennes d'Henderson sont obtenues en minimisant Q parmi toutes les moyennes mobiles d'ordre $2m + 1$ conservant les polynômes de degré 2.

Pour les séries trimestrielles (période 4) on utilise la moyenne d'Henderson à 5 termes :

$$M = \left\{ [5]; \frac{1}{286} [-21, 84, 160] \right\}$$

$\uparrow$
 $2m+1$

$\begin{matrix} 3^0 & 2^1 & 3^2 \\ \text{ } & \text{ } & \text{ } \end{matrix}$

$\underbrace{\hspace{10em}}$

▷ Les propriétés souhaitables des moyennes mobiles pour estimer une tendance déterministe sont :

1. conservation d'une tendance : linéaire, polynômiale, ...
2. annulation d'une saisonnalité,
3. réduction de la variance du bruit,
4. lissage de la série,
5. simplicité des coefficients, ordre le plus petit possible.

EXEMPLE :

La moyenne de Spencer sur 15 points

$$M = \{[15]; \frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74]\}$$

- conserve les polynômes de degré inférieur à 3,
- annule les saisonnalités de périodes 4 et celles de période 5 (saisonnalités dont la période n'est pas rigoureusement 4),
- a un pouvoir de réduction de 0,19,
- a un bon pouvoir de lissage.

▷ Principe du filtrage par moyenne mobile

Supposons que la série se décompose additivement en

$$X_t = m_t + S_t + Y_t, \quad t = 1, \dots$$

où m_t est une tendance, S_t un facteur saisonnier et Y_t une série stationnaire.

Si M_0 est une MM conservant la tendance, annulant la saisonnalité et telle que la série $M_0 Y_t$ est de moyenne proche de 0, alors on pose

$$\hat{m}_t = M_0 X_t. \quad \text{estimation non-paramétrique de } m_t$$

On peut alors retrancher à X_t la tendance estimée

$$X_t - \underbrace{M_0 X_t}_{\hat{m}_t} = (Id - M_0) X_t \simeq \underbrace{S_t + Y_t}.$$

On applique ensuite une moyenne mobile M_1 pour obtenir les coefficients saisonniers, puis à nouveau $(Id - M_0)$ pour s'assurer que la somme des coefficients saisonniers est approximativement nulle.

D'où l'estimateur

$$\hat{S}_t = (Id - M_0)M_1 \underbrace{(Id - M_0)X_t}_{\approx S_t + \gamma_t}$$

estimateur non
paramétrique
de γ

La série corrigée des variations saisonnières (CVS) est définie par

$$X_t - \hat{S}_t.$$

EXEMPLE : Méthode de désaisonnalisation Census X11 (1965) pour une série trimestrielle (version simplifiée)

- Première estimation de la tendance : calcul d'une approximation de la tendance par application d'une MM arithmétique annulant les saisonnalités d'ordre 4 :

$$M_0 = \left\{ [5]; \frac{1}{8}[1, 2, 2] \right\}, \hat{m}_t = M_0 X_t. \quad \checkmark$$

- Première estimation de la série diminuée de sa tendance :

$$X_t - \hat{m}_t = (Id - M_0)X_t.$$

- Première estimation de la saisonnalité : $(Id - M_0)X_t$ est mis en MM sur 5 ans par

$$M_1 = \{ [17]; \frac{1}{9}[1, 0, 0, 0, 2, 0, 0, 0, 3] \}. \quad ,$$

- On applique à nouveau $(Id - M_0)$ pour annuler la somme des coefficients saisonniers :

$$\hat{S}_t = (Id - M_0)M_1(I - M_0)X_t. \quad \checkmark$$

- Première estimation de la série CVS :

$$\hat{X}_t^{CVS} = X_t - \hat{S}_t.$$

- Seconde estimation de la tendance : on lisse la série $\hat{X}_t^{CVS}$. On peut par exemple lui appliquer une MM de Henderson à 5 termes :

$$M_2 = \left\{ [5]; \frac{1}{286}[-21, 84, 160] \right\}, \hat{m}_t^{(2)} = M_2 \hat{X}_t^{CVS}. \quad \checkmark$$

- Seconde estimation de la série diminuée de sa tendance :

$$X_t - \hat{m}_t^{(2)}.$$

- Seconde estimation de la saisonnalité : $X_t - \hat{m}_t^{(2)}$ est mis en MM sur 7 ans par

$$M_3 = \left\{ [25]; \frac{1}{15}[1, 0, 0, 0, 2, 0, 0, 0, 3, 0, 0, 0, 3] \right\}.$$

- On applique à nouveau $(Id - M_0)$ pour annuler la somme des coefficients saisonniers :

$$\hat{S}_t^{(2)} = (Id - M_0)M_3(X_t - \hat{m}_t^{(2)}). \quad \checkmark$$

- Estimation finale de la série CVS :

$$\hat{X}_t^{CVS(2)} = X_t - \hat{S}_t^{(2)}.$$

On remarquera que $\hat{X}_t^{CVS(2)} = MX_t$ avec

$$M = Id - (Id - M_0)M_3\{Id - M_2(Id - M_1(I - M_0)^2)\}.$$

M comporte 57 coefficients et est symétrique : elle porte sur 28 trimestres de part et d'autre du trimestre courant.

Les premiers coefficients sont

$$\begin{aligned}\theta_0 &= 0,856, \theta_1 = 0,051, \theta_2 = 0,041, \theta_3 = 0,050, \theta_4 = -0,140, \\ \theta_5 &= 0,055, \theta_6 = 0,034, \theta_7 = 0,029, \theta_8 = -0,097, \theta_9 = 0,038.\end{aligned}$$

- Sous

Pour appliquer une moyenne mobile à une série, on utilise la fonction

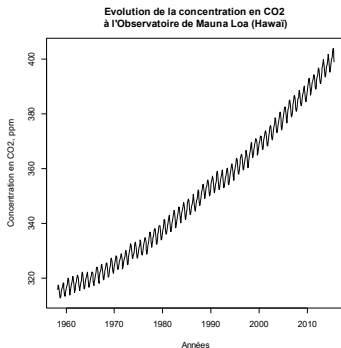
```
filter(x, filter, method = c("convolution", "recursive"),  
      sides = 2, circular = FALSE, init)
```

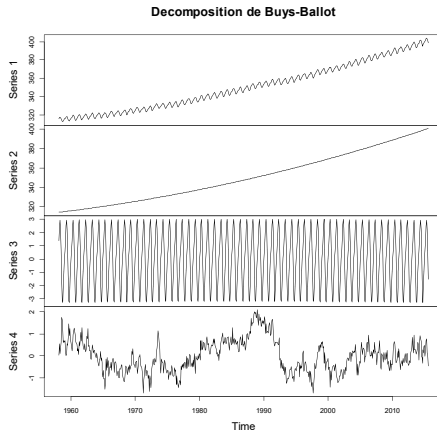
Pour effectuer une décomposition à l'aide de MM de manière automatique, on utilise la fonction

```
decompose(x, type = c("additive", "multiplicative"),  
          filter = NULL)
```

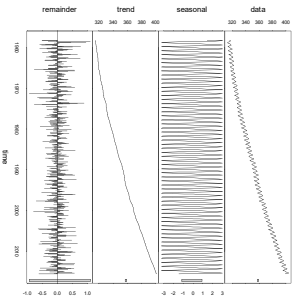
2.3 Exemples

On considère à nouveau la série des concentrations en CO_2

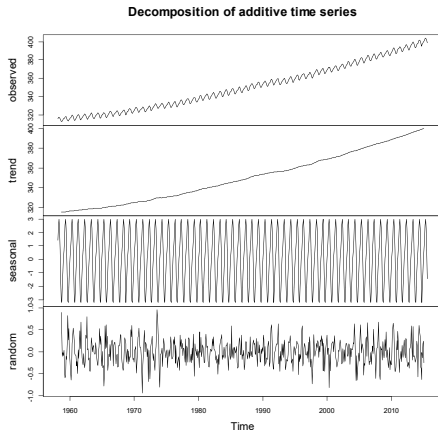




Décomposition de Buys-Ballot généralisée avec tendance quadratique

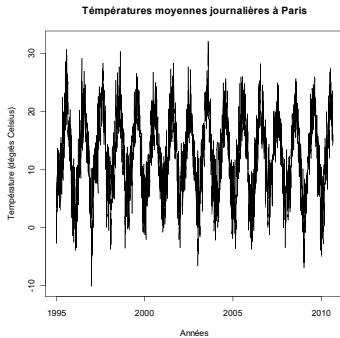


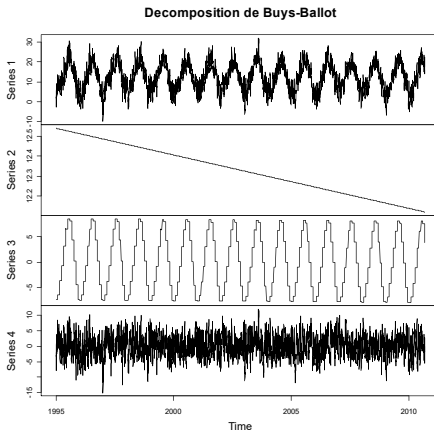
Décomposition par la méthode STL



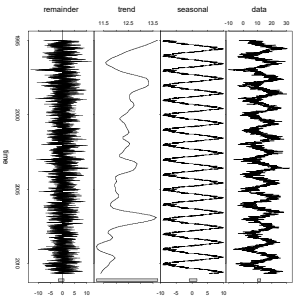
Décomposition par moyenne mobile

On considère à nouveau la série des températures à Paris



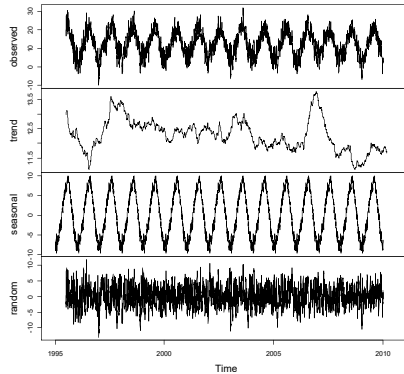


Décomposition de Buys-Ballot généralisée avec tendance linéaire



Décomposition par la méthode STL

Decomposition of additive time series



Décomposition par moyenne mobile

3 Processus linéaires

3.1 L'espace $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$: définition et propriétés

DEFINITION :

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

$$\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P}) = \{X | X : \Omega \rightarrow \mathbb{R}, X \text{ est } \mathcal{A}\text{-mesurable et } \mathbb{E}_{\mathbb{P}}[X^2] < \infty\}.$$

PROPRIETES :

1. L'ensemble des variables aléatoires réelles X de carré intégrable, $\mathbb{E}[X^2] < \infty$, est un espace vectoriel normé sur $\mathbb{R}$.
2. La norme est définie par

$$X \mapsto \|X\|_2^2 = \mathbb{E}[X^2]$$

et le produit scalaire par

$$(X, Y) \mapsto \mathbb{E}[XY].$$

3. L'espace $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ muni du produit scalaire, $(X, Y) \mapsto \mathbb{E}[XY]$, est un espace de Hilbert, c'est-à-dire un espace vectoriel muni d'un produit scalaire et complet pour la norme associée.

Un des intérêts des espaces de Hilbert est qu'ils autorisent la “construction d'approximations” de tous leurs éléments par des éléments qui appartiennent à un sous-ensemble bien défini a priori.

Cette approximation repose sur la notion de projection orthogonale.

DEFINITIONS :

1. (Orthogonalité) Deux variables X et Y de carré intégrables sont dites orthogonales au sens de $\mathcal{L}^2$ si $\mathbb{E}[XY] = 0$.
2. (Convergence) Une suite X_n de variables, vérifiant $\mathbb{E}[X_n^2] < \infty$, converge vers une variable X au sens de $\mathcal{L}^2$ si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_2 = 0$$

c'est-à-dire si

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0.$$

On en déduit en particulier que

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

REMARQUE :

La variable aléatoire $X = \sum_{k=0}^{\infty} X_k$ peut alors être définie comme la limite suivante dans $\mathcal{L}^2$:

$$X = \lim_{n \rightarrow \infty} \sum_{k=0}^n X_k$$

avec $\lim_{n \rightarrow \infty} \left\| \sum_{k=0}^n X_k - X \right\|_2 = 0$.

▷ Définition d'une combinaison linéaire infinie au sens de $\mathcal{L}^2$

Soit la variable aléatoire définie par $Y = \sum_{j \in \mathbb{Z}} a_j X_j$ où les X_j sont indépendants les uns des autres.

Y a un sens dans $\mathcal{L}^2$ si et seulement si :

$$\sum_{j \in \mathbb{Z}} \|a_j X_j\|_2 = \sum_{j \in \mathbb{Z}} |a_j| \|X_j\|_2 < \infty.$$

Dans le cas où (X_t) est un processus stationnaire, on obtient la condition :

$$\sum_{j \in \mathbb{Z}} |a_j| < \infty.$$

puisque :

$$\sum_{j \in \mathbb{Z}} \|a_j X_j\|_2 = \sqrt{\mathbb{E}[X_1^2]} \sum_{j \in \mathbb{Z}} |a_j|.$$

▷ Transformée d'un processus stationnaire par une moyenne mobile infinie

PROPOSITION :

Si $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire et $(a_i)_{i \in \mathbb{Z}}$ est une suite de réels absolument sommable, $\sum_{i \in \mathbb{Z}} |a_i| < \infty$, alors $Y_t = \sum_{i \in \mathbb{Z}} a_i X_{t-i}$ définit un nouveau processus stationnaire.

En particulier, $Y_t \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ et

$$\begin{aligned}\mathbb{E}[Y_t] &= m_Y = m_X \sum_{i \in \mathbb{Z}} a_i \\ \gamma_Y(h) &= \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} a_i a_j \gamma_X(h + j - i).\end{aligned}$$

(à vérifier en exo)

On dit que $(Y_t)_{t \in \mathbb{Z}}$ est la transformée de $(X_t)_{t \in \mathbb{Z}}$ par la moyenne mobile infinie associée aux $(a_i)_{i \in \mathbb{Z}}$.

3.2 Régression linéaire ou affine théorique sur un nombre fini de retards

DEFINITIONS :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

1. La régression linéaire théorique de X_t sur $X_{t-1}, \dots, X_{t-p}$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur $H = Vect(X_{t-1}, \dots, X_{t-p})$. On note :

$$X_t^\times = EL(X_t | X_{t-1}, \dots, X_{t-p}) = P_H X_t$$

où $EL(.|.)$ désigne la régression linéaire.

2. La régression affine théorique de X_t sur $1, X_{t-1}, \dots, X_{t-p}$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur $H^* = Vect(1, X_{t-1}, \dots, X_{t-p})$. On note :

$$X_t^* = EL(X_t | 1, X_{t-1}, \dots, X_{t-p}) = P_{H^*} X_t$$

REMARQUES :

1. Les deux définitions coïncident si $\mathbb{E}[X_t] = 0$.
2. Si $\mathbb{E}[X_t] \neq 0$, on calculera toujours la régression affine théorique notée (par abus de notations) $EL(X_t|X_{t-1}, \dots, X_{t-p})$.
3. $Vect(X_{t-1}, \dots, X_{t-p})$ est un sous-espace vectoriel de dimension finie de $\mathcal{L}^2$ et est donc fermé. L'existence et l'unicité de la projection orthogonale sont donc assurées.
4. Si X_t est un processus Gaussien, alors l'espérance linéaire correspond à l'espérance conditionnelle.

PROPOSITION :

Les coefficients de la régression linéaire affine, $a_0, a_1, \dots, a_p$,

$$P_{H^*} X_t = \underbrace{a_0}_{\text{}} + \sum_{j=1}^p a_j X_{t-j}$$

sont définis par :

$$a_0 = m_X \left(1 - \sum_{j=1}^p a_j \right) \text{ et } \underbrace{\Omega \underline{a} = \underline{\gamma}}$$

où

$$\underline{a} = (a_1, \dots, a_p)^T$$

$$\Omega = (\gamma_X(i-j))_{1 \leq i \leq p, 1 \leq j \leq p}$$

$$\underline{\gamma} = (\gamma_X(1), \dots, \gamma_X(p))^T.$$

DEFINITION :

Le coefficient d'autocorrélation partielle d'ordre k est défini par :

$$r(k) = \text{Cor}(\tilde{X}_t, \tilde{X}_{t-k})$$

où

$$\begin{aligned}\tilde{X}_t &= X_t - EL(X_t | X_{t-1}, \dots, X_{t-k+1}) \\ \tilde{X}_{t-k} &= X_{t-k} - EL(X_{t-k} | X_{t-1}, \dots, X_{t-k+1}).\end{aligned}$$

L'autocorrélation partielle d'ordre k désigne la corrélation entre X_t et X_{t-k} lorsque l'influence des variables X_{t-k+i} avec $i < k$ a été prise en compte.

REMARQUES :

1. Si $\mathbb{E}[X_t] = 0$, on peut montrer que $r(k) = a_k$ où a_k est le coefficient de X_{t-k} dans la régression linéaire $EL(X_t|X_{t-1}, \dots, X_{t-k})$

$$EL(X_t|X_{t-1}, \dots, X_{t-k}) = \sum_{j=1}^k a_j X_{t-j}$$

DEFINITION :

L'autocorrélogramme partiel de $(X_t)_{t \in \mathbb{Z}}$ (processus centré) est la fonction : $k \mapsto r(k)$.

PROPOSITION :

La fonction d'autocorrelation partielle d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ satisfait la relation :

$$r(k) = \frac{\det(\mathcal{R}^*(k))}{\det(\mathcal{R}(k))}$$

où $\mathcal{R}(k)$ est la matrice d'autocorrelation (d'ordre k) définie par $\mathcal{R}(k) = (\rho_X(i - j))_{1 \leq i \leq k, 1 \leq j \leq k}$ et $\mathcal{R}^*(k)$ est la matrice obtenue en remplaçant la dernière colonne de $\mathcal{R}(k)$ par le vecteur des autocorrélations $\underline{\rho} = (\rho_X(1), \dots, \rho_X(k))^T$.

En particulier, en notant $\rho_i = \rho_X(i)$,

$$\begin{aligned} r(1) &= \rho_1, & r(2) &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ r(3) &= \frac{\rho_1^2 - \rho_1 \rho_2 (2 - \rho_2) + \rho_3 (1 - \rho_1^2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}. \end{aligned}$$

3.3 Régression linéaire théorique sur un nombre infini de retards

DEFINITION :

1. La régression linéaire théorique de X_t sur $X_{t-1}, \dots, X_{t-p}, \dots$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur l'adhérence de l'espace engendré par les combinaisons linéaires des X_{t-j} ($j > 0$), $H = Vect(X_{t-1}, X_{t-2}, \dots, X_{t-p}, \dots)$.

On note :

$$X_t^\times = EL(X_t | \underline{X_{t-1}}) = EL(X_t | X_{t-1}, X_{t-2}, \dots, \underline{X_{t-p}, \dots}).$$

2. La régression affine théorique de X_t sur $X_{t-1}, \dots, X_{t-p}, \dots$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur l'adhérence de l'espace $H^* = Vect(1, X_{t-1}, \dots)$.

On note :

$$X_t^* = EL(X_t | \underline{X_{t-1}}) = EL(X_t | 1, X_{t-1}, X_{t-2}, \dots, \underline{X_{t-p}, \dots}).$$

PROPOSITIONS :

1. La régression linéaire théorique sur un nombre infini de retards peut se définir comme suit :

$$EL(X_t|\underline{X_{t-1}})) =_{L^2} \lim_{n \rightarrow \infty} EL(X_t|X_{t-1}, \dots, X_{t-n}).$$

En d'autres termes, une prédiction basée sur un passé limite du processus approxime celle qui résulte de l'usage du passé infini.

2. Soient $(X_t)_{t \in \mathbb{Z}}$ une série stationnaire et $X_t^* = EL(X_t|1, \underline{X_{t-1}}))$. La série $(\varepsilon_t)_{t \in \mathbb{Z}}$ définie par le terme :

$$\varepsilon_t = X_t - X_t^*$$

est un bruit blanc. Cette série est appelée processus des innovations.

Le processus des innovations contient l'information nouvelle présente à chaque date de $(X_t)_{t \in \mathbb{Z}}$ et imprévisible à partir du passé de X_t .

Il est immédiat de montrer que : $\varepsilon_t = x_t - x_t^*$

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

$$\mathbb{E}[\varepsilon_t X_{t-k}] = \text{Cov}(\varepsilon_t, X_{t-k}) = 0 \quad \forall k > 0. \quad (\text{innovation})$$

THEOREME DE WOLD :



$$\varepsilon_t \sim X_t - X_t^*$$

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique stationnaire réel et $(\varepsilon_t)_{t \in \mathbb{Z}}$ le processus des innovations correspondant. Il existe $(a_k)_{k \in \mathbb{N}}$ tel que $\sum_{k=0}^{\infty} |a_k| < \infty$ et

$$X_t = m_X + \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}. \quad (\text{causal})$$

Chaque variable aléatoire X_t d'un processus stationnaire peut s'écrire comme la somme de son espérance et d'une moyenne mobile infinie du processus d'innovation associé.

Il s'agit d'une représentation possible d'un processus stochastique (linéaire) qui exploite la stationnarité à l'ordre 2.

3.4 Densité spectrale et autocorrélations inverses

PROPOSITIONS :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire (moyenne mobile infinie)

$$X_t = m + \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

où $(\varepsilon)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ_ε^2 et où $(a_j)_{j \in \mathbb{Z}}$ est une suite absolument sommable, $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.

La fonction d'autocovariance de $(X_t)_{t \in \mathbb{Z}}$ définie par

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} a_j a_{j-h} \sigma_\varepsilon^2 \quad < c \|a\|_1^2 < +\infty$$

$$\begin{aligned} \gamma_X(h) &= \text{Cov}(X_t, X_{t+h}) = \mathbb{E} \left[\left(\sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j} \right) \left(\sum_{k \in \mathbb{Z}} a_k \varepsilon_{t+h-k} \right) \right] \\ &= \sum_{j, k \in \mathbb{Z}} a_j a_k \gamma_\varepsilon(h+j-k) = \sigma_\varepsilon^2 \sum_{j, k \in \mathbb{Z}} a_j a_k \mathbb{1}_{k=h+j} = \sigma_\varepsilon^2 \sum_{j \in \mathbb{Z}} a_j a_{h+j} \end{aligned}$$

satisfait l'inégalité suivante :

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty.$$

DEFINITION :

On appelle densité spectrale du processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ la fonction définie sur $\mathbb{R}$ par :

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{i\omega h}, \quad \omega \in [-\pi, \pi].$$

La densité spectrale est donc la transformée de Fourier de la suite des autocovariances.

Elle existe car la série de terme général $\gamma_X(h) e^{i\omega h}$ est absolument convergente.

La densité spectrale est une fonction réelle, paire, continue, périodique de période 2π . En particulier :

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) \cos(\omega h).$$

THEOREME D'INJECTIVITE :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique stationnaire. Il est équivalent de connaître la fonction d'autocovariance ou la densité spectrale. On a :

$$\gamma_X(h) = \int_{-\pi}^{\pi} f_X(\omega) e^{-i\omega h} d\omega = \int_{-\pi}^{\pi} f_X(\omega) \cos(\omega h) d\omega.$$

EXEMPLES :

1. Si $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ_ε^2 , alors

$$f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}, \quad \omega \in [-\pi, \pi]. \quad (\text{E}_\text{KO})$$

2. Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus stationnaire. Le processus $(X_t)_{t \in \mathbb{Z}}$ défini par $X_t = m + \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}$ (avec $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$) admet pour fonction de densité spectrale :

où

$$\left[\begin{array}{l} f_X(\omega) = |\Psi(\exp(i\omega))|^2 f_Y(\omega), \quad \omega \in [-\pi, \pi], \\ \Psi(\exp(i\omega)) = \sum_{j=-\infty}^{\infty} \psi_j \exp(ij\omega), \quad \omega \in [-\pi, \pi]. \end{array} \right]$$

3. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus moyenne mobile d'ordre 1 :

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad (\varepsilon_t) \sim BB(0, \sigma_\varepsilon^2).$$

$$\gamma_X(h) = \dots$$

Sa densité spectrale est définie par :

(Exo: ind utilisation
me relation trig)

$$f_X(\omega) = |1 + \theta \exp(-i\omega)|^2 \frac{\sigma_\varepsilon^2}{2\pi} = (1 + 2\theta \cos(\omega) + \theta^2) \frac{\sigma_\varepsilon^2}{2\pi}.$$

4. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus autorégressif d'ordre 1 :

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad (\varepsilon_t) \rightsquigarrow BB(0, \sigma^2), \quad \underline{|\varphi| < 1}.$$

Sa densité spectrale est définie par :

$$f_X(\omega) = |1 - \varphi \exp(-i\omega)|^{-2} \frac{\sigma_\varepsilon^2}{2\pi} = \frac{1}{(1 - 2\varphi \cos(\omega) + \varphi^2)} \frac{\sigma_\varepsilon^2}{2\pi}.$$

DEFINITION : Autocorrélations inverses

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus moyenne mobile infinie, $X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$ où $(\varepsilon)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ_ε^2 et où $(a_j)_{j \in \mathbb{Z}}$ est une suite absolument sommable, $\sum_{j=-\infty}^{\infty} |a_j| < \infty$. Supposons que la fonction :

$$\omega \mapsto \frac{1}{f_X(\omega)} \exp(-i\omega h)$$

soit intégrable sur $[-\pi, \pi]$.

La fonction d'autocovariance inverse (associée au spectre inverse $1/f_X$) est définie par :

inverse ↓

$$\gamma_X^i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} \exp(-i\omega h) d\omega$$

ou

$$\frac{1}{f_X(\omega)} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X^i(h) \exp(i\omega h).$$

3.5 Estimation de la moyenne et des autocovariances pour des processus linéaires

On suppose disposer d'observations $(x_t)_{t=1,\dots,T}$ d'un processus

$$X_t = m_X + \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort de variance σ_ε^2 et où $(a_j)_{j \in \mathbb{Z}}$ est une suite absolument sommable, $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.

▷ Estimateur de la moyenne et loi limite

L'estimateur de la moyenne est :

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t.$$

PROPOSITIONS :

1.

$$T\mathbb{V}ar(\overline{X}_T) = \sum_{|h| < T-1} \left(1 - \frac{|h|}{T}\right) \gamma_X(h) \xrightarrow{n \rightarrow \infty} \underbrace{\sum_{h=-\infty}^{\infty} \gamma_X(h)}_{< +\infty}.$$

2. Si $\sum_{j=-\infty}^{\infty} a_j \neq 0$, alors

$$\sqrt{T}(\overline{X}_T - m_X) \xrightarrow{L} \mathcal{N}\left(0, \underbrace{\sum_{h=-\infty}^{\infty} \gamma_X(h)}_{> 0}\right)$$

avec $\sum_{h=-\infty}^{\infty} \gamma_X(h) = \sigma_\varepsilon^2 \left(\sum_{j=-\infty}^{\infty} a_j\right)^2$.

▷ Estimateur des autocovariances et lois limites

Un estimateur de l'autocovariance d'ordre h est (pour $h < T$) :

$$\hat{\gamma}_X(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (\underbrace{x_t - \bar{x}_T})(\underbrace{x_{t+h} - \bar{x}_T}). \quad \leftarrow$$

PROPOSITIONS :

1. $\hat{\gamma}_X(h)$ est un estimateur presque sûrement convergent, asymptotiquement sans biais.
2. Si de plus (ε_t) est Gaussien et $\sum_{k \in \mathbb{Z}} |k \gamma_X(k)| < \infty$,

$$\sqrt{T} \left((\hat{\gamma}_X(i) - \gamma_X(i))_{i=1, \dots, h} \right) \xrightarrow{L} \mathcal{N}(0, \Omega)$$

où

$$\Omega_{j,k} = \sum_{l=-\infty}^{\infty} \gamma_X(l) (\gamma_X(l-j+k) + \gamma_X(l-j-k)).$$

A partir d'un estimateur des autocovariances, on en déduit

- un estimateur des autocorrélations :

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

- un estimateur 'corrigé' de la densité spectrale :

$$\hat{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-H}^H \left(1 - \frac{|h|}{H+1}\right) \hat{\gamma}_X(h) e^{i\omega h}.$$

où H est suffisamment grand mais pas trop grand !

Un estimateur du coefficient d'autocorrélation partielle d'ordre h est donné par l'estimateur des moindres carrés ordinaires de X_t sur $1, X_{t-1}, \dots, X_{t-h}$.

- Sous

Pour tracer des fonctions d'autocovariance, d'autocovariance partielle et d'autocorrélation empiriques, il faut utiliser la fonction `acf` (après avoir chargé le package `stats`).

```
acf(x, lag.max = NULL,  
    type = c("correlation", "covariance", "partial"),  
    plot = TRUE, na.action = na.fail, demean = TRUE, ...)
```

La fonction renvoie un objet de type "acf" qui est une liste de plusieurs éléments (dont les valeurs calculées pour les différentes valeurs de h).

EXEMPLES AVEC DES PROCESSUS STATIONNAIRES : $T = 250$

BRUIT BLANC :

1. Si (X_t) est un bruit blanc faible, sa fonction d'autocovariance et d'autocovariance partielle sont données par

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}, \quad \underline{r(k) = 0 \quad k \geq 1.}$$

2. Si (X_t) est une moyenne mobile d'ordre 1 (MA(1)) telle que

$$\underline{X_t = m + \varepsilon_t - \theta \varepsilon_{t-1}}, \quad (\varepsilon_t) \rightsquigarrow BB(0, \sigma^2), \quad |\theta| < 1$$

sa fonction d'autocovariance et d'autocovariance partielle sont données par

$$\underline{\rho_X(h) = \begin{cases} 1 & h = 0 \\ -\theta/(1 + \theta^2) & h = \pm 1 \\ 0 & |h| > 1 \end{cases}}, \quad \underline{r_X(k) = \frac{\theta^k(\theta^2 - 1)}{1 - \theta^{2k+2}} \quad k \geq 1.}$$

3. Si (X_t) est un autoregressif d'ordre 1 (AR(1)) tel que

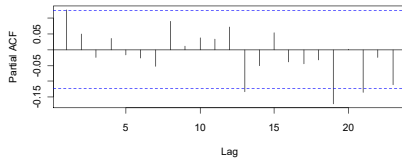
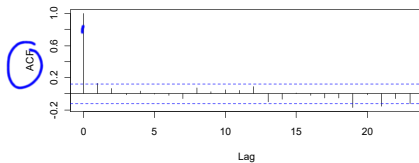
$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad (\varepsilon_t) \rightsquigarrow BB(0, \sigma^2), \quad |\varphi| < 1,$$

et les ε_t est non-corrélé avec les X_{t-i} pour $i > 0$, sa fonction d'autocovariance et d'autocovariance partielle sont données par

$$\rho_X(h) = \varphi^{|h|}, \quad \underline{r_X(1) = \varphi} \text{ et } r_X(k) = 0 \text{ pour } k > 1.$$

Bruit blanc fort

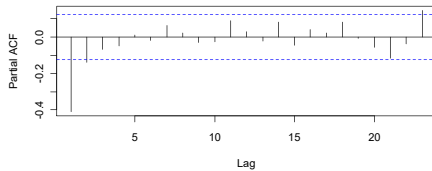
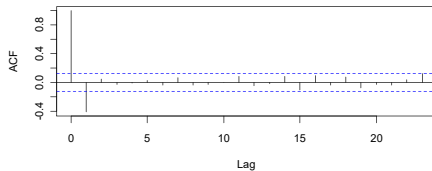
stimulus



$$\varepsilon_t \sim \mathcal{N}(0, 1)$$

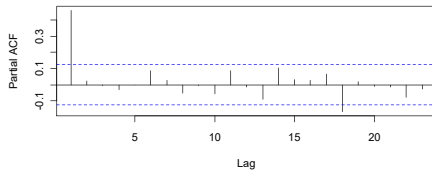
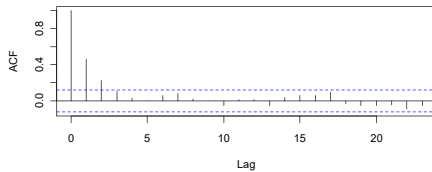


Moyenne Mobile d'ordre 1



$$\underbrace{m = 0}, \underbrace{\theta = 0,5}, \varepsilon_t \sim \mathcal{N}(0,1)$$

Autoregressif d'ordre 1



$$\varphi = 0,5, \varepsilon_t \sim \mathcal{N}(0,1)$$

EXEMPLES AVEC DES PROCESSUS NON-STATIONNAIRES : $T = 250$

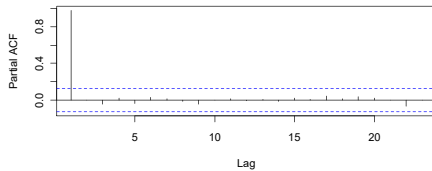
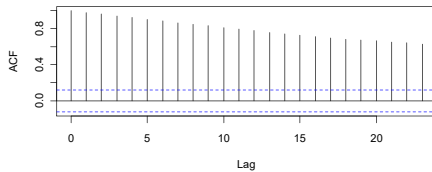
1. (X_t) est une marche aléatoire telle que

$$X_t = \sum_{i=1}^t \varepsilon_i, \quad (\varepsilon_t) \rightsquigarrow BBF(0, \sigma^2).$$

2. (X_t) est un processus avec tendance et variance fonction du temps tel que

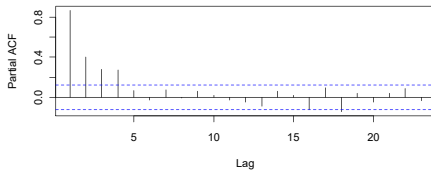
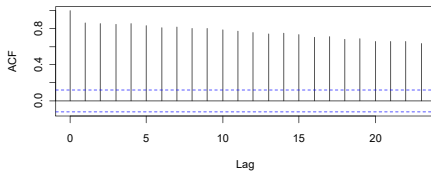
$$X_t = m_t + \kappa_t \varepsilon_t, \quad (\varepsilon_t) \rightsquigarrow BBF(0, \sigma^2).$$

Marche aléatoire



$$\varepsilon_t \sim \mathcal{N}(0, 1)$$

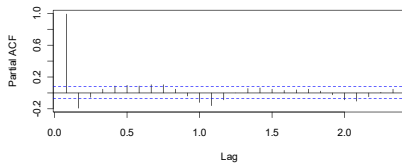
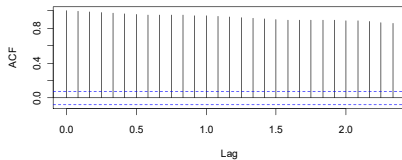
Processus avec tendance et variance fonction du temps



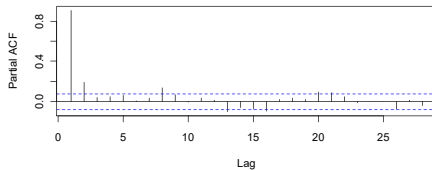
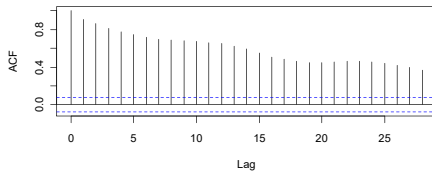
$$m_t = 0,005 \times t, \kappa_t = 0,001 \times t, \varepsilon_t \sim \mathcal{N}(0, 1)$$

EXEMPLES AVEC DES SERIES REELLES :

Concentration en CO2 - Série brute

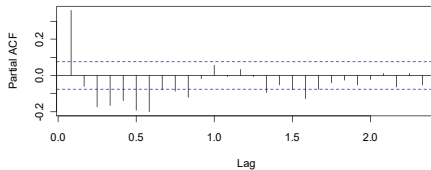
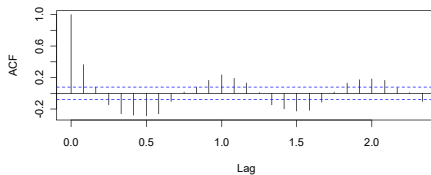


Concentration en CO2 - Série désaisonnalisée



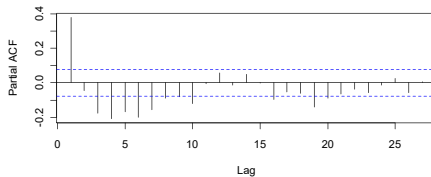
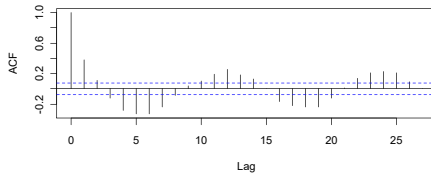
Série désaisonnalisée par la méthode de Buys-Balot

Concentration en CO2 - Série désaisonnalisée



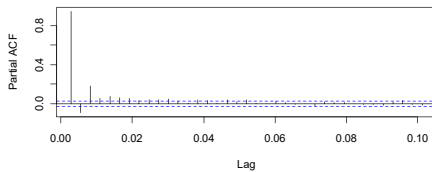
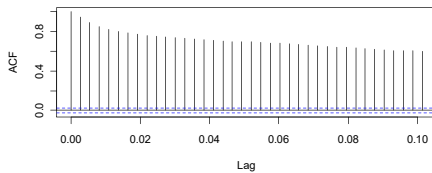
Série désaisonnalisée par la méthode STL

Concentration en CO2 - Série désaisonnalisée

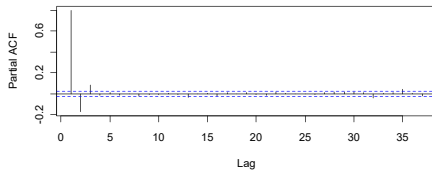
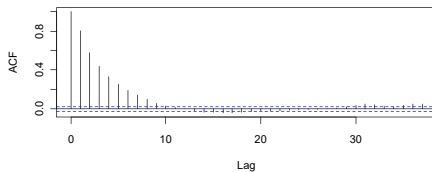


Série désaisonnalisée par la méthode des moyennes mobiles

Températures à Paris - Série brute

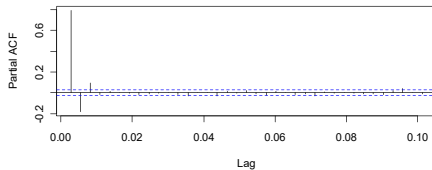
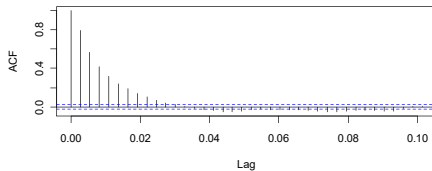


Concentration en CO2 - Série désaisonnalisée



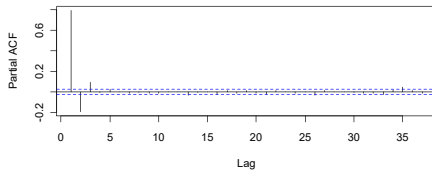
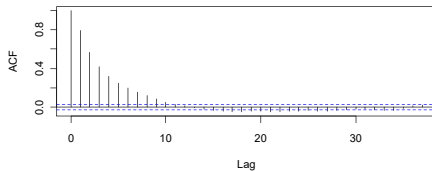
Série désaisonnalisée par la méthode de Buys-Balot

Températures à Paris - Série désaisonnalisée



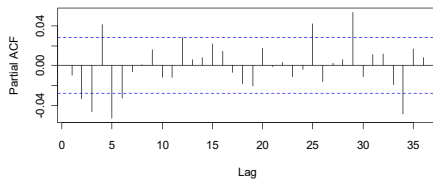
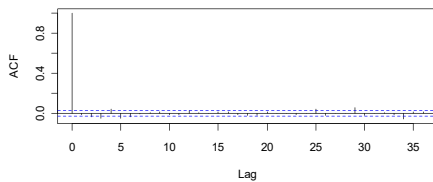
Série désaisonnalisée par la méthode STL

Températures à Paris - Série désaisonnalisée

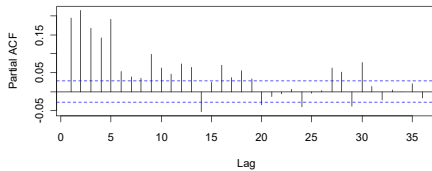
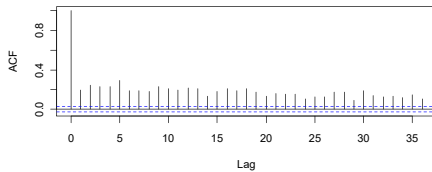


Série désaisonnalisée par la méthode des moyennes mobiles

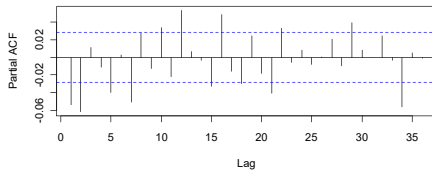
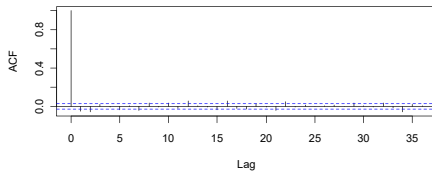
Rentabilités (logarithmiques) du CAC 40



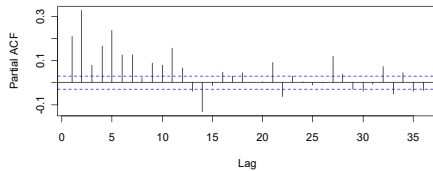
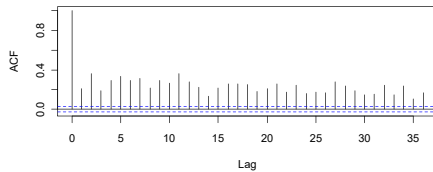
Rentabilités au carré du CAC 40



Rentabilités (logarithmiques) du S&P 500



Rentabilités au carré du S&P 500



4 Processus ARMA et ARIMA

4.1 Polynômes retard et avance

▷ Définitions et propriétés

DEFINITIONS :

1. L'opérateur retard L (lag) ou B (backward) est défini sur la classe des processus stationnaires comme suit :

$$L : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = LX_t = X_{t-1}$$

2. L'opérateur avance F (forward) est défini sur la classe des processus stationnaires comme suit :

$$F : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = FX_t = X_{t+1}$$

PROPOSITIONS :

1. $F = L^{-1}$ et $L = F^{-1}$
2. $L^k = L \cdot \dots \cdot L$ vérifie $L^k X_t = X_{t-k}$.
3. $F^k = F \cdot \dots \cdot F$ vérifie $F^k X_t = X_{t+k}$.

On note $L^0 = Id$.

DEFINITION :

Soit P un polynôme, $P(z) = \sum_{k=0}^p a_k z^k$ avec $a_k \in \mathbb{R}$ ($p \in \tilde{\mathbb{N}}$), on lui associe le polynôme retard défini comme suit :

$$P(L) = \sum_{k=0}^p a_k L^k$$

et

$$P(L)X_t = \left(\sum_{k=0}^p a_k L^k \right) X_t = \sum_{k=0}^p a_k X_{t-k}.$$

Le polynôme avance est défini par :

$$P(F) = \sum_{k=0}^p a_k F^k$$

et

$$P(F)X_t = \left(\sum_{k=0}^p a_k F^k \right) X_t = \sum_{k=0}^p a_k X_{t+k}.$$

PROPOSITIONS :

Soient les polynômes retard de degré "infini" définis par :

$$\underbrace{A(L)}_{\sim} = \sum_{k=0}^{\infty} a_k L^k \quad \underbrace{B(L)}_{\sim} = \sum_{k=0}^{\infty} b_k L^k$$

avec $\sum_{k=0}^{\infty} |a_k| < \infty$ et $\sum_{k=0}^{\infty} |b_k| < \infty$.

1. $\forall \alpha \in \mathbb{R}$

$$\alpha A(L) = (\alpha A)(L) = \sum_{k=0}^{\infty} (\alpha a_k) L^k$$

2.

$$A(L) + B(L) = (A + B)L = \sum_{k=0}^{\infty} (a_k + b_k) L^k$$

3.

$$A(L).B(L) = (A.B)(L) = C(L)$$



où $c_k = \sum_{j=0}^k a_j b_{k-j}$ et $\sum_{k=0}^{\infty} |c_k| < \infty$.


 $a * b$

▷ Inversibilité des polynômes en L

DEFINITION :

1. Un polynôme retard $A(L)$ est inversible si et seulement si il existe un polynôme ~~retard~~ $B(L)$ tel que :

$$A(L).B(L) = B(L).A(L) = Id.$$

PROPOSITIONS :

1. Si $|\lambda| < 1$, alors $1 - \lambda L$ est inversible et

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

2. Si $|\lambda| > 1$, alors $1 - \lambda L$ est inversible et

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

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

PROPOSITIONS :

Soit ϕ un polynôme retard de degré p à coefficients réels :

$$\phi(z) = 1 + \phi_1 z + \dots + \phi_p z^p.$$

1. $\phi(L)$ est inversible si et seulement si ses racines sont de module distinct de 1.

2. Si les modules des racines de ϕ sont tous strictement supérieurs à un, alors $\phi(L)$ est inversible et

$$\phi(L)^{-1} = \sum_{k=0}^{\infty} a_k L^k$$

où $a_0 = 1$, $a_k \in \mathbb{R}$, $\sum_{k=0}^{\infty} |a_k| < \infty$.

REMARQUES :

1. Les racines de l'équation caractéristique (notées z_j) :

$$1 + \phi_1 z + \dots + \phi_p z^p = 0$$

peuvent être aussi caractérisées en déterminant les racines (notées λ_j) de :

$$\lambda^p + \phi_1 \lambda^{p-1} + \dots + \phi_p = 0.$$

2. La condition 2. de la proposition précédente se réécrit :

$$|z_j| > 1 \Leftrightarrow |\lambda_j| < 1.$$

▷ Méthodes pratiques d'inversion de $\phi(L)$

On pose :

$$\phi(L) = \prod_{j=1}^p (1 - \lambda_j L)$$

avec $|\lambda_j| < 1$ pour tout $j = 1, \dots, p$.

METHODE 1 :

Les polynômes $(1 - \lambda_j L)$ sont inversibles, on obtient :

$$\phi(L)^{-1} = \prod_{j=1}^p \left(\sum_{k=0}^{\infty} \lambda_j^k L^k \right)$$

puis on procède par identification...

METHODE 2 :

On explicite l'équation :

$$\phi(z)\phi(z)^{-1} = 1$$

soit

$$\phi(L) \left(\sum_{k=0}^{\infty} a_k L^k \right) = 1.$$

On obtient un système d'équations qui peut être résolu de manière récursive :

$$1 = a_0$$

$$0 = a_1 + \phi_1$$

$$0 = a_2 + a_1\phi_1 + \phi_2$$

$\vdots$

$$0 = a_p + \phi_1 a_{p-1} + \dots + \phi_{p-1} a_1 + \phi_p$$

$$0 = a_k + \phi_1 a_{k-1} + \dots + \phi_p a_{k-p} \text{ pour } k > p$$

METHODE 3 :

Décomposition de la fraction rationnelle $\phi^{-1}(L)$ en éléments simples.

METHODE 4 :

Division selon les puissances croissantes de 1 par $\phi(z)$.

4.2 Processus $MA(q)$

DEFINITION :

Un processus stochastique stationnaire $(X_t)_{t \in \mathbb{Z}}$ est une moyenne mobile d'ordre q s'il satisfait l'équation suivante :

$$X_t = \mu + \Theta(L)\varepsilon_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q},$$

où les θ_j , $j = 1, \dots, q$, sont des nombres réels (avec $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 .

PROPRIETES :

1. $(X_t)_{t \in \mathbb{Z}}$ est nécessairement stationnaire indépendamment des racines du polynôme Θ .
2. $\mathbb{E}[X_t] = \mu$.

▷ Représentation canonique

PROPOSITION :

Soit le processus stochastique $(X_t)_{t \in \mathbb{Z}}$ moyenne mobile d'ordre q

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

pour lequel toutes les racines du polynôme :

$$\Theta(u) = 1 - \theta_1 u - \dots - \theta_q u^q$$

sont de module strictement supérieur à 1, alors $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

REMARQUES :

1. Un processus $MA(q)$ admet plusieurs représentations moyennes mobiles d'ordre q avec différents bruits blancs.
2. Une seule représentation correspond à la représentation canonique ou causale du processus pour laquelle le bruit blanc utilisé est l'innovation.
3. Le résultat précédent reste partiellement valide pour les processus dont le polynôme de la représentation moyenne mobile admet des racines de module unitaire.

▷ Ecriture $AR(\infty)$

i) Si toutes les racines de Θ sont de module strictement supérieur à 1, la représentation canonique $AR(\infty)$ est :

$$\sum_{k=0}^{\infty} a_k X_{t-k} = m + \varepsilon_t$$

avec $m = \mu/\Theta(1)$ et $a_0 = 1$.

ii) Si certaines racines ont un module égal à un,

- $(X_t)_{t \in \mathbb{Z}}$ est un processus stochastique stationnaire ;
- $\Theta(L)$ n'est plus inversible ;
- $(\varepsilon_t)_{t \in \mathbb{Z}}$ reste le processus des innovations.

iii) Si certaines racines de sont de module strictement inférieur à 1, il faut modifier la structure pour obtenir le processus d'innovation.

Supposons :

$$X_t = \prod_{i:|\lambda_i|<1} (1 - \lambda_i L) \prod_{i:|\lambda_i|>1} (1 - \lambda_i L) \varepsilon_t.$$

On définit :

$$\Theta^*(L) = \prod_{i:|\lambda_i|<1} (1 - \lambda_i L) \prod_{i:|\lambda_i|>1} (1 - \frac{1}{\lambda_i} L)$$

et $X_t = \Theta^*(L)\eta_t$, alors $(\eta_t)_{t \in \mathbb{Z}}$ est l'innovation du processus de variance

$$\sigma_\eta^2 = \sigma_\varepsilon^2 \prod_{i:|\lambda_i|>1} |\lambda_i|^2.$$

▷ Caractérisation d'un processus $MA(q)$

On se place dans l'hypothèse où toutes les racines du polynôme P_Θ sont de module strictement supérieur à 1, de telle sorte que $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

PROPOSITION :

1. Sous les hypothèses précédentes, la fonction d'autocovariance d'un processus moyenne mobile d'ordre q est définie par :

$$\begin{aligned}\gamma_X(0) &= \sigma_\varepsilon^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right) \\ \gamma_X(h) &= \sigma_\varepsilon^2 \left(-\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h} \right) \text{ pour } 1 \leq |h| < q \\ \gamma_X(h) &= -\theta_q \sigma_\varepsilon^2 \text{ pour } |h| = q \\ \gamma_X(h) &= 0 \text{ pour } |h| > q.\end{aligned}$$

2. Le processus stochastique $(X_t)_{t \in \mathbb{Z}}$ stationnaire du second ordre à densité spectrale absolument continue et strictement positive est un processus moyenne mobile d'ordre q si et seulement si :

$$\gamma_X(h) = \begin{cases} \neq 0, & |h| = q, \\ = 0, & |h| > q. \end{cases}$$

REMARQUES :

1. Il n'y a pas de résultat particulier pour les autocorrélations partielles.
2. Sous les hypothèses précédentes, la fonction d'autocorrélation inverse décroît exponentiellement avec h .

$$|\gamma_X(h)| \leq c e^{-\hat{c}h}$$

$$c, \hat{c} > 0$$

4.3 Processus $AR(p)$

DEFINITION :

Un processus stochastique $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif d'ordre p s'il est stationnaire et satisfait l'équation suivante :

$$\Phi(L)X_t = \mu + \varepsilon_t$$

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

$$\Phi(L) = 1 + \phi_1 L + \dots + \phi_p L^p$$

où les ϕ_j , $j = 1, \dots, p$, sont des nombres réels (avec $\phi_p \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 .

PROPRIETES :

1. Un processus $(X_t)_{t \in \mathbb{Z}}$ qui satisfait la récurrence $\Phi(L)X_t = \mu + \varepsilon_t$ n'est pas nécessairement stationnaire.
2. $\mathbb{E}[X_t] = m = \mu / \Phi(1)$.

▷ Représentation canonique

PROPOSITION :

Soit le processus stochastique $(X_t)_{t \in \mathbb{Z}}$ processus autorégressif d'ordre p

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t,$$

pour lequel toutes les racines du polynôme :

$$\Phi(u) = 1 - \phi_1 u - \dots - \phi_p u^p$$

sont de module strictement supérieur à 1, alors $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

REMARQUE :

$$EL(X_t | \underline{X_{t-1}}) = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}.$$

▷ Ecriture $MA(\infty)$

i) Si toutes les racines de Φ sont de module strictement supérieur à 1, la représentation canonique $MA(\infty)$ est :

$$X_t = m + \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

avec $m = \mu/\Theta(1)$, $a_0 = 1$ et $\sum_{k=0}^{\infty} |a_k| < \infty$.

ii) Si certaines racines ont un module égal à un,

- $(X_t)_{t \in \mathbb{Z}}$ n'est pas un processus stochastique stationnaire ;
- $\Phi(L)$ n'est plus inversible

iii) Si certaines racines de sont de module strictement inférieur à 1, il faut modifier la structure pour obtenir le processus d'innovation.

Supposons $\Phi(L)X_t = \varepsilon_t$ avec

$$\Phi(L) = \prod_{i:|\lambda_i|<1} (1 - \lambda_i L) \prod_{i:|\lambda_i|>1} (1 - \lambda_i L).$$

On définit :

$$\Phi^*(L) = \prod_{i:|\lambda_i|<1} (1 - \lambda_i L) \prod_{i:|\lambda_i|>1} (1 - \frac{1}{\lambda_i} L)$$

et $\Phi^*(L)X_t = \eta_t$, alors $(\eta_t)_{t \in \mathbb{Z}}$ est l'innovation du processus de variance

$$\sigma_\eta^2 = \prod_{i:|\lambda_i|>1} |\lambda_i|^{-2} \sigma_\varepsilon^2.$$

▷ Caractérisation d'un processus $AR(p)$

On se place dans l'hypothèse où toutes les racines du polynôme Φ sont de module strictement supérieur à 1, de telle sorte que $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

PROPOSITIONS :

1. Le processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire du second ordre est un processus autorégressif d'ordre p si et seulement si :

$$r_X(h) = \begin{cases} \neq 0, & |h| = p, \\ = 0, & |h| > p. \end{cases}$$

2. Le processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire du second ordre est un processus autorégressif d'ordre p si et seulement si :

$$\gamma_X^i(h) = \begin{cases} \neq 0, & |h| = p, \\ = 0, & |h| > p. \end{cases}$$

▷ Autocovariances et équations de Yule-Walker

PROPOSITIONS :

1. Sous les hypothèses précédentes, la suite des autocovariances satisfait une équation de récurrence homogène d'ordre p dont les coefficients sont ceux du polynôme autorégressif $u^p \Phi(u^{-1})$:

$$\gamma_X(h) - \sum_{j=1}^p \phi_j \gamma_X(h-j) = 0 \quad \forall h \geq 0.$$

2. Sous les hypothèses précédentes, la suite des autocorrélations satisfait une équation de récurrence homogène d'ordre p :

$$\rho_X(h) - \sum_{j=1}^p \phi_j \rho_X(h-j) = 0 \quad \forall h \geq 0.$$

▷ Résumé des propriétés $MA(q)$ et $AR(p)$:

– $AR(p)$

1. La suite des autocorrélations $(|\rho_X(h)|)$ décroît exponentiellement vers 0 avec h .
2. Les autocorrélations partielles $(r(h))$ sont nulles pour $|h| > p$.
3. Les autocorrélations inverses $(\rho_X^i(h))$ sont nulles pour $|h| > p$.

– $MA(q)$

1. Les autocorrélations $(\rho_X(h))$ sont nulles pour $|h| > q$.
2. Aucun résultat particulier pour les autocorrélations partielles $(r(h))$.
3. La suite des autocorrélations inverses $(\rho_X^i(h))$ décroît exponentiellement vers 0 avec h .

4.4 Processus $ARMA(p, q)$

DEFINITION :

Un processus stochastique stationnaire $(X_t)_{t \in \mathbb{Z}}$ est un processus $ARMA(p, q)$ s'il satisfait l'équation suivante :

$$\begin{aligned} X_t &= \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \Phi(L)X_t &= \mu + \Theta(L)\varepsilon_t \end{aligned}$$

où les $\phi_j, j = 1, \dots, p$, et les $\theta_j, j = 1, \dots, q$, sont des nombres réels (avec $\phi_p \neq 0$ et $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 .

PROPRIETES :

1. $(X_t)_{t \in \mathbb{Z}}$ n'est pas nécessairement stationnaire.
2. $\mathbb{E}[X_t] = m = \mu / \Phi(1)$.

REMARQUES :

1. En l'absence de contraintes sur Φ et Θ , un processus stationnaire admettant une représentation *ARMA* possède d'autres représentations de ce type. Ces autres représentations sont obtenues en étudiant la forme de la densité spectrale.
2. En particulier, les deux processus stochastiques :

$$\begin{aligned}\Phi(L)X_t &= \mu + \Theta(L)\varepsilon_t \\ \tilde{\Phi}(L)X_t &= \mu + \tilde{\Theta}(L)\tilde{\varepsilon}_t\end{aligned}$$

sont deux représentations du même processus si :

$$f_X(\omega) = \frac{\sigma_\varepsilon^2 |\Theta(\exp(i\omega))|^2}{2\pi |\Phi(\exp(i\omega))|^2} = \frac{\sigma_{\tilde{\varepsilon}}^2 |\tilde{\Theta}(\exp(i\omega))|^2}{2\pi |\tilde{\Phi}(\exp(i\omega))|^2}$$

Cette égalité est vraie (au moins) dans les 2 situations suivantes :

- i) lorsqu'on change les racines de Φ et Θ en leurs inverses ;
- ii) en multipliant ou en divisant et par un même facteur polynômial.

▷ Représentation canonique

PROPOSITION :

Soit le processus stochastique $(X_t)_{t \in \mathbb{Z}}$ $ARMA(p, q)$

$$\begin{aligned} X_t = & \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \\ & + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \end{aligned}$$

pour lequel toutes les racines des polynômes :

$$\Phi(u) = 1 - \phi_1 u - \dots - \phi_p u^p$$

$$\Theta(u) = 1 - \theta_1 u - \dots - \theta_q u^q$$

sont de module strictement supérieur à 1, alors $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.



▷ Ecritures $MA(\infty)$ et $AR(\infty)$

Si toutes les racines de Φ et de Θ sont de module strictement supérieur à 1, la représentation canonique $MA(\infty)$ est

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

avec $a_0 = 1$ et $\sum_{k=0}^{\infty} |a_k| < \infty$, et la représentation canonique $AR(\infty)$ est

$$\varepsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}$$

avec $b_0 = 1$ et $\sum_{k=0}^{\infty} |b_k| < \infty$.

DEFINITION :

La représentation

$$\Phi(L)X_t = \mu + \Theta(L)\varepsilon_t$$

est dite canonique minimale si Φ et Θ sont des polynômes dont toutes les racines sont de module strictement supérieur à un et premiers entre eux et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est son processus des innovations.

REMARQUES :

1. Un processus $ARMA(p, q)$ canonique minimale admet une représentation moyenne mobile infinie dont les coefficients satisfont une équation de récurrence d'ordre p à partir du rang $\max(q + 1, p)$.
2. Un processus $ARMA(p, q)$ canonique minimale admet une représentation autorégressive infinie dont les coefficients satisfont une équation de récurrence d'ordre q à partir du rang $\max(p + 1, q)$.

▷ Caractérisation d'un processus $ARMA(p, q)$

PROPOSITION :

Les autocovariances et autocorrélations d'un processus stationnaire $ARMA(p, q)$ satisfont une équation de récurrence d'ordre p pour $|h| > q$ et décroissent vers 0 exponentiellement vite.



4.5 Processus $ARIMA(p, d, q)$

Ar, delay

$(1-L)^d = \Delta^d$
 Δ : difference

DEFINITION :

Un processus stochastique $(X_t)_{t \geq -p-d}$ est un processus $ARIMA(p, d, q)$ s'il satisfait l'équation suivante :

$$\begin{aligned}(1-L)^d X_t &= \mu + \phi_1(1-L)^d X_{t-1} + \dots + \phi_p(1-L)^d X_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \Phi(L)(1-L)^d X_t &= \mu + \Theta(L)\varepsilon_t\end{aligned}$$

où les ϕ_j , $j = 1, \dots, p$, et les θ_j , $j = 1, \dots, q$, sont des nombres réels (avec $\phi_p \neq 0$ et $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 , et les conditions initiales

$$Z_{-1} = \{X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont telles que :

$$Cov(\varepsilon_t, Z_{-1}) = 0 \quad \forall t \geq 0.$$

▷ Représentation canonique minimale

Si $(X_t)_{t \geq -p-d}$ est un processus $ARIMA(p, d, q)$ pour lequel toutes les racines des polynômes :

$$\Phi(u) = 1 - \phi_1 u - \dots - \phi_p u^p$$

$$\Theta(u) = 1 - \theta_1 u - \dots - \theta_q u^q$$

sont de module strictement supérieur à 1, alors on obtient la représentation canonique minimale.

Si on pose $Y_t = (1 - L)^d X_t$, on obtient

$$\Phi(L)Y_t = \mu + \Theta(L)\varepsilon_t.$$

Mais cela ne signifie pas que Y_t est un processus ARMA...puisque'il n'est pas défini $\forall t \in \mathbb{Z}$! Il y a seulement équivalence asymptotique avec un processus ARMA.

DEFINITION :

Si (i) $(1 - L)^d X_t$ est asymptotiquement équivalent à un processus stochastique stationnaire et (ii) $(1 - L)^{d-1} X_t$ ne l'est pas, alors on dit que le processus stochastique discret (X_t) est intégré d'ordre d et on note $X_t \sim I(d)$.

EXEMPLES :

1. Si $d = 0$, X_t est un $ARMA(p, q)$ (processus stationnaire) et on note $X_t \sim I(0)$.
2. Si $d = 1$, (X_t) est un processus intégré d'ordre 1 (il admet une racine unitaire).
3. Si $d = 2$, (X_t) n'est pas stationnaire, $Y_t = (1 - L)X_t$ non plus. $Z_t = (1 - L)Y_t = (1 - L)^2 X_t$ est asymptotiquement équivalent à un processus stationnaire.

4.6 Processus $SARIMA_s[(p, d, q), (P, D, Q)]$

Les séries dites saisonnières sont caractérisées par de fortes autocorrélations et autocorrélations partielles empiriques entre les variables décalées d'une ou plusieurs saisons. Les modèles ARIMA saisonniers permettent de modéliser une composante saisonnière de nature aléatoire, i.e. qui ne se répète pas à l'identique d'un cycle à l'autre.

DEFINITION :

On appelle processus $SARIMA_s[(p, d, q), (P, D, Q)]$ de période s tout processus (X_t) tel que le processus

↙ élimine les périodicité s

$$Y_t = (I - L)^d (I - L^s)^D X_t$$

est un ARMA

élimine tendance de degré $\leq d$

$$\Phi(L)\phi(L^s)Y_t = \Theta(L)\theta(L^s)\varepsilon_t$$

où Φ , ϕ , Θ et θ sont des polynômes de degrés p , P , q et Q respectivement.

p P q Q

◦ Sous

Pour simuler un processus *ARIMA*, on utilise la fonction

```
arima.sim(model, n, rand.gen = rnorm,  
  innov = rand.gen(n, ...), n.start = NA,  
  start.innov = rand.gen(n.start, ...), ...)
```

Pour obtenir la décomposition $MA(\infty)$ d'un processus *ARMA*, on utilise la fonction

```
ARMAtoMA(ar = numeric(0), ma = numeric(0), lag.max)
```

5 Estimation des modèles ARMA et ARIMA

5.1 Introduction

On dispose des données $x_1, \dots, x_T$ réalisations des variables aléatoires $X_1, \dots, X_T$.

On suppose que ces observations sont compatibles avec un modèle $ARIMA(p, d, q)$:

$$\Phi(L)\Delta^d X_t = \mu + \Theta(L)\varepsilon_t \quad \Delta = 1 - L$$

où

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

$$\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$$

avec $d^\circ(\Phi) = p$, $d^\circ(\Theta) = q$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$, et où les hypothèses sur les polynômes de retard et (ε_t) sont satisfaites.

▷ L'objectif est d'estimer un tel modèle et de déterminer ultérieurement la prévision théorique d'une valeur future sachant un certain ensemble d'information.

Plus précisément, on procède en 4 phases

1. Phase d'identification a priori

(a) Recherche des valeurs plausibles des degrés (p, d, q)

(b) Plusieurs étapes :

- Ordre de différenciation : autocorrélogramme, autocorrélogramme inverse, tests de racines unitaires...

- Ordre p et q : utilisation des propriétés des processus $ARMA(p, q)$, fonctions d'autocorrélation et d'autocorrélation partielle, fonction d'autocorrélation inverse, méthode du coin, etc.

A ce stade, plusieurs modèles sont généralement retenus : (p_j, d_j, q_j) $j = 1, \dots, J$.

2. Phase d'estimation

Pour chaque triplet (p_j, d_j, q_j) , on recherche les valeurs correspondantes des autres paramètres du modèle :

$$(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \mu, \sigma_\varepsilon^2).$$

Plusieurs méthodes sont disponibles :

- maximum de vraisemblance (conditionnelle, non conditionnelle),
- moindres carrés conditionnels,
- méthode des moments généralisés, **MnG**
- filtre de Kalman et modèles espace-état,
- etc.

3. Phase de vérification

Il s'agit de “soumettre” les modèles estimés à des tests pour vérifier les hypothèses sous-jacentes à ces modèles ou tout simplement la significativité des coefficients estimés.

- Les résidus estimés prennent-ils des valeurs correspondant à l'hypothèse d'un bruit blanc ?
- Les paramètres sont-ils significatifs ?
- Peut-on diminuer les valeurs de p et q ?

4. Choix d'un modèle

On peut avoir recours à plusieurs critères : critère d'information, critère de parcimonie, critère de pouvoir prédictif.

5.2 Phase d'identification a priori

Trois questions :

1. Est-ce que la série étudiée est stationnaire ? non-stationnaire ? Faut-il éliminer une tendance déterministe et/ou une saisonnalité déterministes ? Quel est l'ordre de différenciation ?

2. Existe-t-il une composante saisonnière (autorégressive, moyenne mobile), *ARIMA* versus *SARIMA* ?

3. Quels sont les ordres possibles pour p et q ?

▷ Etude de la stationnarité : détermination de d

Deux méthodes :

1. Inspection visuelle de la fonction d'autocorrélation empirique (et fonction d'autocorrélation inverse) ;
2. Tests de racine(s) unitaire(s).

I) Utilisation de l'autocorrélogramme

1. Pour un processus stationnaire, on observe une décroissance de la fonction d'autocorrélation vers zéro avec h .

PROPOSITION :

1. Les autocovariances et autocorrélations d'un processus stationnaire (canonique et minimal) $ARMA(p, q)$ satisfont une équation de récurrence d'ordre p pour $|h| > q$:

$$\gamma_X(h) - \sum_{j=1}^p \phi_j \gamma_X(h-j) = 0 \quad \forall |h| > q,$$

ou

$$\rho_X(h) - \sum_{j=1}^p \phi_j \rho_X(h-j) = 0 \quad \forall |h| > q.$$

2. Les autocovariances et autocorrélations décroissent vers 0 exponentiellement vite avec h .

2. Si un processus stochastique (X_t) est intégré d'ordre d , le processus (X_t) n'est pas stationnaire et on ne peut pas définir la fonction d'autocorrélation. Cependant, la fonction

$$\rho_X(t, h) = \frac{\mathbb{C}ov(X_t, X_{t+h})}{\sqrt{\mathbb{V}ar(X_t)\mathbb{V}ar(X_{t+h})}}$$

a toujours un sens.

EXEMPLE :

Considérons une marche aléatoire sans dérive :

$$X_t = X_{t-1} + \varepsilon_t$$

où $\varepsilon_t \rightsquigarrow BB(0, \sigma_\varepsilon^2)$ et $\mathbb{C}ov(\varepsilon_t, X_0) = 0$ si $t \geq 0$. On a :

$$\rho_X(t, h) = \frac{\mathbb{V}ar(X_0) + t\sigma_\varepsilon^2}{\sqrt{(\mathbb{V}ar(X_0) + t\sigma_\varepsilon^2)(\mathbb{V}ar(X_0) + (t+h)\sigma_\varepsilon^2)}}$$

et pour t grand et $h \ll t$,

$$\rho_X(h) \approx 1 - \frac{h}{2t}.$$

L'examen de la fonction d'autocorrélation aboutit alors à adopter les règles suivantes :

– Règle 1 : Si la fonction d'autocorrélation estimée $\hat{\rho}_X(h)$ reste relativement proche de 1 pour un assez grand nombre de valeurs de h ou décroît linéairement avec h , alors le processus est certainement non stationnaire.

Remarque : Cette règle peut aussi s'interpréter comme suit : le processus est certainement non stationnaire si les premiers $\hat{\rho}_X(h)$ sont proches les uns des autres.

– Règle 2 : Une fois différenciée la série initiale, on étudie l'autocorrélogramme de la série $Y_t = \Delta X_t$ (pour différencier éventuellement à l'ordre 2) et ainsi de suite...

Remarques :

i) L'application des règles précédentes peut cependant conduire à sur-différencier la série initiale. Dans cette perspective, on peut utiliser l'information fournie par l'autocorrélogramme inverse, car en sur-différenciant, les autocorrélations inverses ne décroissent pas vers 0 exponentiellement vite avec h .

ii) La nature de la non-stationarité n'est pas unique : une série non-stationnaire est une série qui n'est pas stationnaire. Si la série étudiée est stationnaire en tendance (déterministe), alors une différenciation à l'ordre d conduira à l'apparition d'une moyenne mobile !

II) Approche des tests

Il est possible d'utiliser des tests (dits) de racine(s) unitaire(s). L'idée de base de ces tests est de savoir si la série une fois (ou plusieurs fois différenciée) est stationnaire. Différents tests ont été proposés dans la littérature : Tests de Dickey-Fuller, Tests de Dickey-Fuller augmentés, Tests de Phillips-Perron, Tests de Schmidt-Phillips, Test KPSS, etc...

▷ Etude de la saisonnalité

- Certaines séries (par exemple, mensuelles) ont un profil saisonnier marqué : les données relatives à un même mois de différentes années ont tendance à se situer de manière semblable par rapport à la moyenne annuelle.
- Cette saisonnalité peut être détectée à partir de :
 - la représentation graphique de la série (et de toute information décrivant la variable),
 - un découpage de la série en plusieurs périodes,
 - la fonction des autocorrélations,
 - la densité spectrale, etc...

▷ Etude de l'ordre p et q

- Il s'agit de déterminer les ordres de la partie autorégressive et de la partie moyenne mobile.
- En pratique, on retiendra une borne supérieure pour chacun de ces ordres (p_{max} et q_{max}). En d'autres termes, différents modèles seront estimés.
- L'identification des ordres p et q repose sur la fonction d'autocorrélations et la fonction d'autocorrélations partielles. Pour un processus autorégressif ou moyenne mobile pur, l'identification est "simple". Cependant, c'est beaucoup plus complexe pour un processus *ARMA*.

1) Les différentes caractérisations

Les propriétés connues des processus *ARMA* peuvent être utilisées.

- Un processus *AR* d'ordre p se caractérise par sa fonction d'autocorrélation partielle qui s'annule à partir de l'ordre $p + 1$.

- Un processus MA d'ordre q se caractérise par sa fonction d'autocorrélation qui s'annule à partir de l'ordre $q + 1$.
- Un processus $ARMA(p, q)$ se caractérise par les propriétés de ces corrélations canoniques.

DEFINITION :

Les corrélations canoniques entre les vecteurs $(X_t, X_{t-1}, \dots, X_{t-j})'$ et $(X_{t-k-1}, \dots, X_{t-k}$ sont les valeurs propres de la matrice :

$$\Gamma(j, -1)^{-1} \Gamma(j, k)' \Gamma(j, -1)^{-1} \Gamma(j, k)$$

où

$$\Gamma(j, k) = \begin{pmatrix} \gamma_X(k+1) & \gamma_X(k+2) & \cdots & \gamma_X(k+1+j) \\ \gamma_X(k) & \gamma_X(k+1) & \cdots & \gamma_X(k+j) \\ \vdots & \vdots & & \vdots \\ \gamma_X(k+1-j) & \cdots & \gamma_X(k) & \gamma_X(k+1) \end{pmatrix}$$

1. En particulier, le produit des diverses corrélations canoniques est donné par :

$$\frac{\Delta(j, k)^2}{\Delta(j, -1)^2}$$

où $\Delta(j, k) = \det \Gamma(j, k)$.

2. On peut montrer que :

$$\Delta(j, k)^2 = \Delta(j + 1, k)\Delta(j - 1, k) + \Delta(j, k - 1)\Delta(j, k + 1).$$

3. Soit un processus (X_t) admettant une représentation canonique minimale $ARMA(p, q)$. $\Delta(j, k)$ vérifie les propriétés suivantes :

- $\Delta(j, k) = 0 \ \forall j \geq p \text{ et } \forall k \geq q,$
- $\Delta(p, q - 1) \neq 0,$
- $\Delta(p - 1, q) \neq 0,$

$$- \Delta(j, q-1) \neq 0 \quad \forall j \geq p-1,$$

$$- \Delta(p-1, k) \neq 0 \quad \forall k \geq q-1.$$

Caractérisation pour différents processus :

1. Pour un processus $ARMA(p, q)$, on obtient le tableau

$$\begin{pmatrix} D_{p,q} & D_q \\ D_p & 0 \end{pmatrix}$$

où $D_{p,q}$ est défini par

j/k	0	1	2	...	q
0	$\Delta(0, 0)$	$\Delta(0, 1)$	$\Delta(0, 2)$	$\dots$	$\Delta(0, q)$
1	$\Delta(1, 0)$	$\Delta(1, 1)$	$\Delta(1, 2)$	$\dots$	$\Delta(1, q)$
2	$\Delta(2, 0)$	$\Delta(2, 1)$	$\Delta(2, 2)$	$\dots$	$\Delta(2, q)$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
p	$\Delta(p, 0)$	$\Delta(p, 1)$	$\Delta(p, 2)$	$\dots$	0

D_q est défini par

$$\begin{array}{ccccc}
 j/k & & q+1 & & q+2 \\
 0 & & \Delta(0, q+1) & & \Delta(0, q+2) \\
 1 & & \Delta(1, q+1) & & \Delta(1, q+2) \\
 2 & & \Delta(2, q+1) & & \Delta(2, q+2) \\
 & & \vdots & & \vdots \\
 p-1 & \Delta(p-1, q+1) & & \Delta(p-1, q+2) & \\
 p & 0 & & 0 &
 \end{array}$$

et D_p est défini par

$$\begin{array}{ccccccc}
 j/k & 0 & 1 & \dots & q-1 & q \\
 p+1 & \Delta(p+1, 0) & \Delta(p+1, 1) & \dots & \Delta(p+1, q-1) & 0 \\
 p+2 & \Delta(p+1, 0) & \Delta(p+2, 1) & \dots & \Delta(p+2, q-1) & 0
 \end{array}$$

2. Dans le cas d'un processus $MA(q)$, on obtient le tableau :

$j/k \cdot \cdot$	0	1	2	...	$q-1$	q	$q+1$
0	$\Delta(0,0)$	$\Delta(1,1)$	$\Delta(1,2)$	...	$\Delta(1,p)$	0	0
1	$\Delta(1,0)$	$\Delta(1,1)$	$\Delta(1,2)$	...	$\Delta(1,p)$	0	0
2	$\Delta(2,0)$	$\Delta(2,1)$	$\Delta(2,2)$	...	$\Delta(2,p)$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
p	$\Delta(p,0)$	$\Delta(q,1)$	$\Delta(q,2)$	...	$\Delta(q,p)$	0	0

3. Dans le cas d'un processus $AR(p)$, on obtient le tableau :

j/k	0	1	2	...	q
1	$\Delta(1,0)$	$\Delta(1,1)$	$\Delta(1,2)$	...	$\Delta(1,q)$
2	$\Delta(2,0)$	$\Delta(2,1)$	$\Delta(2,2)$	...	$\Delta(2,q)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$p-1$	$\Delta(p-1,0)$	$\Delta(p-1,1)$	$\Delta(p-1,2)$	...	$\Delta(p-1,q)$
p	0	0	0	...	0
$p+1$	0	0	0	...	0

II) Utilisation des moments empiriques

- Toutes les caractéristiques précédentes reposent sur les valeurs prises par la fonction d'autocovariances (partielles).
- En utilisant les résultats du chapitre 3 (comportement asymptotique des estimateurs empiriques), des tests peuvent être construits.
- Les deux tests les plus usuels sont ceux de nullité de la fonction d'autocovariance et de la fonction d'autocorrélation partielle.

1. Pour un processus $MA(q)$, on construit une statistique de test à partir de la relation

$$\sqrt{T}(\hat{\gamma}_X(h+q) - \gamma_X(h+q)) \xrightarrow{L} \mathcal{N}\left(0, \gamma_X^2(0) + 2 \sum_{j=1}^q \gamma_X^2(j)\right)$$

ou

$$\sqrt{T}(\hat{\rho}_X(h+q) - \rho_X(h+q)) \xrightarrow{L} \mathcal{N}\left(0, 1 + 2 \sum_{j=1}^q \rho_X^2(j)\right).$$

2. Pour un processus $AR(p)$, on construit une statistique de test à partir de la relation

$$\sqrt{T} \begin{pmatrix} \hat{r}_X(p+1) \\ \hat{r}_X(p+2) \\ \vdots \\ \hat{r}_X(p+h) \end{pmatrix} \xrightarrow{L} \mathcal{N}(0, Id_h)$$

3. Pour un processus ARMA, on peut utiliser la méthode du coin. En utilisant les corrélations canoniques, les ordres p et q sont tels que :

- $\tilde{\Delta}_{j,k} = 0 \quad \forall j \geq p \text{ et } \forall k \geq q,$
- $\tilde{\Delta}_{j,q-1} \neq 0 \quad \forall j \geq p-1,$

$$- \tilde{\Delta}_{p-1,k} \neq 0 \quad \forall k \geq q-1.$$

avec

$$\tilde{\Delta}_{j,k} = \begin{vmatrix} \rho_X(k+1) & \rho_X(k+2) & \dots & \rho_X(k+j+1) \\ \rho_X(k) & \rho_X(k+1) & \dots & \rho_X(k+j) \\ \vdots & & & \\ \rho_X(k+1-j) & \dots & \dots & \rho_X(k+1) \end{vmatrix}$$

– On peut estimer les $\tilde{\Delta}_{j,k}$ en remplaçant les corrélations par leurs estimateurs et sa variance en remarquant qu'il s'agit d'une fonction différentiable du vecteur des autocorrélations. On considère alors les rapports de Student associées aux $\tilde{\Delta}_{j,k}$

$$\left| \frac{\widehat{\tilde{\Delta}}_{j,k}}{\sqrt{\widehat{\text{Var}}(\widehat{\tilde{\Delta}}_{j,k})}} \right|.$$

En comparant à 1,96 la valeur de ce rapport, on obtient un test asymptotique de nullité du coefficient au seuil de 5%.

5.3 Phase d'estimation

▷ Estimation d'un modèle autorégressif

Soit le processus stochastique discret $(X_t)_{t \in \mathbb{Z}}$:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

où $\varepsilon_t \rightsquigarrow BB(0, \sigma_\varepsilon^2)$

I) Equations de Yule-Walker

Les équations de Yule-Walker sont données par

$$\begin{pmatrix} \gamma_X(1) \\ \gamma_X(2) \\ \vdots \\ \gamma_X(p) \end{pmatrix} = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(p-1) \\ \gamma_X(1) & \gamma_X(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \gamma_X(1) \\ \gamma_X(p-1) & \cdots & \gamma_X(1) & \gamma_X(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

et

$$\sigma_{\varepsilon}^2 = \gamma_X(0) - \left(\gamma_X(1) \quad \gamma_X(2) \quad \cdots \quad \gamma_X(p) \right) \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

– Sous forme matricielle :

$$\begin{aligned} \underline{\gamma} &= \Gamma_{p-1} \underline{\phi} \\ \sigma_{\varepsilon}^2 &= \gamma_X(0) - \underline{\gamma}' \underline{\phi} \end{aligned}$$

avec

$$\underline{\gamma} = \begin{pmatrix} \gamma_X(1) \\ \gamma_X(2) \\ \vdots \\ \gamma_X(p) \end{pmatrix} \quad \underline{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

En utilisant les estimateurs des moments empiriques, les estimateurs de Yule-Walker sont donnés par :

$$\begin{cases} \hat{\underline{\phi}} = \hat{\Gamma}_{p-1}^{-1} \hat{\underline{\gamma}} \\ \hat{\sigma}_{\varepsilon}^2 = \hat{\gamma}_X(0) - \hat{\underline{\gamma}}' \hat{\underline{\phi}} \end{cases}$$

PROPOSITION :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique discret admettant une représentation $AR(p)$.
 Sous certaines conditions de régularité, lorsque T tend vers $+\infty$:

$$\sqrt{T}(\hat{\underline{\phi}} - \underline{\phi}) \xrightarrow{L} \mathcal{N}(0, \sigma_{\varepsilon}^2 \Gamma_{p-1}^{-1})$$

et $\hat{\sigma}_{\varepsilon}^2 \xrightarrow{P} \sigma_{\varepsilon}^2$.



II) Estimation par maximum de vraisemblance

La fonction de vraisemblance s'écrit :

$$L(x_1, x_2, \dots, x_{t-1}, x_t; \underline{\phi}, \sigma_\varepsilon^2) \\ = \prod_{t=p+1}^T L(x_t | x_{t-1}, \dots, x_{t-p}; \underline{\phi}, \sigma_\varepsilon^2) \prod_{t=2}^p L(x_t | x_{t-1}, \dots, x_1; \underline{\phi}, \sigma_\varepsilon^2) L(x_1; \underline{\phi}, \sigma_\varepsilon^2)$$

La partie de la vraisemblance qui correspond aux conditions initiales du processus peut être traitée comme une constante (maximum de vraisemblance conditionnelle) ou non (maximum de vraisemblance non conditionnelle).

En pratique elle est négligée dans le calcul de la maximisation et on considère uniquement la fonction de log-vraisemblance conditionnelle qui s'écrit sous l'hypothèse que ε a une distribution Gaussienne (à une constante près) :

$$\begin{aligned} & \log \prod_{t=p+1}^T L(x_t | x_{t-1}, \dots, x_{t-p}; \underline{\phi}, \sigma_\varepsilon^2) \\ &= -\frac{T-p}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^T \left(x_t - \sum_{j=1}^p \phi_j x_{t-j} \right)^2 \end{aligned}$$

L'estimateur du maximum de vraisemblance de $(\underline{\phi}, \sigma_\varepsilon^2)$ est donc celui des moindres carrés ordinaires. L'estimateur du maximum de vraisemblance de σ_ε^2 est donné en particulier par :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_t^2,$$

où $\hat{\varepsilon}_t^2 = x_t - \sum_{j=1}^p \hat{\phi}_j x_{t-j}$.

PROPOSITION :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique discret admettant une représentation $AR(p)$ où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est normalement distribuée de variance σ_ε^2 . Lorsque T tend vers $+\infty$:

$$\sqrt{T}(\hat{\underline{\phi}} - \underline{\phi}) \xrightarrow{L} \mathcal{N}\left(0, \sigma_\varepsilon^2 \Gamma_{p-1}^{-1}\right)$$

et $\hat{\sigma}_\varepsilon^2 \xrightarrow{P} \sigma_\varepsilon^2$.

▷ Estimation d'une moyenne mobile

Soit le processus stochastique discret $(X_t)_{t \in \mathbb{Z}}$:

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

où $\varepsilon_t \rightsquigarrow BB(0, \sigma_\varepsilon^2)$

I) Estimation par maximum de vraisemblance

Il existe des matrices $A_1(\theta_1, \dots, \theta_q)$ de taille $T \times T$ triangulaire inférieure et $A_2(\theta_1, \dots, \theta_q)$ de taille $T \times q$ telles que

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix} = A_1(\theta_1, \dots, \theta_q) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} + A_2(\theta_1, \dots, \theta_q) \begin{pmatrix} \varepsilon_{1-q} \\ \varepsilon_{2-q} \\ \vdots \\ \varepsilon_0 \end{pmatrix}$$

ce qui peut encore se réécrire de la manière suivante

$$\begin{pmatrix} \varepsilon_{1-q} \\ \varepsilon_{2-q} \\ \varepsilon_T \end{pmatrix} = \begin{pmatrix} Id_q & 0 \\ A_1(\theta_1, \dots, \theta_q) & A_2(\theta_1, \dots, \theta_q) \end{pmatrix} \begin{pmatrix} \varepsilon_{0:1-q} \\ \underline{x} \end{pmatrix}$$

avec $\varepsilon_{0:1-q} = (\varepsilon_{1-q}, \dots, \varepsilon_0)'$.

– La densité de probabilité du vecteur $(\varepsilon_{1-q}, \dots, \varepsilon_T)$ sous l'hypothèse Gaussienne est donc :

$$(2\pi\sigma_\varepsilon^2)^{-(q+T)/2} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \begin{pmatrix} \varepsilon'_{0:1-q} & \underline{x}' \end{pmatrix} A' A \begin{pmatrix} \varepsilon_{0:1-q} \\ \underline{x} \end{pmatrix} \right)$$

– La loi marginale de $\underline{x}$ s'en déduit :

$$(2\pi\sigma_\varepsilon^2)^{-T/2} \det(I_q + A_2' A_2)^{-1/2} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \underline{x}' A_1' (Id_T + A_2 A_2') \underline{x} \right)$$

– Il ne reste plus qu'à maximiser par rapport à $(\theta_1, \dots, \theta_q)$ et σ_ε^2 .

PROPOSITION :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique discret admettant une représentation $MA(q)$ où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est normalement distribuée de variance σ_ε^2 . Lorsque T tend vers $+\infty$:

$$\sqrt{T}(\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{L} \mathcal{N}(0, \sigma_\varepsilon^2 \Gamma_{q-1}^{-1})$$

et $\hat{\sigma}_\varepsilon^2 \xrightarrow{P} \sigma_\varepsilon^2$, où Γ_{q-1} est la matrice des autocovariances d'un processus $(Y_t)_{t \in \mathbb{Z}}$ autorégressif d'ordre q qui satisfait une équation de la forme :

$$\eta_t = Y_t + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_q Y_{t-q}$$

où $\eta_t \rightsquigarrow BB(0, \sigma_\eta^2)$.

▷ Estimation d'un modèle $ARMA(p, q)$

I) On procède par une estimation en deux étapes

– Etape 1 : On utilise les équations de Yule-Walker pour estimer les paramètres de la forme autorégressive.

– Etape 2 : On peut alors considérer le modèle moyenne mobile approximativement satisfait par le processus :

$$x_t - \hat{\phi}_1 x_{t-1} - \dots - \hat{\phi}_p x_{t-p}.$$

Cette méthode conduit à des estimateurs moins efficaces que ceux du maximum de vraisemblance.

II) Estimation par maximum de vraisemblance

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique discret admettant une représentation $ARMA(p, q)$ où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est normalement distribuée de variance σ_ε^2 . Lorsque T tend vers $+\infty$:

$$\sqrt{T} \begin{pmatrix} \hat{\phi} - \phi \\ \hat{\theta} - \theta \end{pmatrix} \xrightarrow{L} \mathcal{N}(0, \sigma_\varepsilon^2 \Gamma^{-1})$$

et $\hat{\sigma}_\varepsilon^2 \xrightarrow{P} \sigma_\varepsilon^2$, où Γ est la matrice des autocovariances du vecteur

$$(Y_t, \dots, Y_{t-p+1}, Z_t, \dots, Z_{t-q+1})$$

où $(Y_t)_{t \in \mathbb{Z}}$ est un processus autorégressif d'ordre p qui satisfait une équation de la forme :

$$\eta_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p}$$

et où $(Z_t)_{t \in \mathbb{Z}}$ est un processus autorégressif d'ordre q qui satisfait une équation de la forme :

$$\eta_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$$

où $\eta_t \rightsquigarrow BB(0, \sigma_\eta^2)$.

5.4 Phase de vérification

Deux principaux types de tests :

1. les tests concernant l'estimation des paramètres,
2. les tests concernant les hypothèses sur ε .

▷ Tests sur les ordres p et q

Considérons un modèle $ARMA(p, q)$ de la forme

$$\Phi(L)X_t = \mu + \Theta(L)\varepsilon_t$$

où les polynômes Φ et Θ ont toutes leurs racines de module strictement supérieur à un et sont premiers entre eux, et ε_t est de variance σ_ε^2 .

L'idée générale est de faire un test de significativité standard (asymptotiquement normal) des coefficients ϕ_p et θ_q :

$$\begin{cases} H_0 : \phi_p = 0, \\ H_1 : \phi_p \neq 0, \end{cases}$$

ou

$$\begin{cases} H_0 : \theta_q = 0, \\ H_1 : \theta_q \neq 0. \end{cases}$$

Pour comparer une spécification $ARMA(p, q)$ avec une formulation $ARMA(p', q')$, il est utile de se placer dans le cas où l'un des modèles est un cas particulier de l'autre.

Il n'est pas possible de tester la nécessité d'augmenter simultanément les degrés des polynômes autorégressif et moyenne mobile.

▷ Les tests concernant les hypothèses sur ε .

L'objectif est de vérifier si la série des résidus estimés est cohérente avec l'hypothèse de bruit blanc (Gaussien).

– Deux types de tests :

1. Tests d'absence de corrélation des résidus,
2. Tests de normalité des résidus.

1) Tests d'absence de corrélation des résidus

- Le test de Box-Pierce permet d'identifier les processus de bruit blanc (i.e. les processus aléatoires de moyenne nulle, de variance constante et non autocorrélés)

$$H_0 : \rho_\varepsilon(1) = \rho_\varepsilon(2) = \dots = \rho_\varepsilon(h) = 0,$$

$$H_1 : \text{il existe } i \text{ tel que } \rho_\varepsilon(i) \neq 0.$$

- La statistique de Box-Pierce est donnée par :

$$Q_h = T \sum_{k=1}^h \hat{\rho}_\varepsilon^2(k).$$

Cette statistique est d'autant plus grande qu'il existe une corrélation non nulle entre les composantes du processus (ε_t) entre l'ordre 1 et l'ordre h .

Sous l'hypothèse de bruit blanc

$$Q_h \xrightarrow{L} \chi^2(h - p - q).$$

– Cette statistique peut être corrigée comme suit (statistique de Ljung-Box) :

$$Q_h = T(T + 2) \sum_{k=1}^h (T - h)^{-1} \hat{\rho}_{\varepsilon}^2(k).$$

II) Tests de normalité

On pourra utiliser par exemple le statistique de Jarque-Bera ou utiliser le test de Kolmogorof-Smirnov.

5.5 Choix d'un modèle

Plusieurs modèles (très différents) peuvent co-exister après les phases précédentes. Cependant, il reste un certain nombre de critères de choix :

- critère de parcimonie,
- critère de pouvoir prédictif,
- critère d'information.

▷ Critère de parcimonie

- Privilégier un modèle avec $p + q$ minimal.

▷ Critère de pouvoir prédictif

– Chercher le modèle qui donne les erreurs de prévisions les plus faibles. Il existe divers critères :

1. l'estimation $\hat{\sigma}_T^2$ de la variance,
2. le coefficient de détermination, $R^2 = 1 - \hat{\sigma}_T^2 / \hat{\gamma}_Y(0)$, $\hat{\gamma}_Y(0)$ étant la variance empirique de la série initiale différenciée à l'ordre d , avec $Y_t = (1 - L)^d X_t$,
3. le coefficient de détermination modifié

$$R^2 = 1 - \frac{\hat{\sigma}_T^2 / (T - p - q)}{\hat{\gamma}_Y(0) / (T - 1)},$$

4. la statistique de Fisher

$$F = \frac{(\hat{\gamma}_Y(0) - \hat{\sigma}_T^2) / (T - p - q)}{\hat{\sigma}_T^2 / (T - 1)}.$$

Le but est alors de minimiser (1) ou de maximiser (2), (3) ou (4).

▷ Critère d'information

Un critère d'information mesure la qualité de l'ajustement d'un modèle en assignant un coût d'information au nombre de paramètres à estimer (il existe toujours un arbitrage entre p et q en choisissant les ordres autorégressif et moyenne mobile).

– Le critère d'information le plus naïf est celui de la variance résiduelle :

$$\hat{\sigma}_{p,q}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2(p, q)$$

où $\hat{\varepsilon}_t(p, q)$ est le terme d'erreur ajusté d'un modèle $ARMA(p, q)$.

– L'inconvénient de ce critère est qu'il tend à favoriser les modèles avec beaucoup de paramètres (avec un risque de sur-ajustement) par rapport aux modèles avec peu de paramètres (principe de parcimonie).

- L'idée est alors d'introduire des critères de pénalité pour le nombre de paramètres : les critères d'information diffèrent dans la manière de pénaliser le nombre de paramètres estimés.
- L'objectif étant de pouvoir comparer des modèles qui ne sont pas nécessairement emboîtés les uns dans les autres, on se fonde sur une mesure de l'écart entre la vraie loi inconnue (du modèle implicite satisfaisant les données ou loi du processus générateur des données) et le modèle proposé (les modèles *ARMA* par lesquels on cherche à approximer le modèle implicite des données).
- Il s'agit alors de sélectionner la meilleure approximation, i.e. celle pour laquelle la distance au vrai modèle est la plus faible.
- La mesure retenue est habituellement la quantité d'information de Kullback entre les densités de probabilité de deux modèles.

DEFINITION :

Soit f la densité de probabilité inconnue des observations et $\{f(x), f \in \mathcal{F}_{p,q}\}$ la famille de densités correspondant au modèle $ARMA(p, q)$, l'écart entre la vraie loi et le modèle est mesuré par :

$$d_K(f_0, \mathcal{F}_{p,q}) = \min_{f \in \mathcal{F}_{p,q}} \int \log \left(\frac{f_0(x)}{f(x)} \right) f_0(x) dx.$$

Cette quantité est toujours positive et ne s'annule que si f_0 appartient à $\mathcal{F}_{p,q}$.

Cette quantité est inconnue et on cherche donc à minimiser un estimateur de cette mesure.

Plusieurs estimateurs de la quantité d'information ont été proposés :

1. Le critère d'Akaike :

$$AIC(p, q) = \log(\hat{\sigma}_{p,q}^2) + 2\frac{p+q}{T}.$$

Le modèle (c'est-à-dire la combinaison (p, q)) est choisi par minimisation du critère AIC.

Remarques :

1. Il peut exister plusieurs minima (i.e., combinaisons (p, q) qui donnent la même valeur pour le critère AIC) ;
2. Le critère AIC dépend de l'hypothèse de normalité ;
3. Le critère AIC tend à sur-paramétriser (à partir de simulations de Monte Carlo) ;

4. Le critère d'Akaike n'est pas convergent (la probabilité de sélectionner le bon modèle ne converge pas vers 1 lorsque le nombre d'observations tend vers l'infini).

2. Le critère de Schwartz :

$$BIC(p, q) = \log(\hat{\sigma}_{p,q}^2) + (p + q) \frac{\log T}{T}.$$

– Le modèle (c'est-à-dire la combinaison (p, q)) est choisi en minimisant le critère BIC .

– Le critère BIC tend à pénaliser plus les modèles avec un plus grand nombre de paramètres que le critère AIC , i.e. le critère BIC tend à choisir des modèles avec moins de paramètres.

– Le critère BIC est (fortement) convergent (la probabilité de sélectionner le bon modèle converge (presque sûrement) vers 1 lorsque le nombre d'observations tend vers l'infini).

3. Le critère d'Akaike corrigé :

$$AICC(p, q) = \log(\hat{\sigma}_{p,q}^2) + 2 \frac{p + q + 1}{T - p - q - 2}.$$

- L'objectif est de corriger le "biais de surparamétrisation" du critère AIC.
- Les critères AIC et $AICC$ sont asymptotiquement équivalents.
- Ce critère est très utile lorsque la taille de l'échantillon est faible.
- Le critère de pénalité est compris entre celui du critère AIC et celui du critère BIC .

4. Le critère d'Hannan-Quinn :

$$HN(p, q) = \log(\hat{\sigma}_{p,q}^2) + \frac{p + q}{T} c \log \log T$$

avec $c > 2$.

- Le critère HN est (fortement) convergent.
- Les critères AIC et BIC sont généralement préférés.

En pratique, on utilise souvent la règle suivante :

1. Fixer les bornes supérieures des ordres AR et MA : p_{max} et q_{max}
2. Estimer tous les modèles $ARMA(p, q)$ (où $0 \leq p \leq p_{max}$ et $0 \leq q \leq q_{max}$) en utilisant un échantillon commun.
3. Choisir le ou les meilleurs modèles en minimisant les critères d'information précédents.

– Remarques :

1. La détermination du modèle final ne se fonde pas nécessairement sur un de ces critères (le meilleur modèle in-sample n'est pas nécessairement le meilleur modèle out-of-sample) ;
2. Les critères d'information peuvent recommander des modèles différents.
3. Il peut être utile de considérer les modèles “adjacents” (i.e. ceux dont la valeur du critère d'information est proche du minimum).

o Sous

Pour estimer les paramètres d'un processus $ARIMA(p, d, q)$, on utilise la fonction

```
arima(x, order = c(0, 0, 0),  
      seasonal = list(order = c(0, 0, 0), period = NA),  
      xreg = NULL, include.mean = TRUE,  
      transform.pars = TRUE,  
      fixed = NULL, init = NULL,  
      method = c("CSS-ML", "ML", "CSS"),...)
```

Pour estimer un processus AR , on utilise la fonction

```
ar(x, aic = TRUE, order.max = NULL,  
    method=c("yule-walker", "burg", "ols", "mle", "yw"),  
    na.action, series, ...)
```

5.6 Exemples

▷ Simulation d'un $AR(1)$ avec $\varphi = 0,7$

```
ts.sim <- arima.sim(list(order = c(1,0,0), ar = 0.7), n = 500)
```

– Estimation automatique à l'aide du critère AIC

```
rec1 <- ar(ts.sim)
```

```
rec1
```

```
rec1$x.mean
```

```
sqrt(diag(rec1$asy.var.coef))
```

```
> rec1  
Call :  
ar(x = ts.sim)  
Coefficients :  
1  
0.6698  
Order selected 1 sigma^2 estimated as 0.914  
> rec1$x.mean  
[1] 0.07958582  
> sqrt(diag(rec1$asy.var.coef))  
[1] 0.03327462
```

– Estimation des coefficients d'un $AR(1)$

```
rec2 <-arima(ts.sim, order = c(1,0,0))
```

```
rec2
```

```
> rec2
```

```
Call :
```

```
arima(x = ts.sim, order = c(1, 0, 0))
```

```
Coefficients :
```

ar1	intercept
-----	-----------

0.6698	0.0739
--------	--------

s.e. 0.0331	0.1286
-------------	--------

```
sigma^2 estimated as 0.909 : log likelihood = -685.9, aic = 1377.8
```

– Estimation des coefficients d'un $AR(2)$

```
rec3 <-arima(ts.sim, order = c(2,0,0))
```

```
rec3
```

```
> rec3
```

```
Call :
```

```
arima(x = ts.sim, order = c(2, 0, 0))
```

```
Coefficients :
```

	ar1	ar2	intercept
	0.6518	0.0268	0.0735
s.e.	0.0447	0.0447	0.1320

```
sigma^2 estimated as 0.9083 : log likelihood = -685.72, aic = 1379.45
```

▷ Simulation d'un $ARIMA(1, 1, 1)$ avec $\varphi = -0,3$, $\theta = 0,5$

```
ts.sim <- arima.sim(list(order = c(1,1,1), ar = -0.3, ma=0.5), n = 500)
```

– Estimation des coefficients d'un $ARIMA(1, 1, 1)$

```
> rec1 <- arima(ts.sim, order = c(1,1,1))
```

```
> rec1
```

Call :

```
arima(x = ts.sim, order = c(1, 1, 1))
```

Coefficients :

ar1	ma1
-----	-----

-0.2226	0.4267
---------	--------

s.e. 0.1888	0.1748
-------------	--------

σ^2 estimated as 0.9484 : log likelihood = -696.25, aic = 1398.49

– Estimation des coefficients d'un *ARIMA*(0, 1, 1)

```
> rec2 <- arima(ts.sim, order = c(0,1,1))
```

```
> rec2
```

Call :

```
arima(x = ts.sim, order = c(0, 1, 1))
```

Coefficients :

ma1

0.2137

s.e. 0.0461

sigma^2 estimated as 0.9507 : log likelihood = -696.84, aic = 1397.68

– Estimation des coefficients d'un $ARIMA(1, 1, 0)$

```
> rec3 <-arima(ts.sim, order = c(1,1,0))
```

```
> rec3
```

Call :

```
arima(x = ts.sim, order = c(1, 1, 0))
```

Coefficients :

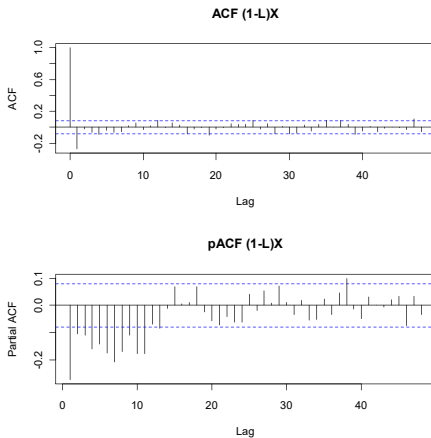
ar1

0.1866

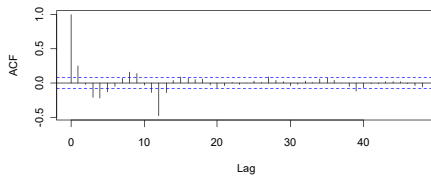
s.e. 0.0440

sigma^2 estimated as 0.9556 : log likelihood = -698.13, aic = 1400.27

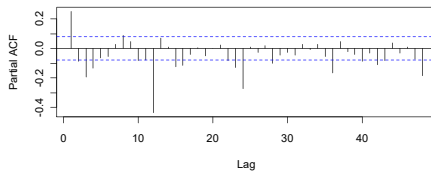
▷ Concentrations en CO_2 - Modélisation des résidus issus des moyennes mobiles



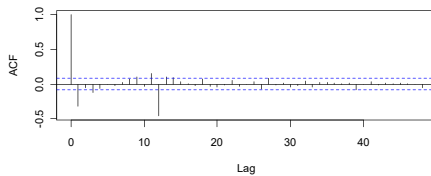
ACF (1-L¹²)X



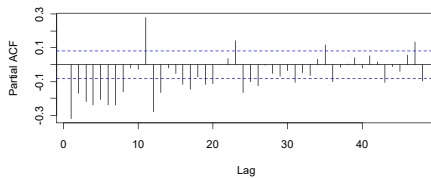
pACF (1-L¹²)X



ACF $(1-L)(1-L^{12})X$



pACF $(1-L)(1-L^{12})X$



– Estimation des coefficients d'un $SARIMA(1, 1, 6)(0, 1, 1)$

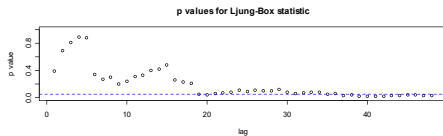
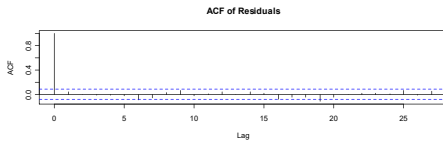
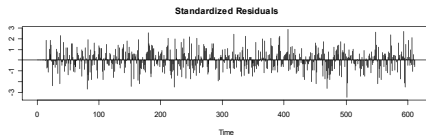
```
arima(x = resi, order = c(1, 1, 6),  
seasonal = list(order = c(0, 1, 1), period = 12))
```

Coefficients :

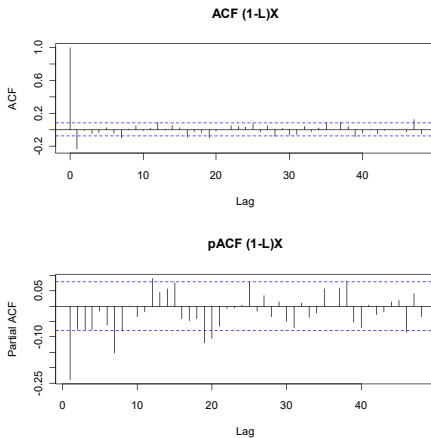
	ar1	ma1	ma2	ma3
	0.5621	-1.4565	0.2992	0.0003
s.e.	0.2622	0.2829	0.2633	0.0929

	ma4	ma5	ma6	sma1
	0.0075	0.0949	0.0551	-0.8848
s.e.	0.0909	0.1556	0.2139	0.0211

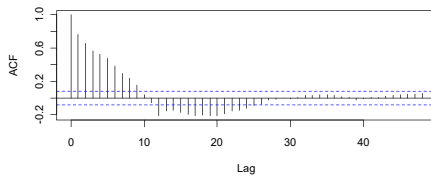
sigma² estimated as 0.05189 : log likelihood = 18.18, aic = -18.36



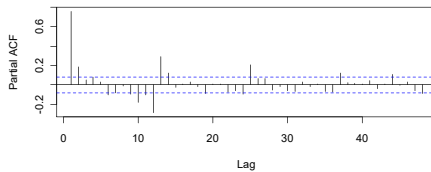
▷ Concentrations en CO_2 - Modélisation des résidus Buys Ballot



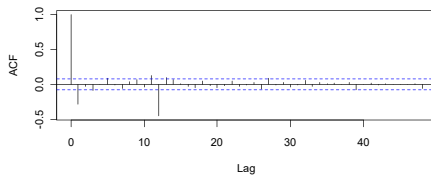
ACF (1-L¹²)X



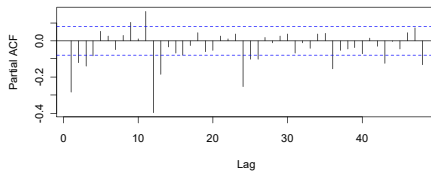
pACF (1-L¹²)X



ACF (1-L)(1-L¹²)X



pACF (1-L)(1-L¹²)X



– Estimation des coefficients d'un $SARIMA(0, 1, 7)(1, 0, 0)$

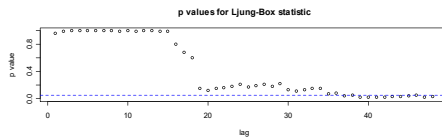
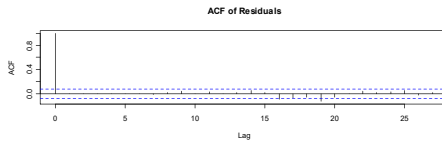
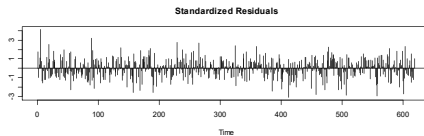
```
arima(x = resi, order = c(0, 1, 7),  
      seasonal = list(order = c(1, 0, 0), period = 12))
```

Coefficients :

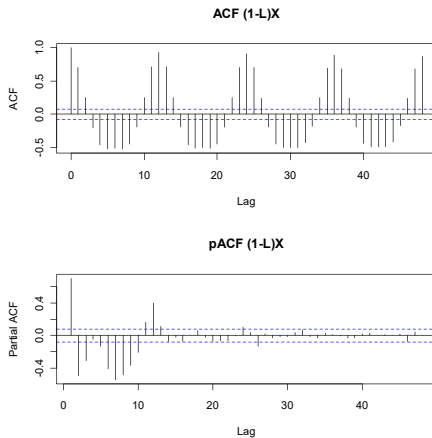
	ma1	ma2	ma3	ma4
	-0.2928	-0.0361	-0.0867	-0.0370
s.e.	0.0405	0.0420	0.0422	0.0414

	ma5	ma6	ma7	sar1
	0.0183	-0.0704	-0.0869	0.1039
s.e.	0.0441	0.0450	0.0415	0.0416

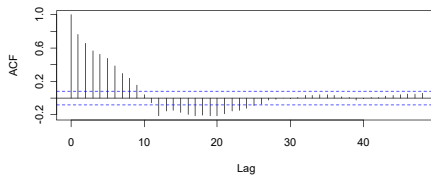
σ^2 estimated as 0.0848 : log likelihood = -114.64, aic = 247.29



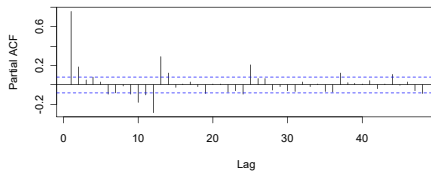
▷ Concentrations en CO_2 - Série - tendance par la méthode de Buys Ballot



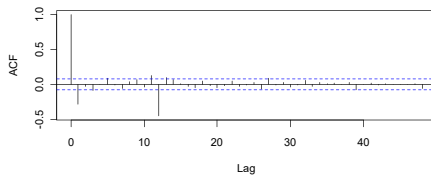
ACF (1-L¹²)X



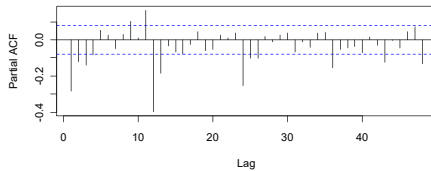
pACF (1-L¹²)X



ACF (1-L)(1-L¹²)X



pACF (1-L)(1-L¹²)X



– Estimation des coefficients d'un $SARIMA(0, 1, 1)(0, 1, 1)$

```
arima(x = resi, order = c(0, 1, 1),  
      seasonal = list(order = c(0, 1, 1), period = 12))
```

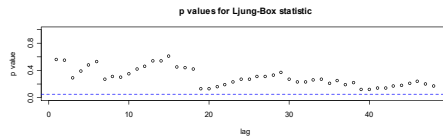
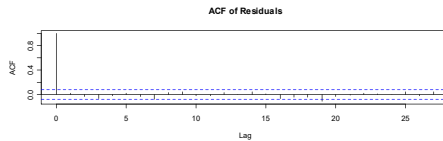
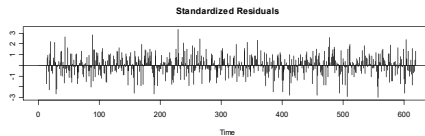
Coefficients :

ma1	sma1
-----	------

-0.3809	-0.8765
---------	---------

s.e. 0.0419	0.0207
-------------	--------

σ^2 estimated as 0.08836 : log likelihood = -133.55, aic = 273.1



6 Prévisions

6.1 Le lissage exponentiel

Il s'agit d'une méthodologie permettant des calculs de prévisions ne reposant sur aucun (ou presque) modèle.

1. Lissage exponentiel simple

On dispose d'observations d'une série temporelle : $X_1, \dots, X_T$. On cherche à prédire les valeurs futures. Pour tout coefficient $\alpha \in (0, 1)$, appelé constante de lissage, on définit la série lissée

$$\begin{aligned}\hat{X}_t &= (1 - \alpha)(X_t + \alpha X_{t-1} + \alpha^2 X_{t-2} + \dots + \alpha^{t-1} X_1) \\ &= (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j X_{t-j}, \quad t = 1, \dots, T.\end{aligned}$$

REMARQUES :

– Les observations les plus récentes ont l'influence la plus grande (décroissance exponentielle) :

- si α est proche de 0, il y a une faible influence du passé éloigné (prévision souple)

- si α est proche de 1, il y a une forte influence du passé éloigné (prévision rigide)

– Formule indépendante de h . Notation : $\hat{X}_T$.

▷ Formules de mise à jour

On a directement

$$(1) : \hat{X}_T = \alpha \hat{X}_{T-1} + (1 - \alpha) X_T$$

$$(2) : \hat{X}_T = \hat{X}_{T-1} + (1 - \alpha)(X_T - \hat{X}_{T-1})$$

Formule (1) : moyenne pondérée entre la dernière observation et la prévision faite en $T - 1$.

Formule (2) : mécanisme de correction d'erreur. Nécessité d'initialiser les algorithmes de mise à jour (par exemple $\hat{X}_1 = X_1$).

▷ Interprétation

Supposons que la tendance soit à peu près constante sur la période d'observation :

$$X_t = a + u_t, \quad (u_t) \sim BB(0, \sigma^2), \quad t = 1, \dots, T.$$

On peut estimer a par moindres carrés en considérant le programme

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

Plus généralement, considérons le programme (moindres carrés pondérés)

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

On obtient facilement la solution

$$\hat{a} = \frac{(1 - \alpha)}{(1 - \alpha^T)} \sum_{j=0}^{t-1} \alpha^j X_{t-j}.$$

Donc pour T assez grand, on a $\hat{a} \simeq \hat{X}_T$. $\hat{X}_T$ s'interprète donc comme la constante qui approxime le mieux la série "au voisinage" de T . Il s'agit donc d'une méthode utilisable lorsque la série est approximativement constante au voisinage de T (localement constante)

▷ Choix de la constant de lissage

- Critères subjectifs : plus ou moins grande souplesse. Choix conseillé par Brown (inventeur de la méthode) : $\alpha = 0,7$.

- Critère objectif : on choisit la constante qui minimise la somme des carrés des erreurs de prévision aux dates $1, \dots, T - 1$:

$$\sum_{t=1}^{T-1} (X_{t+1} - \hat{X}_t)^2 = \sum_{t=1}^{T-1} (X_{t+1} - (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j X_{t-j})^2.$$

2. Lissage exponentiel double

Cette méthode est plus générale, et plus adaptée aux séries temporelles présentant une tendance non constante.

Supposons que l'on ajuste la série, au voisinage de T , par une droite :

$$X_t = a_1 + (t - T)a_2 + u_t, \quad (u_t) \sim BB(0, \sigma^2), \quad t = 1, \dots, T.$$

De manière analogue, on considère le programme :

$$\min_{a_1, a_2} \sum_{j=0}^{T-1} \alpha^j (X_{T-j} - a_1 + a_2 j)^2$$

pour une prévision à horizon h :

$$\hat{X}_T(h) = \hat{a}_1(T) + h\hat{a}_2(T),$$

où $\hat{a}_1(T)$ et $\hat{a}_2(T)$ sont les solutions du programme. La résolution du programme donne (pour T grand) :

$$\begin{aligned}\hat{a}_1(T) &= 2S_1(T) - S_2(T) \\ \hat{a}_2(T) &= \frac{1-\alpha}{\alpha}[S_1(T) - S_2(T)]\end{aligned}$$

avec

$$\begin{aligned}S_1(T) &= (1-\alpha) \sum_{j=0}^{T-1} \alpha^j X_{T-j} \quad (\text{série lissée}) \\ S_2(T) &= (1-\alpha) \sum_{j=0}^{T-1} \alpha^j S_1(T-j) \quad (\text{série doublement lissée}).\end{aligned}$$

Ces formules ne sont généralement pas utilisées en pratique.

▷ Formules de mise à jour

On obtient (après quelques calculs) :

$$\hat{a}_1(T) = \hat{a}_1(T-1) + \hat{a}_2(T-1) + (1 - \alpha^2)[X_T - \hat{X}_{T-1}(1)]$$

$$\hat{a}_2(T) = \hat{a}_2(T-1) + (1 - \alpha)^2[X_T - \hat{X}_{T-1}(1)]$$

Pour utiliser ces formules il faut initialiser l'algorithme ; par exemple $\hat{a}_1(2) = X_2$ et $\hat{a}_2(2) = X_2 - X_1$.

▷ Interprétation des formules de mise à jour

Supposons que l'on ait fait une prévision parfaite à la date $t - 1$: $X_T = \hat{X}_{T-1}(1)$.

On aurait

$$\hat{a}_1(T) = \hat{a}_1(T-1) + \hat{a}_2(T-1)$$

$$\hat{a}_2(T) = \hat{a}_2(T-1)$$

Les droites de prévision sont donc les mêmes aux dates $T - 1$ et T . Dans les formules de mise à jour, il y a en plus des termes proportionnels à la dernière erreur de prévision.

▷ Autres formules de mise à jour

On a

$$\hat{X}_{T-1}(1) = \hat{a}_1(T-1) + \hat{a}_2(T-1).$$

Donc

$$\begin{aligned}\hat{a}_1(T) &= (1 - \alpha^2)X_T + \alpha^2\hat{X}_{T-1}(1) \\ \hat{a}_2(T) &= \frac{(1 - \alpha)^2}{(1 - \alpha^2)}(\hat{a}_1(T) - \hat{a}_1(T-1)) + \frac{2\alpha}{1 + \alpha}\hat{a}_2(T-1).\end{aligned}$$

3. Méthode de Holt-Winters non saisonnière

Comme pour le lissage exponentiel double, on suppose que la série s'ajuste, au voisinage de T , par une droite

$$X_t = a_1 + (t - T)a_2 + u_t.$$

On introduit 2 constantes de lissage $0 < \beta < 1$, $0 < \gamma < 1$. On choisit les formules de mise à jour :

$$\hat{a}_1(T) = (1 - \beta)X_T + \beta(\hat{a}_1(T - 1) + \hat{a}_2(T - 1))$$

$$\hat{a}_2(T) = (1 - \gamma)(\hat{a}_1(T) - \hat{a}_1(T - 1)) + \gamma\hat{a}_2(T - 1)$$

pour une prévision à horizon h :

$$\hat{X}_T(h) = \hat{a}_1(T) + h\hat{a}_2(T).$$

Cette méthode est plus flexible que le lissage exponentiel double, mais il faut choisir deux constantes de lissage.

On peut minimiser en β et γ le critère

$$\sum_{t=1}^{T-1} (X_{t+1} - \hat{X}_t(1))^2.$$

Pour initialiser les constantes, on peut prendre par exemple $\hat{a}_1(2) = X_2$ et $\hat{a}_2(2) = X_2 - X_1$.

4. Méthode de Holt-Winters saisonnière

La composante saisonnière est introduite de manière additive. On suppose qu'au voisinage de T , la série peut être approchée par

$$X_t = a_1 + (t - T)a_2 + S_t + u_t, \quad (u_t) \sim BB(0, \sigma^2), \quad t = 1, \dots, T,$$

où S_t est un facteur saisonnier prenant s valeurs différentes (inconnues) : $S_1, \dots, S_s$.

On impose, pour des raisons d'identifiabilité, $S_1 + \dots + S_s = 0$.

▷ Formules de mise à jour

$$\hat{a}_1(T) = (1 - \beta)(X_T - \hat{S}_{T-s}) + \beta(\hat{a}_1(T-1) + \hat{a}_2(T-1))$$

$$\hat{a}_2(T) = (1 - \gamma)(\hat{a}_1(T) - \hat{a}_1(T-1)) + \gamma\hat{a}_2(T-1)$$

$$\hat{S}_T = (1 - \delta)(X_T - \hat{a}_1(T)) + \delta\hat{S}_{T-s}$$

où $0 < \beta < 1$, $0 < \gamma < 1$, $0 < \delta < 1$ sont des constantes. La prévision est :

$$\begin{aligned}\hat{X}_T(h) &= \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-s}, & \text{pour } 1 \leq h \leq s \\ &= \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-2s}, & \text{pour } s < h \leq 2s, \\ &= \text{etc...}\end{aligned}$$

▷ Problèmes pratiques

Choix des 3 constantes de lissage : l'estimation par balayage peut être assez longue

Initialisation des formules de mise à jour : prenons par exemple $s = 4$; on commence par fixer :

- $\hat{a}_1(1) = \hat{a}_1(2) = \hat{a}_1(3) = \hat{a}_1(4) = \text{moyenne de la première année}$;
- $\hat{a}_2(1) = \hat{a}_2(2) = \hat{a}_2(3) = \hat{a}_2(4) = 0$.

Puis on initialise les coefficients saisonniers :

$$\hat{S}_4 = X_4 - \hat{a}_1(4); \hat{S}_3 = X_3 - \hat{a}_1(3) \text{ etc } \dots$$

Avantages des méthodes de lissage :

- mise en oeuvre simple : les équations de prévision sont faciles à interpréter et les calculs rapides,
- méthodes ne reposant sur aucun modèle,
- méthodes parfois aussi performantes que des méthodes plus sophistiquées,
- méthodes "robustes" : elles permettent de travailler sur des séries courtes ou changeant de structure.

Inconvénients des méthodes de lissage :

- méthodes adaptées à certains types de séries seulement,
- arbitraire (choix des constantes de lissage),
- optimalité des prévisions obtenues ?

6.2 Prédiction dans un modèle ARIMA

1. Introduction

On dispose des données $x_1, \dots, x_T$ réalisations des variables aléatoires $X_1, \dots, X_T$. Ces observations sont compatibles avec un modèle $ARIMA(p, d, q)$:

$$\Phi(L)\Delta^d X_t = \mu + \Theta(L)\varepsilon_t$$

où

$$\begin{aligned}\Phi(L) &= 1 - \phi_1 L - \dots - \phi_p L^p \\ \Theta(L) &= 1 - \theta_1 L - \dots - \theta_q L^q\end{aligned}$$

avec $d^\circ(\Phi) = p$, $d^\circ(\Theta) = q$, $\mathbb{V}ar(\varepsilon_t) = \sigma_\varepsilon^2$, et où les hypothèses sur les polynômes de retard et (ε_t) sont satisfaites.

Les paramètres du modèle ont été estimés et on cherche à prédire la valeur de $X_{T+1}, \dots, X_{T+h}$ sachant un certain ensemble d'information à la date T .

2. Prévisions dans un $AR(p)$

Soit le processus stochastique (X_t) autorégressif d'ordre p satisfaisant l'équation :

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

où les ϕ_j , $j = 1, \dots, p$, sont des nombres réels (avec $\phi_p \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 . Les racines du polynôme :

$$\Phi(u) = 1 - \phi_1 u - \dots - \phi_p u^p$$

sont de module strictement supérieur à 1, et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

DEFINITION :

La prévision linéaire optimale de X_{T+h} sachant $\underline{X_T}$ est définie par :

$${}_T X_{T+h}^* = EL(X_{T+h} | 1, \underline{X_T}).$$

En particulier,

$$\begin{aligned} {}_T X_{T+1}^* &= \mu + \phi_1 X_T + \dots + \phi_p X_{T+1-p} \\ {}_T X_{T+2}^* &= \mu + \phi_1 {}_T X_{T+1}^* + \phi_2 X_T + \dots + \phi_p X_{T+2-p} \end{aligned}$$

et

$${}_T X_{T+h}^* = \mu + \sum_{j=1}^{h-1} \phi_j {}_T X_{T+h-j}^* + \sum_{j=h}^p \phi_j X_{T+h-j}$$

On obtient ainsi une équation de récurrence :

$$\Phi(L) {}_T X_{T+h}^* = \mu \Leftrightarrow \Phi(L) {}_T Y_{T+h}^* = 0$$

avec ${}_T Y_{T+h}^* = {}_T X_{T+h}^* - \mu / \Phi(1)$. ${}_T Y_{T+h}^*$ est la solution de l'équation de récurrence de polynôme caractéristique $z^p \Phi(z^{-1})$. On en déduit que ${}_T Y_{T+h}^*$ est de la forme :

$${}_T Y_{T+h}^* = \sum_{i=1}^r \sum_{j=1}^{m_i-1} \alpha_{ij} \lambda_i^h h^j$$

où les $1/\lambda_i$ sont les racines distinctes de $\Phi(z)$ et sont de multiplicité m_i , et les α_{ij} sont obtenus à partir des p conditions initiales (observations ou prévisions).

DEFINITION :

L'erreur de prédiction à l'horizon h est définie par :

$$Te_{T+h} = X_{T+h} - {}^T X_{T+h}^* = \sum_{j=0}^{h-1} a_j \varepsilon_{T+h-j}$$

où les a_j sont les coefficients de la représentation moyenne mobile infinie du processus :

$$X_t = m + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}.$$

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

$$\text{Var}(Te_{T+h}) = \sigma_{\varepsilon}^2 \sum_{j=0}^{h-1} a_j^2.$$

3. Prévisions dans un $MA(q)$

Soit le processus stochastique (X_t) moyenne mobile d'ordre q satisfaisant l'équation :

$$X_t = \mu + \Theta(L)\varepsilon_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q},$$

où les θ_j , $j = 1, \dots, q$, sont des nombres réels (avec $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 . Les racines du polynôme :

$$\Theta(u) = 1 - \theta_1 u - \dots - \theta_q u^q$$

sont de module strictement supérieur à 1, alors $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation du processus.

DEFINITION :

La prévision linéaire optimale de X_{T+h} sachant $\underline{X_T}$ est définie par :

$${}_T X_{T+h}^* = EL(X_{T+h} | \underline{X_T}) = EL(X_{T+h} | \underline{\varepsilon_T}).$$

Cette prévision optimale est donnée par :

$$\begin{aligned} {}_T X_{T+h}^* &= \mu \text{ si } h > q \\ {}_T X_{T+h}^* &= \mu - \sum_{j=h}^q \theta_j \varepsilon_{T+h-j} \text{ si } h \leq q. \end{aligned}$$

Cette forme est exacte mais n'est pas utilisable en pratique car les ε_{t-k} ne sont pas observés pour $k \geq 0$.

Pour déterminer de manière pratique les prédictions linéaires d'un processus $MA(q)$, on peut utiliser la représentation autorégressive infinie :

$$\Theta^{-1}(L)X_t = m + \varepsilon_t$$

avec $m = \mu/\Theta(1)$. On obtient :

$$X_T = m - \sum_{k=1}^{\infty} a_k X_{T-k} + \varepsilon_T \quad X_{T+h} = m - \sum_{k=1}^{\infty} a_k X_{T+h-k} + \varepsilon_{T+h}$$

et on en déduit donc :

$$\begin{aligned} {}_T X_{T+1}^* &= m - \sum_{k=1}^{\infty} a_k X_{T+1-k} \\ {}_T X_{T+h}^* &= m - \sum_{k=1}^{h-1} a_k {}_T X_{T+h-k}^* - \sum_{k=h}^{\infty} a_k X_{T+h-k}. \end{aligned}$$

En pratique, on n'observe pas les X_t pour $t < 0$. On obtient une prévision approchée en tronquant :

$${}_T \hat{X}_{T+h}^* = m - \sum_{k=1}^{h-1} a_k {}_T \hat{X}_{T+h-k}^* - \sum_{k=h}^{T+h} a_k X_{T+h-k}.$$

DEFINITION :

L'erreur de prédiction à l'horizon h est définie par :

$${}^T e_{T+h} = X_{T+h} - {}_T X_{T+h}^* = \sum_{j=0}^{\min(h,q)} \theta_j \varepsilon_{T+h-j}.$$

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

$$\mathbb{V}ar(Te_{T+h}) = \sigma_{\varepsilon}^2 \sum_{j=0}^{\min(h,q)} \theta_j^2.$$

4. Prévisions d'un $ARMA(p, q)$

Soit le processus stationnaire, (X_t) défini par :

$$\begin{aligned}X_t &= \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \\&\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \Phi(L)X_t &= \mu + \Theta(L)\varepsilon_t\end{aligned}$$

où les ϕ_j , $j = 1, \dots, p$, et les θ_j , $j = 1, \dots, q$, sont des nombres réels (avec $\phi_p \neq 0$ et $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 . Φ et Θ sont des polynômes de retard dont toutes les racines sont de module strictement supérieur à un et premiers entre eux et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est donc l'innovation du processus.

DEFINITION :

La prévision linéaire optimale de X_{T+h} sachant $\underline{X_T}$ est définie par :

$${}_T X_{T+h}^* = EL(X_{T+h} | \underline{X_T}).$$

Cette prévision optimale est donnée par :

$${}_T X_{T+h}^* = \mu + \sum_{j=1}^{h-1} \phi_j {}_T X_{T+h-j}^* + \sum_{j=h}^p \phi_j X_{T+h-j} - \mathbb{I}_{h \leq q} \sum_{j=h}^q \theta_j \varepsilon_{T+h-j}.$$

Si $h > q$, on obtient une équation de récurrence comme dans le cas d'un processus $AR(p)$

Si $h \leq q$, la prévision linéaire optimale est exacte mais n'est pas exploitable. Dans cette perspective, on peut utiliser l'écriture $AR(\infty)$ et obtenir ainsi des prévisions approchées en tronquant.

DEFINITION :

L'erreur de prédiction à l'horizon h est définie par :

$$T e_{T+h} = X_{T+h} - {}_T X_{T+h}^* = \sum_{j=0}^{h-1} \delta_j \varepsilon_{T+h-j}$$

où les δ_j sont les coefficients de la représentation moyenne mobile infinie du processus :

$$X_t = m + \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j}.$$

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

$$\mathbb{V}ar(Te_{T+h}) = \sigma_{\varepsilon}^2 \sum_{j=0}^{h-1} \delta_j^2.$$

5. Prévisions d'un $ARIMA(p, d, q)$

Soit le processus stochastique $(X_t)_{t \geq -p-d}$ défini par :

$$\begin{aligned}(1-L)^d X_t &= \mu + \phi_1(1-L)^d X_{t-1} + \dots + \phi_p(1-L)^d X_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \\ \Phi(L)(1-L)^d X_t &= \mu + \Theta(L)\varepsilon_t\end{aligned}$$

où les ϕ_j , $j = 1, \dots, p$, et les θ_j , $j = 1, \dots, q$, sont des nombres réels (avec $\phi_p \neq 0$ et $\theta_q \neq 0$) et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ_ε^2 , et les conditions initiales

$$Z_{-1} = \{X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$$

sont telles que :

$$Cov(\varepsilon_t, Z_{-1}) = 0 \quad \forall t \geq 0.$$

On suppose que toutes les racines des polynômes Φ et Θ sont de module strictement supérieur à 1, et on obtient ainsi la représentation canonique minimale. On peut toujours supposer que $\mu = 0$ en considérant le processus

$$\Psi(L)Y_t = \Theta(L)\varepsilon_t$$

avec $\Psi(L) = \Phi(L)(1 - L)^d$ et $Y_t = X_t - m$.

DEFINITION :

La prévision linéaire optimale de Y_{T+h} sachant $\underline{Y_T}, Z_{-1}$ est définie par :

$${}_TY_{T+h}^* = EL(Y_{T+h}|\underline{Y_T}, Z_{-1}).$$

On peut utiliser l'approximation autorégressive :

$$Y_t = - \sum_{j=1}^t a_j Y_{t-j} + H'(t)Z_{-1} + \varepsilon_t$$

avec $H(t) \in \mathbb{R}^{p+d+q}$ et $\lim_{t \rightarrow \infty} \|H(t)\| = 0$. On obtient :

$$Y_{T+h} = - \sum_{j=1}^{T+h} a_j Y_{T+h-j} + H'(T+h)Z_{-1} + \varepsilon_{T+h}$$

$${}_T Y_{T+h}^* = - \sum_{j=1}^{T+h} a_j {}_T Y_{T+h-j}^* + H'(T+h)Z$$

avec ${}_T Y_{T+h-j}^* = Y_{T+h-j}$ si $j \leq h$. Une approximation est donnée par

$${}_T \hat{Y}_{T+h}^* = - \sum_{j=1}^{T+h} a_j {}_T \hat{Y}_{T+h-j}^*.$$

Pour $h > q$, on obtient une équation de récurrence de polynôme caractéristique $z^{p+d}\Psi(z^{-1})$. D'où

$${}_T Y_{T+h}^* = \sum_{i=1}^r \sum_{j=1}^{m_i-1} \alpha_{ij} \lambda_i^h h^j.$$

○ Sous

Pour prédire les valeurs d'un processus $ARIMA(p, d, q)$, on utilise la fonction

```
predict(object, newdata, se.fit = FALSE, scale = NULL, df = Inf,  
interval = c("none", "confidence", "prediction"),  
level = 0.95, type = c("response", "terms"),  
terms = NULL, na.action = na.pass,  
pred.var = res.var/weights, weights = 1, ...)
```

DONNEES REELLES :

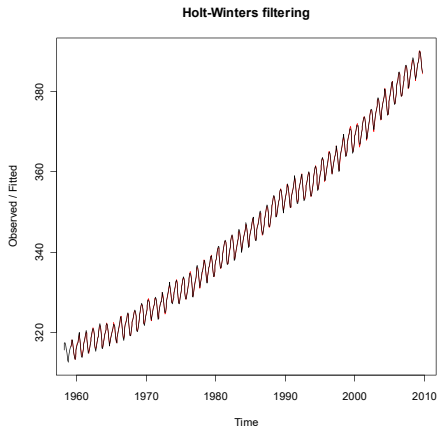
▷ Concentrations en CO₂

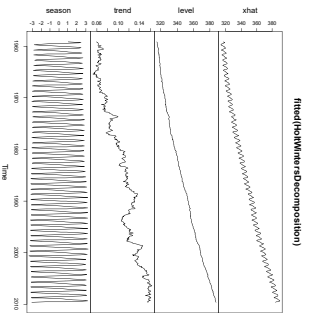
On a retiré les 12 dernières données et on a estimé le modèle. On compare ensuite les valeurs observées des 12 derniers mois aux valeurs données par les différents modèles.

On considère les prévisions obtenues par :

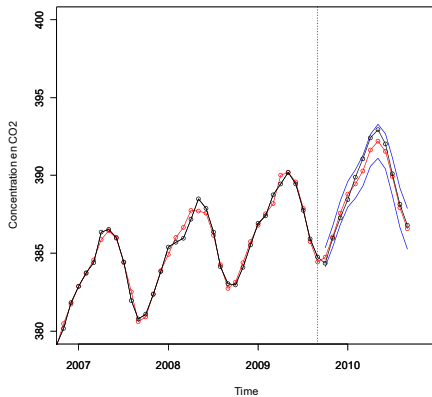
- la méthode de Holt-Winters saisonnière,
- la méthode basée sur un modèle $SARIMA(0, 1, 1)(0, 1, 1)$ estimé sur la série : données initiales - tendance issue de la méthode de Buys-Ballot.

▷ Prédiction par la méthode de Holt-Winters saisonnière

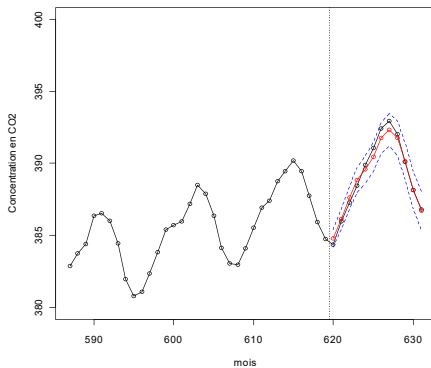




Holt-Winters filtering



▷ Prédiction par la méthode *SARIMA*



7 Fonctions R pour l'analyse des séries temporelles

(extrait de <http://cran.r-project.org/doc/contrib/Ricci-refcard-ts.pdf>)

Here are some helpful R functions for time series analysis. They belong from stats, tseries, ast and lmtest packages and grouped by their goal.

INPUT

- cycle() : gives the positions in the cycle of each observation (stats)
- deltat() : returns the time interval between observations (stats)
- end() : extracts and encodes the times the last observation were taken (stats)
- frequency() : returns the number of samples per unit time (stats)
- read.ts() : reads a time series file (tseries)
- start() : extracts and encodes the times the first observation were taken (stats)
- time() : creates the vector of times at which a time series was sampled (stats)
- ts() : creates time-series objects (stats)

- `window()` : is a generic function which extracts the subset of the object 'x' observed between the times 'start' and 'end'. If a frequency is specified, the series is then re-sampled at the new frequency (stats)

TS DECOMPOSITION

- `decompose()` : decomposes a time series into seasonal, trend and irregular components using moving averages. Deals with additive or multiplicative seasonal component (stats)
- `filter()` : linear filtering on a time series (stats)
- `HoltWinters()` : computes Holt-Winters Filtering of a given time series (stats)
- `sfilter()` : removes seasonal fluctuation using a simple moving average (ast)
- `spectrum()` : estimates the spectral density of a time series (stats)
- `stl()` : decomposes a time series into seasonal, trend and irregular components using 'loess' (stats)
- `tsr()` : decomposes a time series into trend, seasonal and irregular. Deals with additive and multiplicative components (ast)

TESTS

- `adf.test()` : computes the Augmented Dickey-Fuller test for the null that 'x' has a unit root (`tseries`)
- `Box.test()` : computes the Box-Pierce or Ljung-Box test statistic for examining the null hypothesis of independence in a given time series (`stats`)
- `bds.test()` : computes and prints the BDS test statistic for the null that 'x' is a series of i.i.d. random variables (`tseries`)
- `bptest()` : performs the Breusch-Pagan test for heteroskedasticity of residuals (`lmtest`)
- `dwtest()` : performs the Durbin-Watson test for autocorrelation of residuals (`lmtest`)
- `jarque.bera.test()` : Jarque-Bera test for normality (`tseries`)
- `kpss.test()` : computes KPSS test for stationarity (`tseries`)
- `shapiro.test()` : Shapiro-Wilk Normality Test (`stats`)

STOCHASTIC MODELS

- `ar()` : fits an autoregressive time series model to the data, by default selecting the complexity by AIC (stats)
- `arma()` : fits an ARIMA model to a univariate time series (stats)
- `arma.sim()` : simulate from an ARIMA model (stats)
- `arma()` : fits an ARMA model to a univariate time series by conditional least squares (tseries)
- `garch()` : fits a Generalized Autoregressive Conditional Heteroscedastic GARCH(p, q) time series model to the data by computing the maximum-likelihood estimates of the conditionally normal model (tseries)

GRAPHICS

- `lag.plot` : plots time series against lagged versions of themselves. Helps visualizing "auto-dependence" even when auto-correlations vanish (stats)
- `monthplot()` : plots a seasonal (or other) subseries of a time series (stats)
- `plot.ts()` : plotting time-series objects (stats)

- `seaplot()` : plotting seasonal sub-series or profile (ast)
- `seqplot.ts()` : plots a two time series on the same plot frame (tseries)
- `tsdiag()` : a generic function to plot time-series diagnostics (stats)
- `ts.plot()` : plots several time series on a common plot. Unlike 'plot.ts' the series can have a different time bases, but they should have the same frequency (stats)

MISCELLANEOUS

- `acf()`, `pacf()`, `ccf()` : the function 'acf' computes (and by default plots) estimates of the autocovariance or autocorrelation function. Function 'pacf' is the function used for the partial autocorrelations. Function 'ccf' computes the cross-correlation or cross-covariance of two univariate series (stats)
- `diff.ts()` : returns suitably lagged and iterated differences (stats)
- `lag()` : computes a lagged version of a time series, shifting the time base back by a given number of observations (stats)

Table des matières

1	Introduction aux séries temporelles : définitions et exemples	4
1.1	Définitions	4
1.2	Problèmes statistiques	13
1.3	Modélisation et modèle classique de décomposition	14
1.4	Processus stationnaires	17

2	Trend et saisonnalités	32
2.1	Méthodes par régression	33
2.2	Méthodes par moyenne mobiles	41
2.3	Exemples	61
3	Processus linéaires	69
3.1	L'espace $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$: définition et propriétés	69
3.2	Régression linéaire ou affine théorique sur un nombre fini de retards .	75
3.3	Régression linéaire théorique sur un nombre infini de retards	81

3.4	Densité spectrale et autocorrélations inverses	85
3.5	Estimation de la moyenne et des autocovariances pour des processus linéaires	91

4 Processus ARMA et ARIMA 116

4.1	Polynômes retard et avance	116
4.2	Processus $MA(q)$	126
4.3	Processus $AR(p)$	132
4.4	Processus $ARMA(p, q)$	139
4.5	Processus $ARIMA(p, d, q)$	145
4.6	Processus $SARIMA_s[(p, d, q), (P, D, Q)]$	148

5	Estimation des modèles ARMA et ARIMA	150
5.1	Introduction	150
5.2	Phase d'identification a priori	154
5.3	Phase d'estimation	170
5.4	Phase de vérification	181
5.5	Choix d'un modèle	186
5.6	Exemples	197

6	Prévisions	219
6.1	Le lissage exponentiel	219
6.2	Prédiction dans un modèle ARIMA	233
7	Fonctions R pour l'analyse des séries temporelles	253