

Introduction aux séries temporelles

Julien JACQUES

Table des matières

1	Introduction et premières définitions	4
1.1	Tendances et composantes saisonnières	6
1.2	Indices descriptifs d'une série temporelle	7
1.2.1	Indices de tendances centrales	7
1.2.2	Indices de dispersion	7
1.2.3	Indices de dépendance	7
1.3	Mise en oeuvre sous R	9
2	Lissages exponentiels	10
2.1	Lissage exponentiel simple	10
2.2	Lissage exponentiel double	11
2.3	Méthode de Holt-Winters	11
2.3.1	Méthode non saisonnière	11
2.3.2	Méthode saisonnière additive	12
2.3.3	Méthode saisonnière multiplicative	12
2.4	Mise en oeuvre sous R	13
3	Estimation et élimination de la tendance et de la saisonnalité	18
3.1	Bruit blanc	18
3.2	Processus stationnaire	18
3.3	Une estimation paramétrique de la tendance (<i>trend</i>)	18
3.4	Estimation non paramétrique : moyenne mobile	19
3.5	Élimination de la tendance et de la saisonnalité par la méthode des différences	22
3.6	Test sur la série résiduelle	22
3.6.1	Comment tester si on est en présence d'un bruit blanc ?	24
3.7	Mise en oeuvre sous R	24
4	Modélisation des séries stationnaires	25
4.1	Auto-corrélation partielle	25
4.2	Les processus auto-régressifs AR_p	25
4.2.1	Exercice : cas particulier de l' AR_1	26
4.2.2	Illustrations	27
4.3	Les processus en moyenne mobile MA_q	27
4.3.1	Exercice : cas particulier du MA_1	27
4.3.2	Illustrations	27
4.4	Les processus mixtes $ARMA_{p,q}$	35
4.4.1	Exercice : le processus $ARMA_{1,1}$	35
4.4.2	Illustrations	35
4.5	Récapitulatif des propriétés des processus MA_q , AR_p et $ARMA_{p,q}$	37
4.6	Estimation, choix de modèle et prévision	37
4.6.1	Estimation	37
4.6.2	Choix de modèle	37
4.6.3	Prévision	38
5	Processus non stationnaire : ARIMA et SARIMA	39
5.1	Exemple	39
5.2	Les processus ARIMA et SARIMA	39
5.3	Mise en oeuvre sous R : processus ARMA, ARIMA, SARIMA	40

6	Processus ARCH et GARCH	41
6.1	Définitions	42
6.2	Quelques rappels de probabilités	42
6.3	Propriétés des processus ARCH	42
6.4	Processus GARCH et propriétés	43
6.5	Mise en oeuvre sous R	43

Ce document est uniquement un support de cours, et ne constitue pas à lui seul un cours sur les séries temporelles. Les différentes démonstrations, applications et exercices nécessaires à la compréhension des notions présentées dans ce document n'y figurent pas (ou alors sous la forme d'exercice), mais seront développés en cours. A la fin de chaque chapitre figurent les commandes R nécessaires à l'application des notions du chapitre.

1 Introduction et premières définitions

Une **série temporelle** (ou série chronologique) à temps discret est une suite réelle finie $(x_t)_{1 \leq t \leq n}$, où t représente le temps (en minute, jour, année...).

Voici quelques exemples de séries temporelles :

- Ex. 1 : Nombre de morts accidentelles aux Etats-Unis de 1973 à 1978 (figure 1).

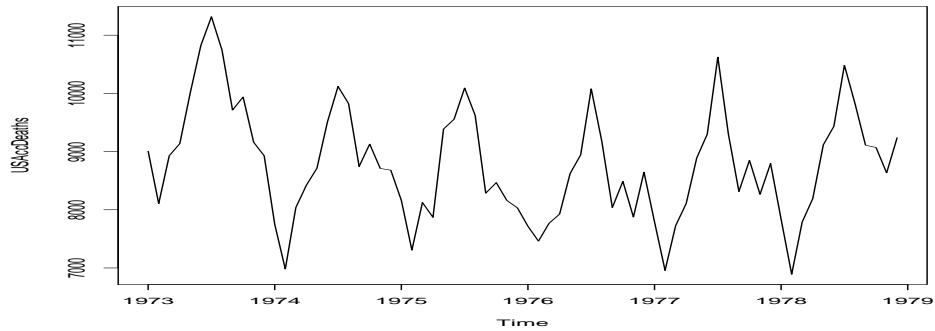


FIG. 1 – Nombre de morts accidentelles aux Etats-Unis de 1973 à 1978

- Ex. 2 : Nombre de passagers par mois (en milliers) dans les transports aériens, de 1949 à 1960 (figure 2).

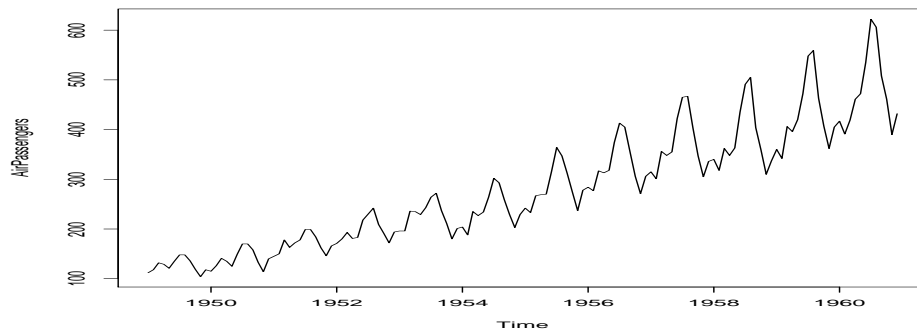


FIG. 2 – Nombre de passagers (en milliers) dans les transports aériens

- Ex. 3 : Nombre annuel de tâches solaires observées à la surface du soleil de 1700 à 1980 (figure 3).

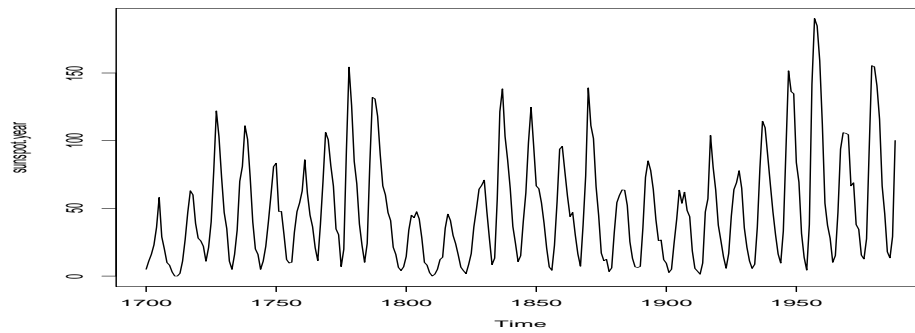


FIG. 3 – Nombre annuel de tâches solaires

- Ex. 4 : Taille de la population française (en milliers) de 1985 à 2005 (figure 4).

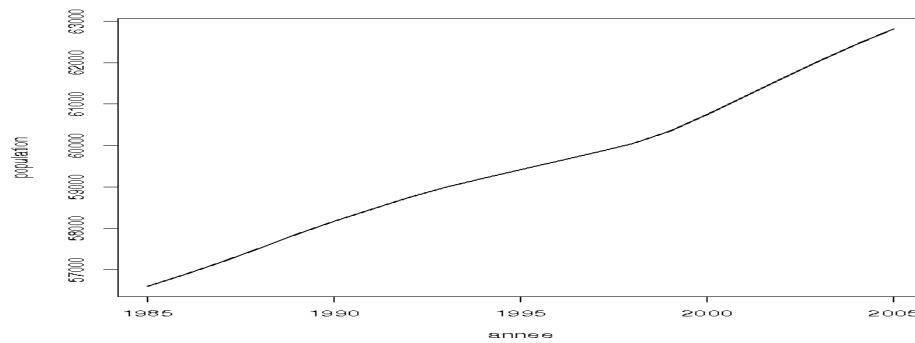


FIG. 4 – Population française de 1985 à 2005

- Ex. 5 : Valeurs de clôtures journalières du CAC40 de 1991 à 1998 (figure 5).

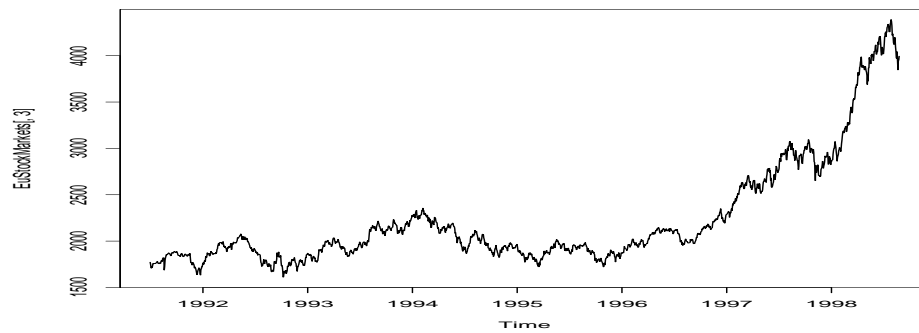


FIG. 5 – Valeurs de clôtures journalières du CAC40 de 1991 à 1998

Excepté l'exemple 4, ces données sont disponibles dans le logiciel R sous les noms : `EuStockMarkets`, `USAccDeaths`, `AirPassengers` et `sunspot.year`.

Exercice 1. Repérer les tendances (croissance, décroissance, linéaire, quadratique...) et saisonnalités (périodicités) de chacune de ces séries.

Un des objectifs principaux de l'étude d'une série temporelle est la **prévision** des réalisations futures, très souvent pour des raisons économiques (prévoir l'évolution de la vente d'un produit pour ajuster au mieux les moyens de production, prévoir l'évolution d'un marché financier ...).

Bien entendu, aucun modèle ne correspond exactement à la réalité, et il est impossible de prévoir parfaitement le devenir d'une série temporelle. Lorsque cela sera possible, nous donnerons des intervalles de prévisions, afin de pouvoir apporter une information quant à la précision de la prévision.

Pour ce faire, il existe un large choix de modèle utilisable :

- les modèles de régression, comme par exemple :

$$x_t = \alpha_1 t^2 + \alpha_2 t + \alpha_3 + \epsilon_t, \quad t = 1, \dots, n.$$

Une fois les coefficients de ce modèle estimés, la prévision de x_{t+1} sera $\hat{x}_{t+1} = \hat{\alpha}_1(t+1)^2 + \hat{\alpha}_2(t+1) + \hat{\alpha}_3$.

- les lissages exponentiels qui sont très simples à mettre en oeuvre, et qui feront l'objet d'un chapitre suivant,
- les modèles de type ARMA, qui consistent à enlever de la série les tendances et saisonnalités (ou périodicités) évidentes et à modéliser le résidu restant. Ces méthodes sont plus sophistiquées et plus lourdes numériquement (temps de calcul) que les précédentes, mais également plus performantes.

Parmi les 5 exemples précédents, celui relatif au nombre de passagers dans les transports aériens (figure 2) est une série assez typique de ce que l'on rencontre en économétrie, et elle donne lieu à de bonnes prévisions pour toutes les méthodes classiques. Au contraire, l'évolution des marchés boursiers (figure 5) est beaucoup plus difficile à prévoir.

Les défis que nous allons devoir relever sont les suivants :

- définir un modèle avec un nombre fini de paramètres,
- estimer les paramètres de ce modèle,
- vérifier la qualité d'ajustement du modèle, comparer différents modèles (partage de l'échantillon d'observations en 80% pour l'apprentissage et 20% pour le test),
- effectuer des prédictions.

1.1 Tendances et composantes saisonnières

On parle de **tendance** lorsque la série $(x_t)_{1 \leq t \leq n}$ peut s'écrire, à une erreur d'ajustement ϵ_t près, comme une combinaison linéaire de m fonctions du temps, choisies a priori (par exemple fonction puissance, exponentielle, logarithmique...) :

$$x_t = \sum_{j=1}^m \alpha_j f_j(t) + \epsilon_t \quad 1 \leq t \leq n.$$

Lorsque $x_t = \alpha t + \beta + \epsilon_t$ la tendance est **linéaire** ($m = 1$ et $f(t) = \alpha t + \beta$).

Une tendance **polynomiale** se traduira par $x_t = \alpha_1 t^p + \alpha_{p-1} t^{p-1} + \dots + \alpha_{p+1} + \epsilon_t$.

Exercice 2. Comment semble être la tendance dans l'exemple 5 ?

On parle de **composante périodique** lorsque la série $(x_t)_{1 \leq t \leq n}$ peut se décomposer en :

$$x_t = s_t + \epsilon_t \quad 1 \leq t \leq n,$$

où s_t est périodique, c'est-à-dire $s_{t+T} = s_t$, avec T la période (supposée entière).

Lorsque la période est de 6 mois ou 1 an, on parle généralement de composante saisonnière.

Enfin, il est fréquent qu'une série comporte à la fois une tendance et une composante périodique (cf. exemple 2).

1.2 Indices descriptifs d'une série temporelle

1.2.1 Indices de tendances centrales

Nous utilisons comme indicateur de la tendance centrale la **moyenne** :

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^n x_t.$$

1.2.2 Indices de dispersion

Nous utilisons comme indicateur de dispersion la **variance empirique** (et sa racine carrée, l'écart-type empirique) :

$$\hat{\sigma}_n(0) = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}_n)^2.$$

1.2.3 Indices de dépendance

Ces notions, plus spécifiques à l'étude de série temporelle, renseignent sur la dépendance entre les données x_t .

Auto-covariance L'**auto-covariance empirique** d'ordre 1 renseigne sur la dépendance entre deux données successives :

$$\hat{\sigma}_n(1) = \frac{1}{n-1} \sum_{t=1}^{n-1} (x_t - \bar{x}_n)(x_{t+1} - \bar{x}_n),$$

l'auto-covariance empirique d'ordre 2 renseigne sur la dépendance entre deux données écartées de deux pas de temps :

$$\hat{\sigma}_n(2) = \frac{1}{n-2} \sum_{t=1}^{n-2} (x_t - \bar{x}_n)(x_{t+2} - \bar{x}_n),$$

et ainsi de suite. Pour des raisons de bon sens statistique, nous ne considérerons les covariances empiriques que jusqu'à un ordre h pas trop grand.

On appelle **fonction d'auto-covariance** (empirique) la fonction qui à h associe $\hat{\sigma}_n(h)$.

Auto-corrélation Les **auto-corrélations empiriques** sont les quotients des covariances empiriques par la variance empirique :

$$\hat{\rho}_n(h) = \frac{\hat{\sigma}_n(h)}{\hat{\sigma}_n(0)}.$$

Ce sont les auto-corrélations empiriques que nous utiliserons pour caractériser la dépendance entre les variables.

On appelle **fonction d'auto-corrélation** (empirique) la fonction qui à h associe $\hat{\rho}_n(h)$.

Visualisation de l'auto-corrélation d'ordre 1 La représentation graphique des nuages de points (x_t, x_{t+1}) , pour $t = 1, \dots, n-1$, est une bonne illustration de la valeur de l'auto-corrélation d'ordre 1 $\hat{\rho}_n(1)$: plus le nuage sera arrondi, plus $\hat{\rho}_n(1)$ sera proche de 0, et plus le nuage sera allongé, plus $\hat{\rho}_n(1)$ sera proche de 1. Ceci est vrai pour toutes les auto-corrélations $\hat{\rho}_n(k)$, $1 \leq k \leq n$.

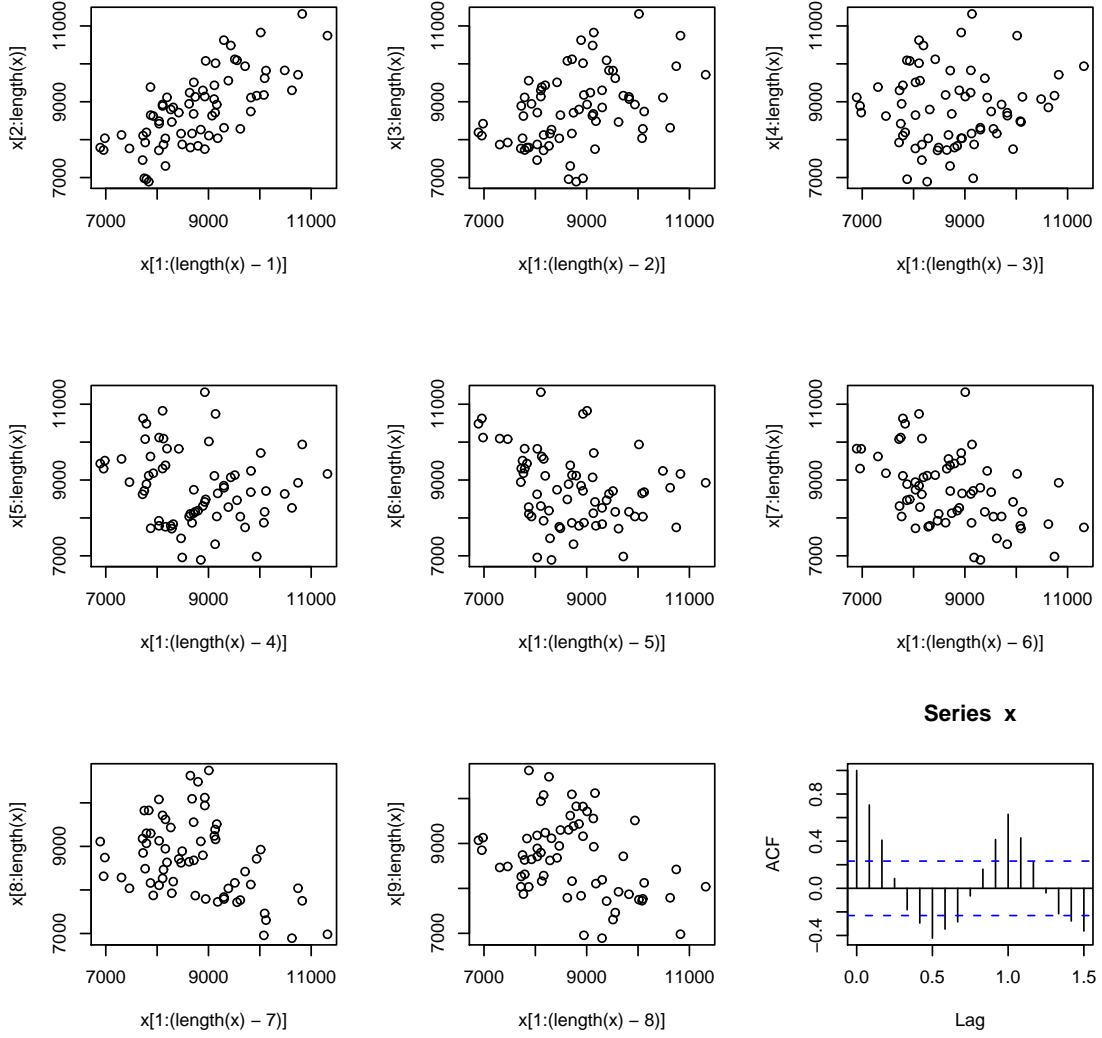


FIG. 6 – Nuages de points (x_t, x_{t+k}) pour $k = 1, \dots, 8$ et auto-corrélation pour la série temporelle du nombre de morts accidentelles aux Etats-Unis de 1973 à 1978

Première analyse de la série à l'aide des auto-corrélations

Proposition 1. Si la série $(x_t)_{1 \leq t \leq n}$ est une tendance linéaire pure $x_t = at + b, t = 1, \dots, n$, alors on a pour h fixé :

$$\hat{\rho}_n(h) \xrightarrow[n \rightarrow \infty]{} 1.$$

Exercice 3. Faire la preuve.

Proposition 2. Si la série $(x_t)_{1 \leq t \leq n}$ est une série périodique pure $x_t = a \cos \frac{2t\pi}{T}, t = 1, \dots, n$, on a pour h fixé :

$$\hat{\rho}_n(h) \xrightarrow[n \rightarrow \infty]{} \cos \frac{2h\pi}{T}.$$

En interprétant l'auto-corrélation grâce à ces deux propositions, il sera possible de « deviner » si une série temporelle admet une tendance (l'auto-corrélation tend vers 1) ou une saisonnalité (la saisonnalité se voit sur l'auto-corrélation).

1.3 Mise en oeuvre sous R

Quelques fonctions R utiles à l'étude des séries temporelles :

- Lire un fichier de données en sautant les k premières lignes : `data=scan(file="donnee.dat", skip=k)`.
- Créer un objet de type série temporelle : `serie <- ts(data, start, end, frequency)`.
data contient le vecteur des données (un fichier contenant les données peut être mentionné en remplaçant data par file="donnees.dat"), start et end mentionne les dates de début et de fin de la série (ex : start=c(1990, 1) et end=c(1999, 6) pour des données allant de janvier 90 à juin 99), et enfin frequency mentionne le nombre de données par unité de temps (par exemple, si les dates de début et de fin sont des années, et que les données sont mensuelles, il faudra indiquer frequency=12).
- Représenter graphiquement un objet de type série temporelle : `plot.ts(serie)`
- La fonction `acf(x, lag.max = 10, type = c("correlation", "covariance"), plot = TRUE)` calcule (et trace si l'option plot est à TRUE) les lag.max premières auto-corrélations et auto-covariances.

Quelques conseils utiles pour les graphiques en R :

- pour représenter plusieurs courbes sur le même graphique, tracer la première à l'aide de la commande `plot` qui crée la fenêtre graphique et y insère la courbe, puis tracer les autres courbes à l'aide de la commande `lines` qui trace une courbe sur une fenêtre graphique existante.
- pour partager la fenêtre graphique en plusieurs ($n \times p$) sous-graphes, utiliser la commande `par(mfrow=c(n, p))`.
- préciser les limites des axes des graphiques : `plot(..., xlim=c(0, 10), ylim=c(-10, 10))`.
- pour exporter les graphiques en *jpeg* (idem pour *bmp*, *png*), il faut lui procéder de la sorte
 1. `jpeg(filename="nomfichier%d.jpeg")`,
 2. réaliser le graphique,
 3. la commande `dev.off()` permet enfin de rediriger le dernier graphique tracé vers le fichier `nomfichier1.jpeg`, et ainsi de suite après chaque graphique. Le nom de fichier sera automatiquement incrémenté.

2 Lissages exponentiels

Les méthodes de lissages exponentiels constituent un outil permettant de réaliser des prévisions à partir de l'observation d'une série temporelle. Ces méthodes étant relativement basiques et simples de mise en oeuvre, elles sont souvent utilisées dans l'industrie, notamment lorsque le nombre de prévisions à réaliser est important (par exemple, prévisions des ventes de centaines de produits dans une grande surface).

Nous présentons trois types de lissage exponentiel :

- le lissage exponentiel simple qui consiste à ajuster localement à la série temporelle une constante,
- le lissage exponentiel double qui ajuste quant à lui une droite,
- le lissage exponentiel de Holt-Winters qui considère des fonctions plus complexes (polynomiales, périodiques...).

2.1 Lissage exponentiel simple

Disposant d'une série temporelle x_1, \dots, x_n , l'objectif du lissage exponentiel est d'estimer la valeur x_{n+h} non encore observée. Nous noterons $\hat{x}_{n,h}$ cette prévision.

Etant donnée une constante de lissage $0 < \alpha < 1$, on définit la **prévision par lissage exponentiel simple** :

$$\hat{x}_{n,h} = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}. \quad (1)$$

La prévision est une moyenne de toutes les observations passées, pondérée de sorte à ce que plus l'observation soit ancienne moins elle ait d'importance.

Une constante de lissage α proche de 0 (≤ 0.3) donne une importance significative aux observations éloignées, tandis qu'un α proche de 1 (≥ 0.7) tend à négliger ces observations éloignées.

Remarque : la prévision $\hat{x}_{n,h}$ ne dépend pas de h !

Formules récursives de mise à jour La définition (1) vérifiant la formule récursive suivante

$$\hat{x}_{n,h} = \alpha x_n + (1-\alpha)\hat{x}_{n-1,h},$$

la prévision $\hat{x}_{n,h}$ peut être obtenue immédiatement à partir de la connaissance de :

- 1- la prévision $\hat{x}_{n-1,h}$ basée sur les $n-1$ -èmes premières observations,
- 2- l'observation x_n .

L'utilisation de cette récurrence permet de réaliser des algorithmes très rapides d'estimation de la prévision par lissage exponentiel (en initialisant à $\hat{x}_{1,h} = x_1$).

Exercice 4. *Ecrire et interpréter la valeur de $\hat{x}_{n,1}$ à partir de l'équation de récurrence.*

Exercice 5. *Montrer que $\hat{x}_{n,h}$ défini en (1) est solution asymptotique d'un problème de moindres carrés pondérés.*

Choix de la constante de lissage Pour choisir la constante de lissage, une solution pragmatique consiste à tester plusieurs valeurs et à choisir celle minimisant un critère d'erreur minimale. Pour cela on partage l'échantillon d'observations en un échantillon d'apprentissage (les 80% premières observations : x_1, \dots, x_m où m est par exemple l'entier le plus proche de $\frac{8}{10}n$) et un échantillon test (les 20% dernières : x_{m+1}, \dots, x_n), on estime le modèle de lissage exponentiel à partir de l'échantillon d'apprentissage, et on évalue l'erreur sur l'échantillon test :

$$\text{erreur} = \sum_{h=1}^{n-m} (\hat{x}_{t,h} - x_{t,h})^2$$

On répète cette opération pour plusieurs valeurs de la constante de lissage α , et on choisit celle conduisant à l'erreur la plus petite.

2.2 Lissage exponentiel double

On ajuste au voisinage de l'instant n une droite d'équation $y_t = a_1 + a_2(t - n)$.
La **prévision par lissage exponentiel double** est :

$$\hat{x}_{n,h} = \hat{a}_1 + \hat{a}_2 h$$

où \hat{a}_1 et \hat{a}_2 sont solution de

$$\inf_{a_1, a_2 \in \mathbb{R}} \sum_{j=0}^{n-1} (1 - \alpha)^j (x_{n-j} - (a_1 + a_2 j))^2.$$

Les solutions de cette équation sont

$$\hat{a}_1 = 2L_1(n) - L_2(n) \quad \text{et} \quad \hat{a}_2 = \frac{\alpha}{1 - \alpha} (L_1(n) - L_2(n))$$

où $L_1(n) = \alpha \sum_{j=0}^{n-1} (1 - \alpha)^j x_{n-j}$ et $L_2(n) = \alpha \sum_{j=0}^{n-1} (1 - \alpha)^j L_1(n - j)$ sont deux lissages exponentiels simples successifs.

Remarque : comme pour le lissage exponentiel simple, l'estimateur de la prévision est la meilleure approximation au sens des moindres carrés pondérés.

Formules récursives de mise à jour

$$\begin{aligned} \hat{a}_1(n) &= \hat{a}_1(n-1) + \hat{a}_2(n-1) + \alpha(2 - \alpha)(x_n - \hat{x}_{n-1,1}), \\ \hat{a}_2(n) &= \hat{a}_2(n-1) + \alpha(2 - \alpha)(x_n - \hat{x}_{n-1,1}), \end{aligned}$$

où $\hat{a}_1(n)$ et $\hat{a}_2(n)$ sont les estimations des paramètres a_1 et a_2 lorsque l'on a observé la série jusqu'à la n -ème réalisation. Les valeurs initiales étant $\hat{a}_1(0) = x_1$ et $\hat{a}_2(0) = x_2 - x_1$.

2.3 Méthode de Holt-Winters

2.3.1 Méthode non saisonnière

Comme la méthode de lissage exponentiel double, celle de Holt-Winters non saisonnière revient à estimer au voisinage de l'instant n une droite

$$y_t = a_1 + a_2(t - n).$$

La prévision prend la forme

$$\hat{x}_{n,h} = \hat{a}_1 + \hat{a}_2 h.$$

La variante par rapport à la méthode de lissage exponentiel double est au niveau des formules de mise à jour dans l'estimation des paramètres a_1 et a_2 .

Soient deux constantes de lissages $0 < \alpha < 1$ et $0 < \beta < 1$. Les formules de mise à jour sont :

$$\begin{aligned} \hat{a}_1(n) &= \alpha x_n + (1 - \alpha)[\hat{a}_1(n-1) + \hat{a}_2(n-1)], \\ \hat{a}_2(n) &= \beta[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1). \end{aligned}$$

Exercice 6. Montrer que les formules de mise à jour du lissage exponentiel double sont un cas particulier de ces dernières.

Remarque :

- l'introduction de deux constantes rend la méthode plus souple que le lissage exponentiel double : la constante α joue un rôle dans l'estimation de l'ordonnée à l'origine de la droite, a_1 , et la constante β dans celle de la pente de la droite, a_2 .
- si α et β sont petits le lissage est important car on tient compte du passé lointain.

2.3.2 Méthode saisonnière additive

On cherche maintenant à ajuster au voisinage de l'instant n une droite d'équation

$$y_t = a_1 + a_2(t - n) + s_t,$$

où s_t est une composante périodique de période T .

Les formules récursives de mise à jour sont :

$$\begin{aligned}\hat{a}_1(n) &= \alpha(x_n - \hat{s}_{n-T}) + (1 - \alpha)[\hat{a}_1(n-1) + \hat{a}_2(n-1)], \\ \hat{a}_2(n) &= \beta[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1), \\ \hat{s}_n &= \gamma[x_n - \hat{a}_1(n)] + (1 - \gamma)\hat{s}_{n-T}.\end{aligned}$$

Les prévisions sont de la forme :

$$\begin{aligned}\hat{x}_{n,h} &= \hat{a}_1 + \hat{a}_2 h + \hat{s}_{n+h-T} \quad 1 \leq h \leq T, \\ \hat{x}_{n,h} &= \hat{a}_1 + \hat{a}_2 h + \hat{s}_{n+h-2T} \quad T+1 \leq h \leq 2T.\end{aligned}$$

et ainsi de suite pour $h \geq 2T$.

Les trois constantes de lissage, α , β et γ ont le même effet que précédemment, plus elles sont petites et plus l'importance des données éloignées est significative. Elles agissent respectivement sur les paramètres a_1 , a_2 et s_t .

Se référer à Gouriéroux et Monfort 1983 [4] pour les valeurs d'initialisation.

2.3.3 Méthode saisonnière multiplicative

On ajuste au voisinage de l'instant n une droite d'équation

$$y_t = [a_1 + a_2(t - n)] \times s_t,$$

où s_t est une composante périodique de période T .

Les formules récursives de mise à jour sont :

$$\begin{aligned}\hat{a}_1(n) &= \alpha \frac{x_n}{\hat{s}_{n-T}} + (1 - \alpha)[\hat{a}_1(n-1) + \hat{a}_2(n-1)], \\ \hat{a}_2(n) &= \beta[\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \beta)\hat{a}_2(n-1), \\ \hat{s}_n &= \gamma \frac{x_n}{\hat{a}_1(n)} + (1 - \gamma)\hat{s}_{n-T}.\end{aligned}$$

Les prévisions sont de la forme :

$$\begin{aligned}\hat{x}_{n,h} &= [\hat{a}_1 + \hat{a}_2 h] \hat{s}_{n+h-T} \quad 1 \leq h \leq T, \\ \hat{x}_{n,h} &= [\hat{a}_1 + \hat{a}_2 h] \hat{s}_{n+h-2T} \quad T+1 \leq h \leq 2T.\end{aligned}$$

Se référer également à [4] pour les valeurs d'initialisation.

2.4 Mise en oeuvre sous R

Les méthodes de lissages exponentiels sont disponibles sous R, grâce à la fonction `HoltWinters`.

Pour une série temporelle x , cette procédure permet :

- un lissage exponentiel simple :
`xlisse <- HoltWinters(x, alpha= α , beta=0, gamma=0),`
- un lissage de Holt-Winters sans composante saisonnière :
`xlisse <- HoltWinters(x, alpha= α , beta= β , gamma=0),`
- un lissage Holt-Winters additif :
`xlisse <- HoltWinters(x, alpha= α , beta= β , gamma= γ , seasonal="add"),`
- un lissage Holt-Winters multiplicatif :
`xlisse <- HoltWinters(x, alpha= α , beta= β , gamma= γ , seasonal="mul").`

A noter que pour un lissage de Holt-Winters avec composante saisonnière la série temporelle x doit être un objet de type série temporelle, défini avec la fonction `ts` en précisant la saisonnalité.

L’affichage et la visualisation des résultats peuvent être réalisés à l’aide des commandes :

- `summary(xlisse)` : description de l’objet `xlisse` obtenu précédemment par la procédure `HoltWinters`,
- `plot(xlisse)` : représentation des valeurs observées et des valeurs lissées,
- `plot(xlisse$fitted[,1])` : représentation de l’ajustement de la série remis à jour à chaque observation.

Les prévisions à l’horizon h sont réalisées à l’aide de la fonction `predict` :

`p<-predict(xlisse,n.ahead=h).`

Un intervalle de confiance (dont le fondement théorique n’a pas été étudié dans ce cours) peut être obtenu en validant (à TRUE) l’option `prediction.interval`.

Remarque : lorsqu’aucune valeur n’est précisée pour les constantes de lissage, un algorithme interne à la procédure `HoltWinters` se charge d’estimer la meilleur constante possible à partir de la série des observations.

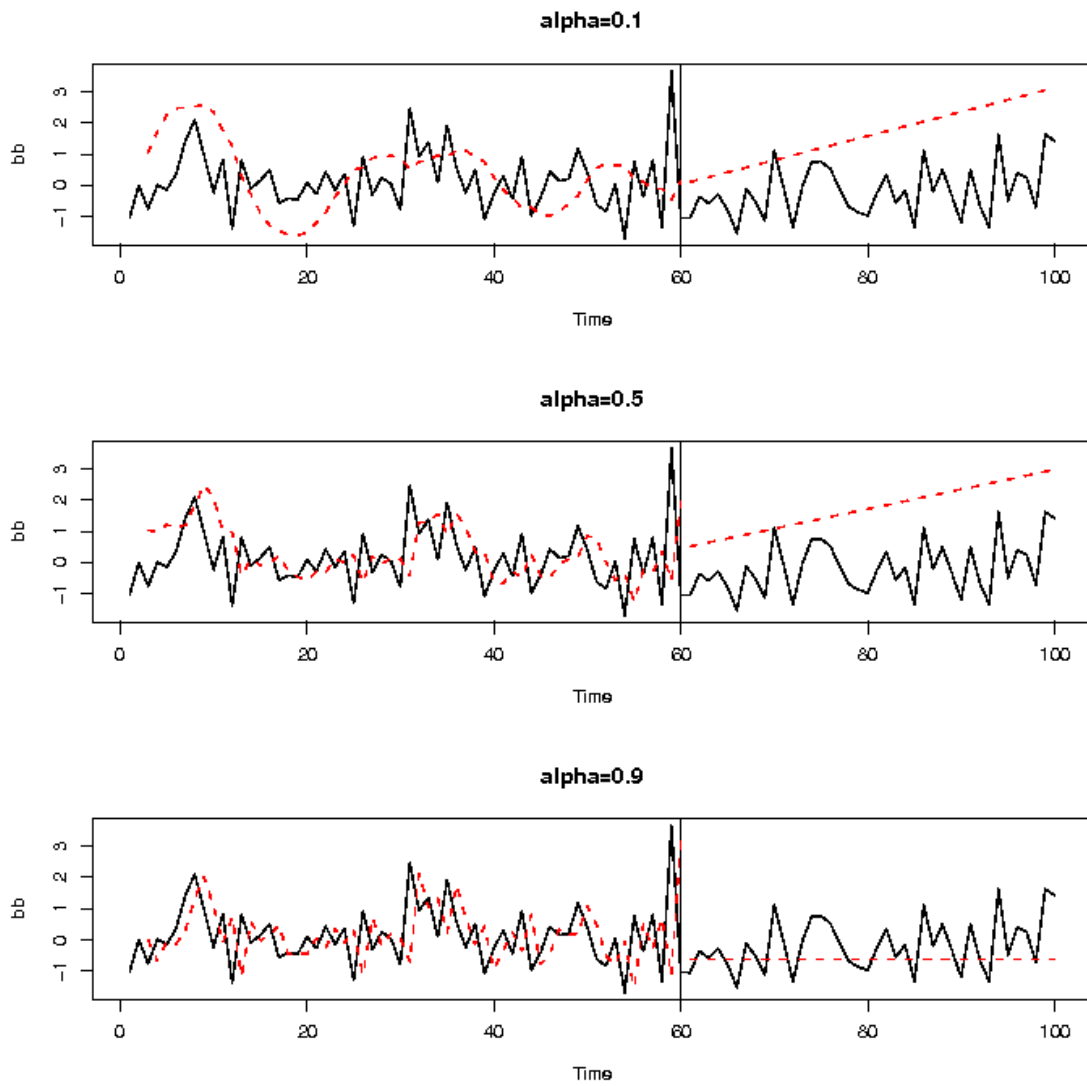


FIG. 7 – Lissage et prévision par lissage exponentiel double d'un bruit blanc gaussien

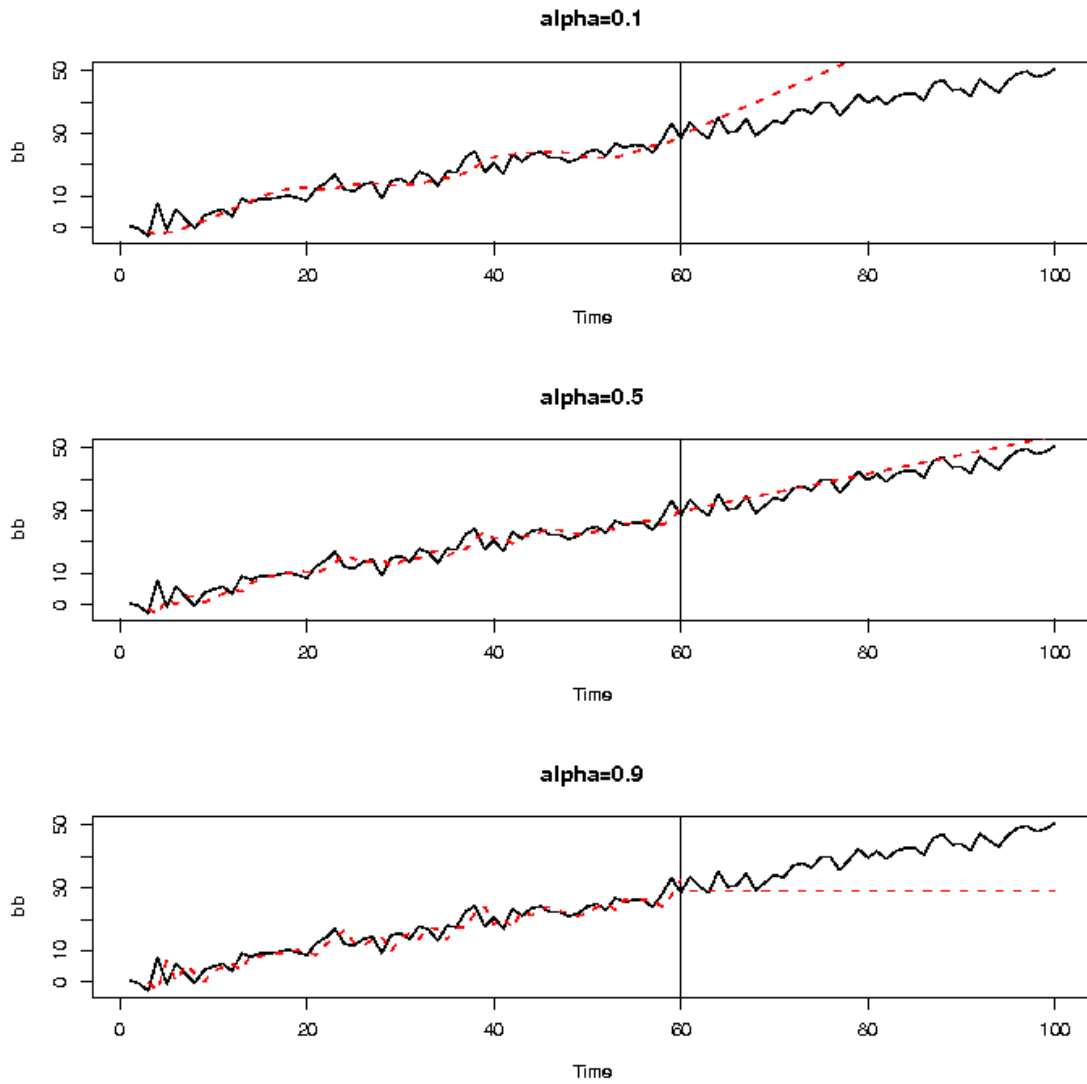


FIG. 8 – Lissage et prévision par lissage exponentiel double de la série $X(t) = 0.5t + 2\epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$

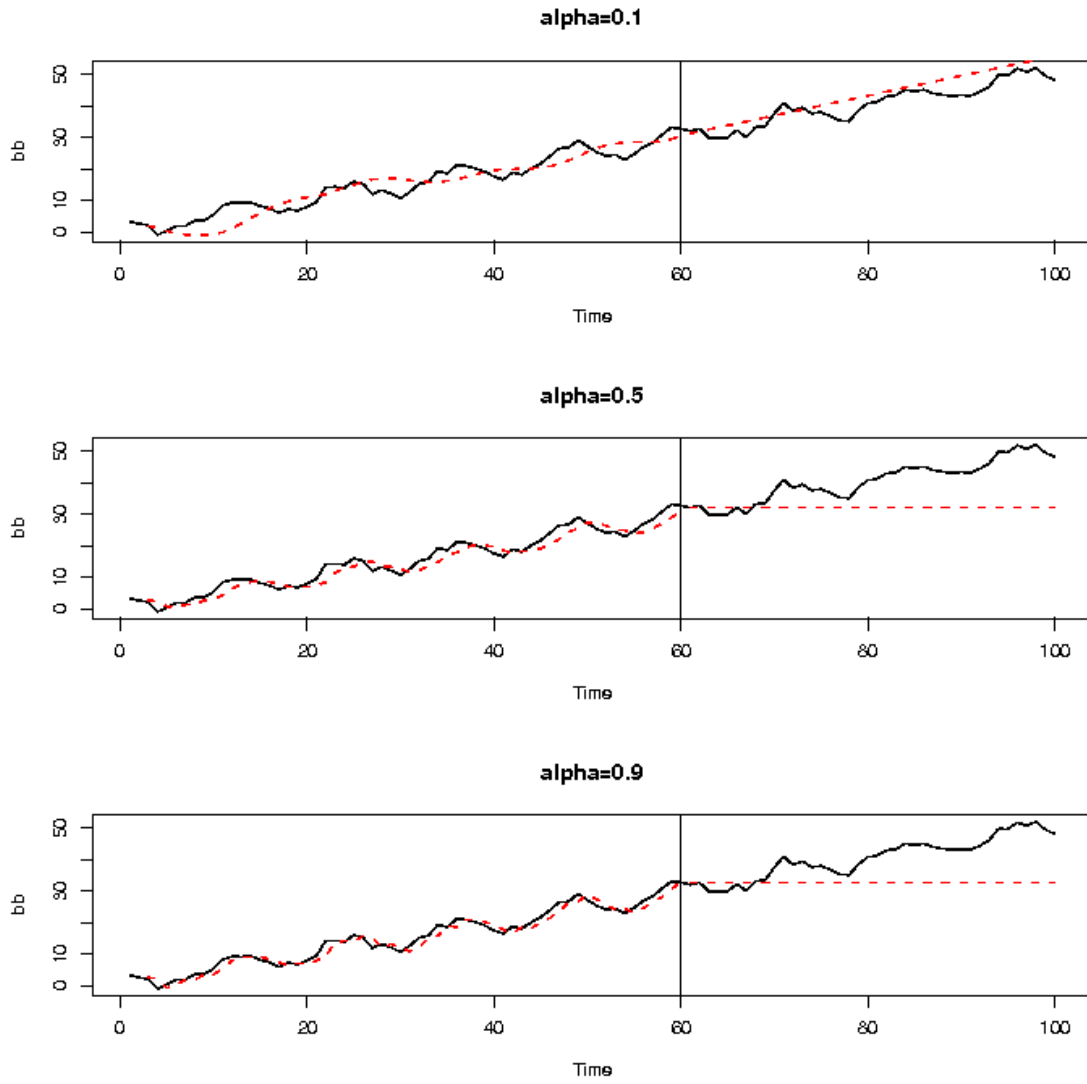


FIG. 9 – Lissage et prévision par lissage exponentiel double de la série $X(t) = 0.5t + \epsilon_t + 3 \cos\left(t \frac{\pi}{6}\right)$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$

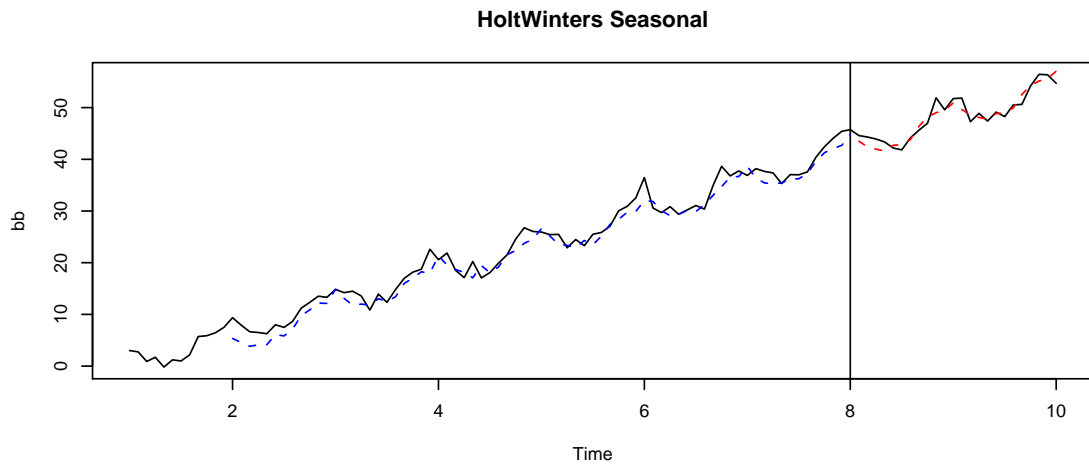
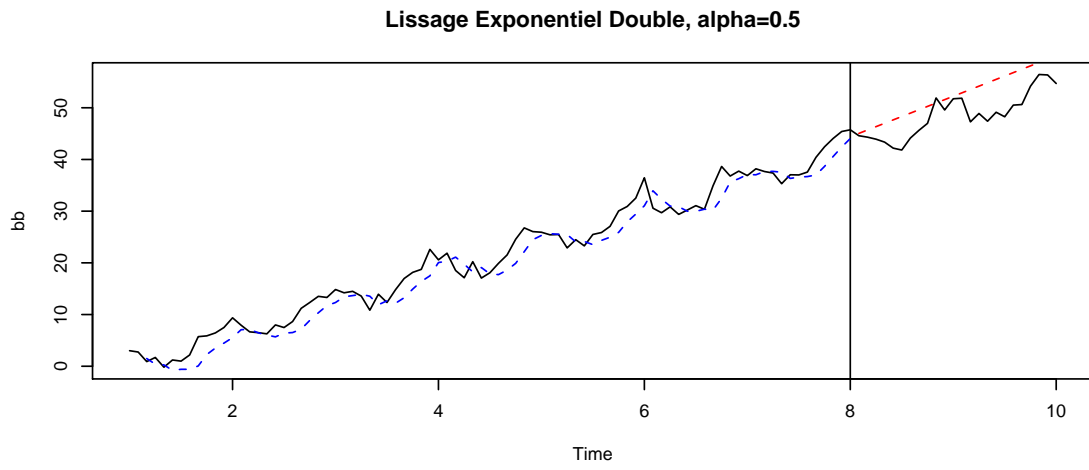
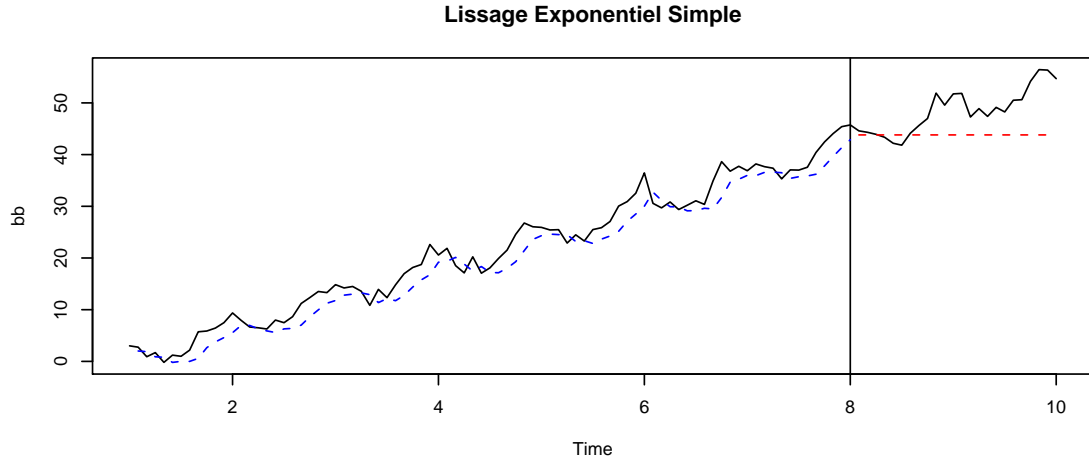


FIG. 10 – Lissage et prévision par lissage exponentiel simple, double, et Holt-Winters avec composante saisonnière de la série $X(t) = 0.5t + \epsilon_t + 3 \cos\left(t\frac{\pi}{6}\right)$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$

3 Estimation et élimination de la tendance et de la saisonnalité

Une **série temporelle** $(x_t)_{1 \leq t \leq n}$ est l'observation des n premières réalisations d'un processus stochastique $(X_t)_t$. C'est ce processus que l'on cherche désormais à modéliser. Pour cela, la démarche suivante doit être adoptée :

- représenter graphiquement la série afin de repérer les tendances et saisonnalités,
- estimer et supprimer les tendances et saisonnalités (partie déterministe du processus stochastique),
- choisir un modèle pour les résidus (partie aléatoire du processus stochastique) et l'estimer,
- prédire les réalisations futures à l'aide de ce modèle.

L'objectif de cette section est de donner quelques méthodes pour estimer et supprimer les tendances et saisonnalités. La fin de ce cours sera concentré sur la modélisation de processus stationnaires.

3.1 Bruit blanc

Définition 1. Un processus de **bruit blanc** est une suite de variables aléatoires $(X_t)_t$ indépendantes, d'espérance et de variance constantes. Si l'espérance est nulle, le bruit blanc est **centré**, et si les variables aléatoires sont gaussiennes, le bruit blanc est **gaussien**.

3.2 Processus stationnaire

Un processus aléatoire $(X_t)_t$ est **stationnaire** s'il est d'espérance constante

$$E[X_t] := \mu \quad \forall t,$$

et si les covariances sont stables par translation dans le temps, c'est-à-dire, pour tout h

$$\text{Cov}(X_t, X_{t+h}) := \sigma(h) \quad \forall t.$$

On appelle **fonction d'auto-covariance** du processus stationnaire la suite $\sigma(h)$, et **fonction d'auto-corrélation** du processus stationnaire la suite $\rho(h) := \frac{\sigma(h)}{\sigma(0)}$.

Exercice 7. Montrer que $\sigma(h) = \sigma(-h)$.

3.3 Une estimation paramétrique de la tendance (*trend*)

Nous supposons que la série temporelle étudiée soit la réalisation d'un processus stochastique composé d'une tendance déterministe m_t et d'une partie aléatoire ϵ_t (supposée de moyenne nulle) :

$$X_t = m_t + \epsilon_t.$$

Une méthode simple consiste à supposer que cette tendance est linéaire :

$$m_t = a + bt,$$

et d'estimer les paramètres a et b par moindres carrés.

Ainsi, si on observe la série x_1, \dots, x_n , il faut trouver a et b qui minimisent la quantité :

$$\sum_{t=1}^n (x_t - a - bt)^2$$

Les solutions de ce problème sont :

$$\begin{aligned} \hat{a} &= \frac{6}{n(n-1)} \left(-\sum_{t=1}^n tx_t + \frac{2n+1}{3} n\bar{x} \right), \\ \hat{b} &= \frac{12}{n(n^2-1)} \left(\sum_{t=1}^n tx_t - \frac{n+1}{2} n\bar{x} \right). \end{aligned}$$

Exercice 8. *Ecrire le problème sous forme matricielle... et le résoudre.*

L'hypothèse de linéarité de la tendance convient très bien à certaines séries temporelles : par exemple, celles des figures 1, 2 (et encore...), 4... Mais ce n'est pas le cas de toutes les séries : voir par exemple celle représentant le cours du CAC40, figure 5.

Il est alors possible de supposer que la tendance soit de forme polynomiale :

$$m_t = a + bt + ct^2$$

et d'estimer les paramètres a , b et c par moindres carrés.

Mais il est parfois difficile d'estimer le degré du polynôme, et lorsque le degré est trop important, le nombre de paramètres à estimer devient grand et les calculs fastidieux. Dans cette situation, on a recours à une méthode d'estimation non paramétrique.

3.4 Estimation non paramétrique : moyenne mobile

Tendance Supposons que la tendance m_t soit linéaire dans un petit intervalle $[t - q, t + q]$ autour de t . Dans ce cas, un bon estimateur de la tendance est la *moyenne* sur cet intervalle :

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

On peut donc estimer la tendance à chaque temps t en calculant la moyenne sur les observations étant dans une fenêtre de largeur $2q + 1$ autour de t : c'est ce que l'on appelle une estimation par **moyenne mobile**.

Pour éviter les problèmes de bord, on suppose que $x_t = x_1$ si $t < 1$ et $x_t = x_n$ si $t > n$.

Tendance et saisonnalité Supposons désormais que le processus ne comporte pas uniquement une tendance, mais également une saisonnalité :

$$X_t = m_t + s_t + \epsilon_t,$$

avec s_t une fonction T -périodique.

Le principe d'estimation est (en simplifiant légèrement) le suivant : on estime la tendance moyenne sur une période, puis on estime la composante saisonnière en moyennant sur toutes les périodes les écarts à la tendance moyenne de la période.

Application à la série du nombre de morts accidentelles aux Etats-Unis La figure 11 représente la série du nombre de morts accidentelles aux Etats-Unis de 1973 à 1978 (déjà présentée en introduction), la figure 12 représente la tendance (estimée par moyenne mobile) et la figure 13 représente la composante saisonnière. La figure 14 représente la série après élimination de la tendance et la figure 15 représente la série après élimination de la tendance et de la composante saisonnière.

La série 15 ainsi obtenue est une série (supposée) stationnaire, sur laquelle nous chercherons plus tard à ajuster un modèle.

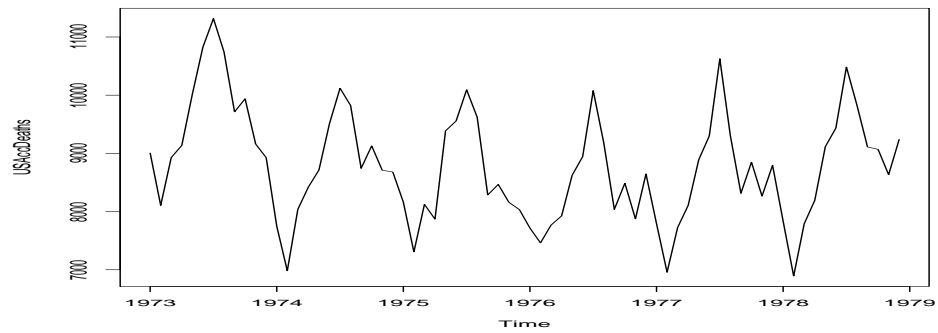


FIG. 11 – Série *USAccDeaths* : nombre de morts accidentelles aux Etats-Unis de 1973 à 1978



FIG. 12 – Tendence de la série *USAccDeaths*

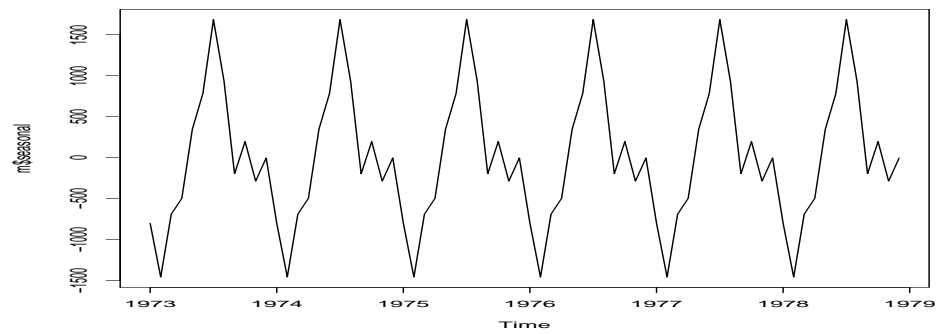


FIG. 13 – Composante saisonnière de la série *USAccDeaths*

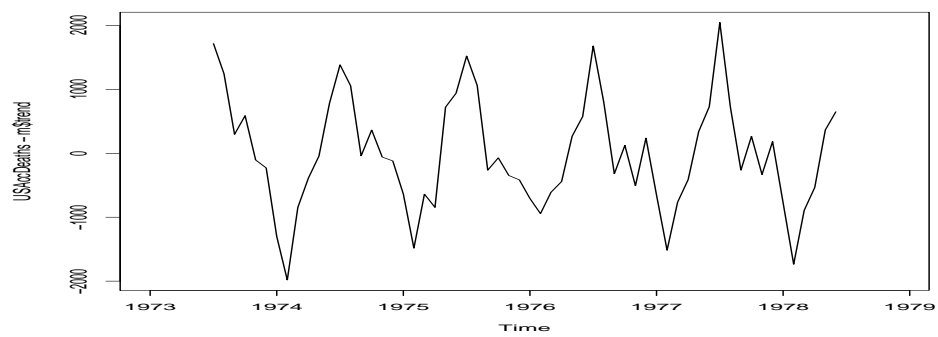


FIG. 14 – Série *USAccDeaths* après élimination de la tendance

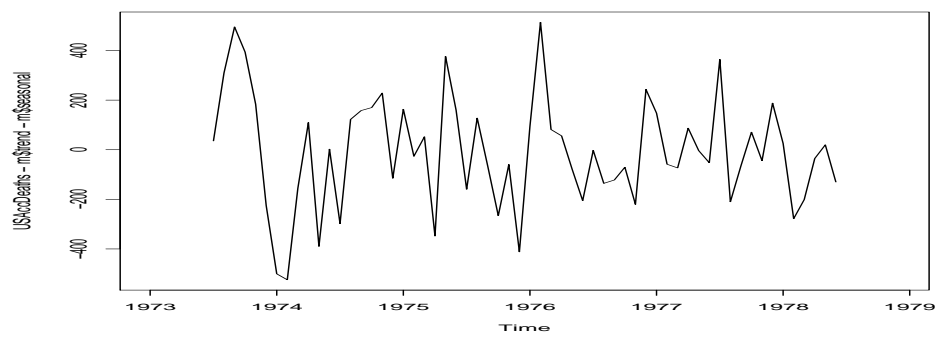


FIG. 15 – Série *USAccDeaths* après élimination de la tendance et de la composante saisonnière

3.5 Elimination de la tendance et de la saisonnalité par la méthode des différences

Cette méthode permet de supprimer les tendance et saisonnalité d'une série temporelle sans les estimer. Soit Δ_T l'opérateur qui associe $(X_t - X_{t-T})$ à (X_t) :

$$\Delta_T X_t = X_t - X_{t-T}.$$

On note Δ l'opérateur Δ_1 , et Δ_T^k l'opérateur $\underbrace{\Delta_T \circ \dots \circ \Delta_T}_{k \text{ fois}}$.

Proposition 3. Soit un processus admettant une tendance polynomiale d'ordre k :

$$X_t = \underbrace{\sum_{j=0}^k a_j t^j}_{m_t} + \epsilon_t.$$

Le processus ΔX_t admet une tendance polynomiale d'ordre $k - 1$.

Exercice 9. Faire la preuve.

Ainsi, en appliquant k fois Δ , on élimine la tendance.

Remarque : il est important de remarquer que si l'on applique Δ_t quelque soit t , le résultat est le même quant à l'élimination de la tendance.

Comme en pratique il n'est pas évident de connaître le degré k , on appliquera l'opérateur Δ jusqu'à ce que la moyenne du processus soit nulle (k sera généralement 1, 2 ou 3).

Proposition 4. Soit un processus admettant une tendance m_t et une saisonnalité, de période T :

$$X_t = m_t + s_t + \epsilon_t.$$

Dans ce cas,

$$\Delta_T X_t = (m_t - m_{t-T}) + (\epsilon_t - \epsilon_{t-T})$$

est un processus désaisonnalisé.

De plus, si la tendance du processus est linéaire, elle est également supprimée.

Exercice 10. Faire la preuve.

Si la tendance est plus que linéaire, il suffit d'appliquer la procédure précédente pour finir de supprimer la tendance, et obtenir ainsi un processus que l'on supposera stationnaire.

La figure 16 illustre l'élimination de la tendance linéaire et de la saisonnalité de la série $x_t = \frac{t}{2} + 3 \cos \frac{t\pi}{6} \epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$.

3.6 Test sur la série résiduelle

L'objectif des techniques présentées dans la section 3.5 est d'obtenir une série stationnaire (ou tout au moins le plus stationnaire possible), et en particulier sans tendance ou saisonnalité. L'étape suivante consiste à modéliser la série résiduelle obtenue. La première chose à faire est donc de tester s'il y a dépendance entre les termes de cette série. Si ce n'est pas le cas, on dit que la série résiduelle (stationnaire) est un bruit blanc (définition 1).

Si la série résiduelle obtenue après désaisonnalisation et élimination de la tendance, est un bruit blanc, il n'est donc pas utile d'aller plus loin dans la modélisation si ce n'est d'estimer la moyenne et variance du bruit blanc.

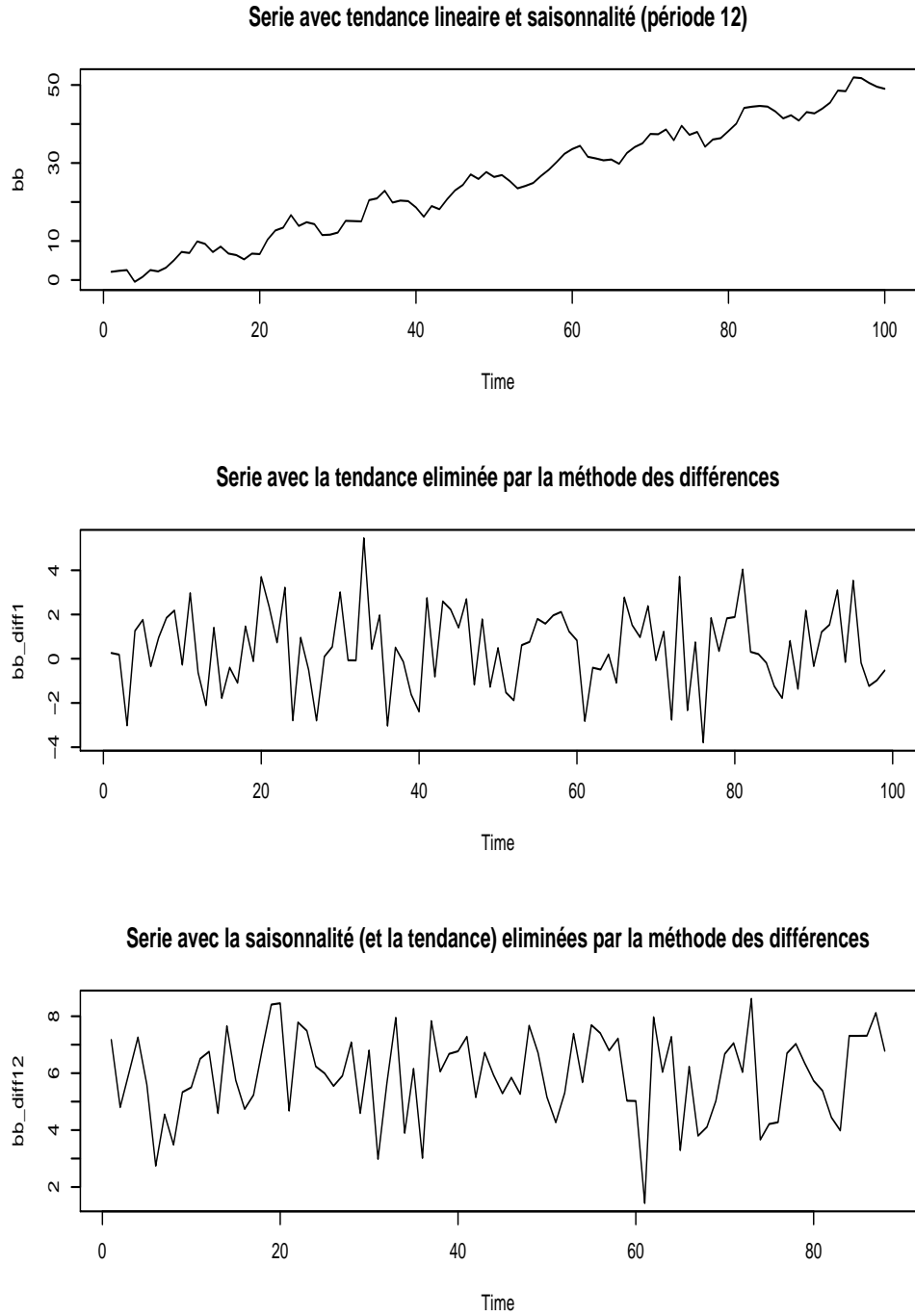


FIG. 16 – Elimination de la tendance et de la saisonnalit  par la m thode des diff rences (figure du haut : s rie $x_t = \frac{t}{2} + 3 \cos \frac{t\pi}{6} \epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$; figure du milieu : s rie $x_t - x_{t-1}$; figure du bas : s rie $x_t - x_{t-12}$)

3.6.1 Comment tester si on est en présence d'un bruit blanc ?

Par l'étude de la fonction d'auto-corrélation empirique Lorsque n est assez grand, les auto-corrélations d'un bruit blanc sont approximativement indépendantes et de loi $\mathcal{N}(0, \frac{1}{n})$. Ainsi, 95% des auto-corrélations devraient se trouver dans l'intervalle $[\frac{-1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}]$, et en traçant les 40 premières auto-corrélations il ne devrait pas y en avoir plus de 2 voir 3 en dehors de ces limites.

A noter que ces bornes sont tracées lorsque l'on demande à R de représenter graphiquement les auto-corrélations.

A l'aide du test du portemanteau Plutôt que de regarder si chaque auto-corrélation est dans les bornes de l'intervalle précédent, nous considérons la statistique définie par la somme des h premières auto-corrélations au carré

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j).$$

D'après la remarque précédente sur la normalité des auto-corrélations, la statistique Q suit une loi du χ^2 à h degrés de liberté. Il est donc possible de construire un test qui consistera à rejeter l'hypothèse nulle (la série est un bruit blanc) si Q est supérieur au quantile $\chi_{h,1-\alpha}^2$.

Ljung et Box (1978) ont amélioré ce test en considérant la statistique

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j},$$

dont la distribution est mieux approximée que la précédente par une loi du χ^2 à h degrés de liberté. C'est ce test qui est implémenté dans la fonction `Box.test` de R.

3.7 Mise en oeuvre sous R

La fonction `decompose` permet d'extraire d'une série temporelle (via la méthode de la moyenne mobile) :

```
serie_decomp<-decompose(serie,type=c("additive","multiplicative"))
```

- la composante saisonnière : `serie_decomp$seasonal`, que l'on suppose additive ou multiplicative dans l'option `type`,
- la tendance : `serie_decomp$trend`,
- le partie aléatoire stationnaire de la série : `serie_decomp$random`.

La fonction `diff.ts(serie, lag=T, difference=k)` permet d'appliquer l'opérateur de différenciation Δ_T^k .

La fonction `Box.test(serie, lag=H)` examine l'hypothèse nulle de nullité des H première auto-covariance, à l'aide du test du portemanteau. Par défaut H est fixé à 1, et seule la nullité de l'auto-covariance d'ordre 1 est testée. Pour tester si la série peut-être apparentée à un bruit blanc, nous fixerons arbitrairement un H de l'ordre de 20 (nous considérerons abusivement que si les 20 premières auto-corrélations sont nulles, la série est indépendante).

4 Modélisation des séries stationnaires

Nous présentons dans cette section comment modéliser une série, qui une fois tendance et saisonnalité supprimées, est stationnaire. A noter que le seul fait de supprimer la tendance et la saisonnalité ne rend pas la série nécessairement stationnaire, puisque cela n'affecte pas la variance et l'auto-covariance, qui doivent être constantes pour un processus stationnaire.

4.1 Auto-corrélation partielle

Le **coefficient de corrélation partielle** entre les deux variables X_1 et X_n d'un processus stochastique $(X_t)_t$ est le coefficient de corrélation entre les deux variables auxquelles on a retranché leurs meilleures explications en terme de X_2, \dots, X_{n-1} :

$$r_{X_2, \dots, X_{n-1}}(X_1, X_n) = \text{corr}(X_1 - P_{X_2, \dots, X_{n-1}}(X_1), X_n - P_{X_2, \dots, X_{n-1}}(X_n)),$$

où corr est le coefficient de corrélation classique (quotient de la covariance par le produit des écarts-types), et où $P_{X_2, \dots, X_{n-1}}(X_1)$ est la projection¹ de la variable X_1 dans l'espace vectoriel engendré par les variables X_2, \dots, X_{n-1} . Ce coefficient exprime la dépendance entre les variables X_1 et X_n qui n'est pas due aux autres variables X_2, \dots, X_{n-1} . La **fonction d'auto-corrélation partielle** $r(h)$ d'un processus stationnaire est définie par :

$$\begin{aligned} r(h) &= r_{X_2, \dots, X_h}(X_1, X_{h+1}) \quad \forall h \geq 2 \\ r(h) &= r(-h) \quad \forall h \neq 0 \\ r(1) &= \rho(1) \end{aligned}$$

L'algorithme de Durbin-Watson, que nous ne présentons pas ici, permet d'estimer les auto-corrélations partielles d'un processus stationnaire.

Dans le logiciel R, la fonction `pacf` permet ces estimations.

4.2 Les processus auto-régressifs AR_p

Les premiers modèles que nous présentons sont les processus auto-régressifs, construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes.

Définition 2. On dit que (X_t) est un processus auto-régressif d'ordre p (centré) s'il s'écrit

$$X_t = \epsilon_t + \sum_{j=1}^p a_j X_{t-j}, \quad (2)$$

où ϵ_t est un bruit blanc centré de variance σ^2 .

L'observation X_t au temps t est alors la somme d'un choc aléatoire à l'instant t , ϵ_t , indépendant de l'historique, et d'une fonction linéaire de son passé $\sum_{j=1}^p a_j X_{t-j}$, qui peut être vue comme la prédiction de X_t à partir des p dernières observations passées.

ϵ_t est l'**innovation** contenue dans le processus au temps t . On dit que (ϵ_t) est le processus d'innovation.

Les coefficients a_j doivent vérifier la contrainte suivante pour assurer la stationnarité du processus : le polynôme caractéristique du processus (2), $A(z) = 1 - a_1 z - \dots - a_p z^p$, ne doit avoir que des racines (réelles ou complexes) de module strictement supérieur à 1. La démonstration sort du cadre de ce cours.

Remarque 1 : en prenant l'espérance de (2) et en utilisant la stationnarité du processus, on obtient que l'espérance μ du processus vérifie $\mu(1 - \sum_{j=1}^p a_j) = 0$. Comme 1 ne peut être racine du polynôme $A(z)$, on a nécessairement $\mu = 0$.

Remarque 2 : dans ce cours, conformément à la définition (2), nous ne considérons que des processus AR_p centrés. Un processus AR_p non centré serait défini par $X_t = \mu + \epsilon_t + \sum_{j=1}^p a_j X_{t-j}$.

¹la projection $P_{X_2, \dots, X_{n-1}}(X_1) = \gamma_2 X_2 + \dots + \gamma_{n-1} X_{n-1}$ peut être interprétée comme la meilleure explication linéaire de X_1 en fonction de X_2, \dots, X_{n-1} .

Exercice 11. Vérifier la stationnarité et expliciter l'expression du processus autorégressif de polynôme caractéristique $A(z) = 1 - z - \frac{1}{2}z^2$.

Proposition 5. La variance du processus AR_p défini en (2) est :

$$\sigma(0) = \sigma^2 + \sum_{j=1}^p a_j \sigma(j)$$

et l'auto-covariance, pour tout $h > 0$:

$$\sigma(h) = \sum_{j=1}^p a_j \sigma(h-j).$$

Exercice 12. Faire la démonstration.

L'auto-covariance d'un processus AR_p vérifie donc la formule de récurrence d'ordre p suivante :

$$\sigma(h) - a_1 \sigma(h-1) - \dots - a_p \sigma(h-p) = 0. \quad (3)$$

Le polynôme caractéristique de cette relation de récurrence est :

$$z^p - a_1 z^{p-1} - \dots - a_{p-1} z - a_p = z^p \left(1 - \frac{a_1}{z} - \dots - \frac{a_p}{z^p} \right) = z^p A \left(\frac{1}{z} \right).$$

Les racines λ_i de ce polynôme sont les inverses des racines du polynôme A , et sont donc de module strictement inférieur à 1. En supposant que A a toute ses racines distinctes, les solutions de l'équation de récurrence précédente sont de la forme :

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

Ainsi, puisque $|\lambda_i| < 1$, **l'auto-covariance décroît exponentiellement avec h .**

Ces résultats s'étendent immédiatement à l'auto-corrélation d'un processus AR_p .

Quant à **l'auto-corrélation partielle, elle est nulle à tout ordre strictement supérieur à p , et vaut a_p à l'ordre p :**

$$\begin{aligned} r(h) &= 0 & \forall h > p, \\ r(p) &= a_p. \end{aligned}$$

La démonstration de ce résultat très utile ne sera pas abordée dans ce cours.

4.2.1 Exercice : cas particulier de l' AR_1

Soit le processus AR_1 (centré) suivant

$$X_t = aX_{t-1} + \epsilon_t,$$

où ϵ_t est un bruit blanc centré de variance σ^2 .

1. Quelle condition sur a doit on imposer pour que ce processus soit stationnaire ?
2. Calculer la variance de ce processus.
3. Montrer que l'auto-covariance d'un tel processus est

$$\sigma(h) = \sigma^2 \frac{a^h}{1 - a^2}$$

4. En déduire la décroissance vers 0 de l'auto-covariance lorsque h tend vers l'infini.
5. Calculer l'auto-corrélation partielle.

4.2.2 Illustrations

Voici quelques simulations de processus ainsi que leurs auto-corrélation et auto-corrélation partielle empiriques :

- AR_1 avec coefficient positif puis négatif : figures 17 et 18,
- AR_2 : figures 19 et 20.

4.3 Les processus en moyenne mobile MA_q

La seconde catégorie de modèles classiques regroupe les processus en moyenne mobile.

Définition 3. On appelle **moyenne mobile** (Moving Average) d'ordre q un processus de la forme

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

où les ϵ_j pour $t - q \leq j \leq t$ sont des bruits blancs centrés de variance σ^2 .

On notera parfois $X_t = \sum_{j=0}^q b_j\epsilon_{t-j}$ en imposant $b_0 = 1$.

A noter que pour l'instant aucune condition n'est nécessaire sur les b_i pour que le processus stationnaire ².

Un processus moyenne mobile est nécessairement centré, et son auto-covariance vérifie la proposition suivante.

Proposition 6. L'auto-covariance d'un processus moyenne mobile $X_t = \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q}$ est

$$\sigma(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} b_k b_{k+h} & \forall h \leq q \\ 0 & \forall h > q \end{cases} \quad \text{où } b_0 = 1$$

Exercice 13. Faire la démonstration (simple).

L'auto-corrélation est donc également nulle au dessus de l'ordre q . On retrouve le même comportement que l'auto-corrélation partielle d'un AR_p .

L'auto-corrélation partielle d'un processus moyenne mobile est compliquée à calculer, et sans grand intérêt. Néanmoins, il est important de savoir qu'elle tend vers 0 à vitesse exponentielle lorsque h tend vers l'infini.

Ce comportement symétrique entre les processus moyenne mobile et auto-régressif est du à la propriété suivante :

Proposition 7. Un processus auto-régressif est un processus moyenne mobile d'ordre infini, et réciproquement un processus moyenne mobile est un processus auto-régressif d'ordre infini.

Exercice 14. De la définition d'un processus moyenne mobile, extraire la valeur de l'innovation au temps t , puis ré-écrire la définition d'un MA_q en remplaçant les innovations par leur valeur. Conclure.

4.3.1 Exercice : cas particulier du MA_1

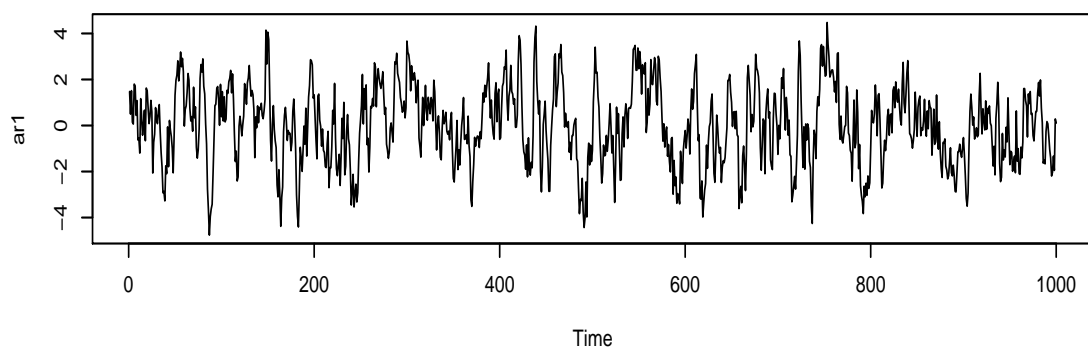
Calculer les coefficients d'auto-corrélation d'un tel processus.

4.3.2 Illustrations

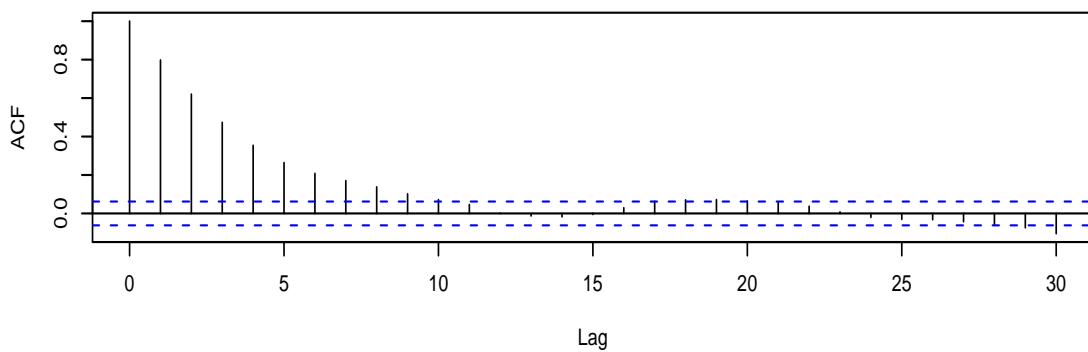
Voici quelques simulations de processus ainsi que leurs auto-corrélation et auto-corrélation partielle empiriques :

- MA_1 avec coefficient positif puis négatif : figures 21 et 22,
- MA_3 : figure 23.

²pour des questions de prévision, nous serons amené à supposer que le polynôme caractéristique $B(z) = 1 + b_1z + \dots + b_qz^q$ a toutes ses racines de module strictement supérieures à 1.



Series ar1



Series ar1

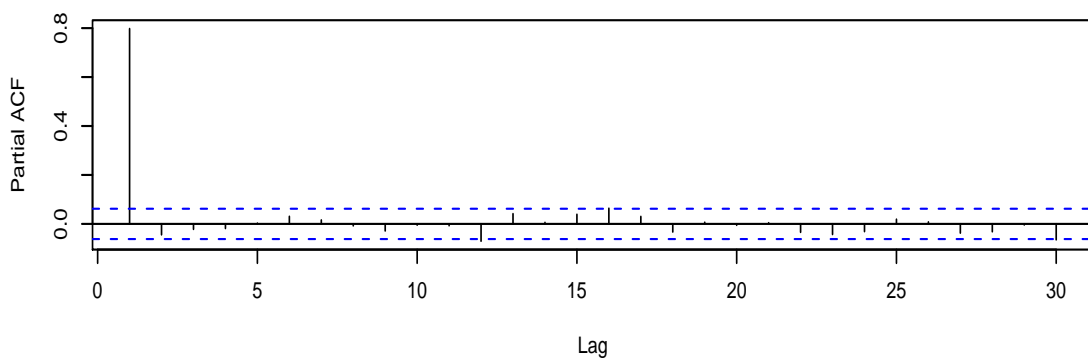
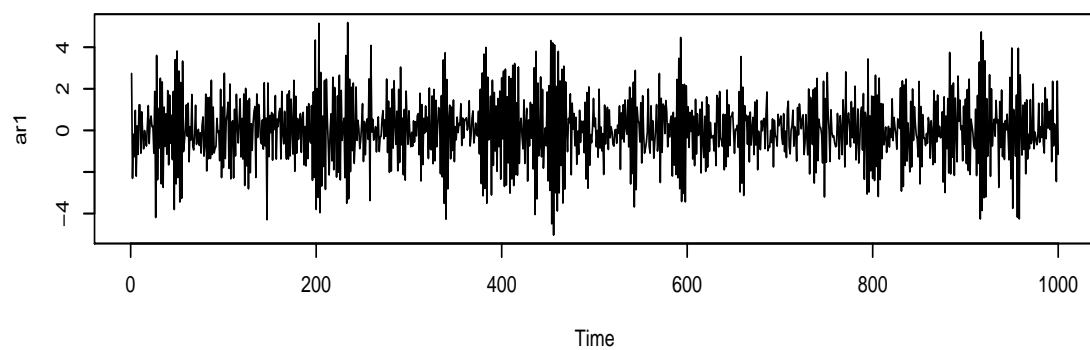
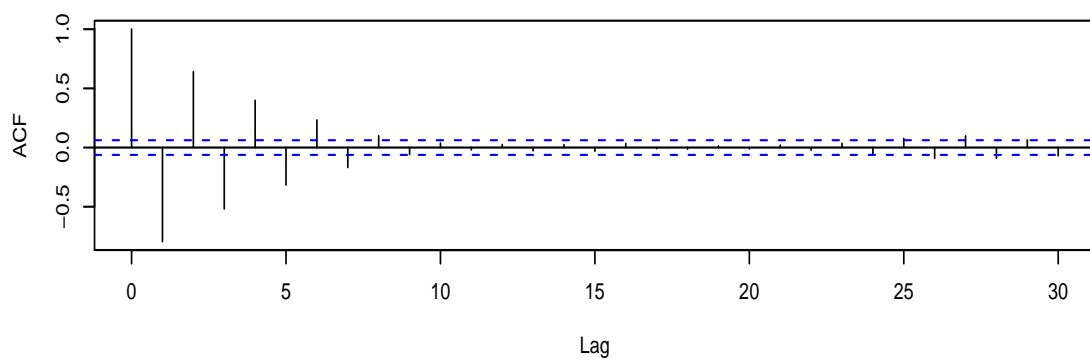


FIG. 17 – Simulation d'un $AR1$: $X_t = 0.8X_{t-1} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar1



Series ar1

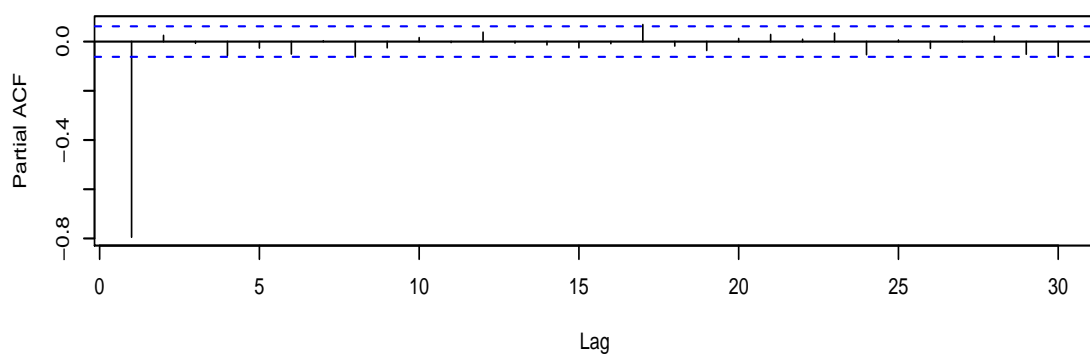
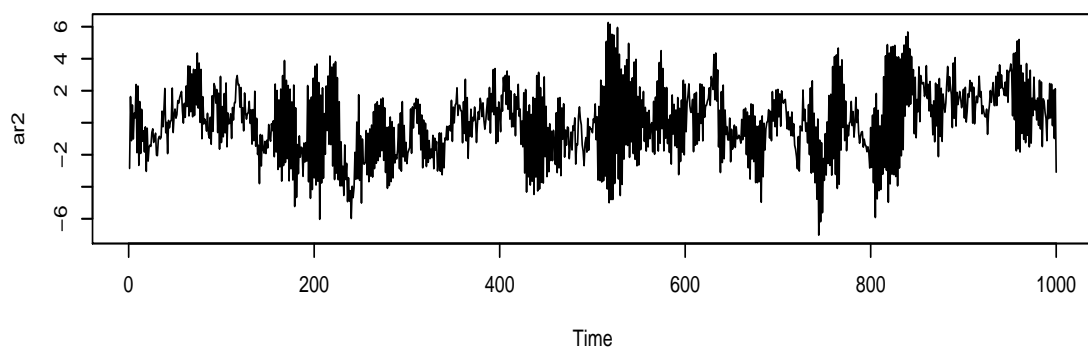
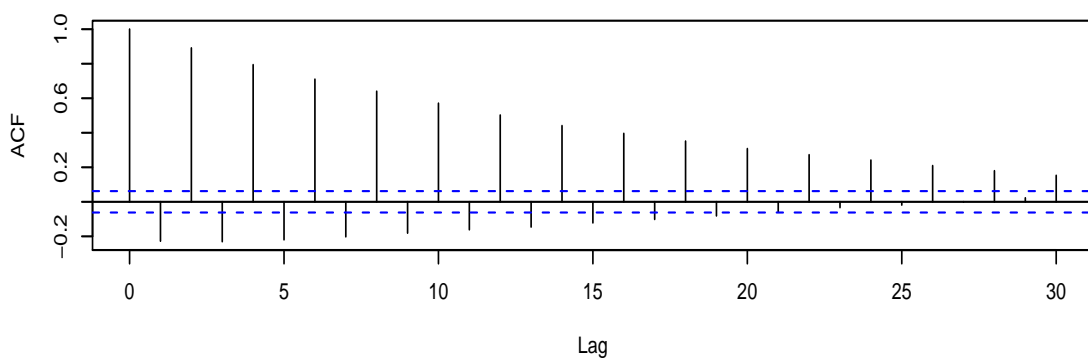


FIG. 18 – Simulation d'un $AR1$: $X_t = -0.8X_{t-1} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar2



Series ar2

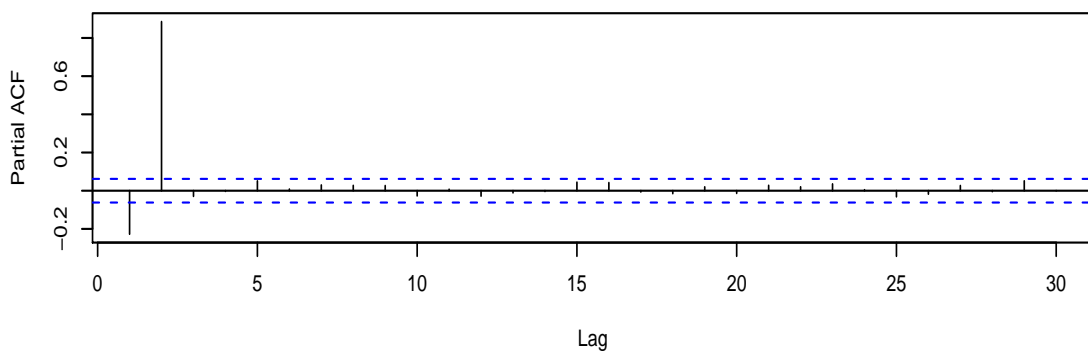
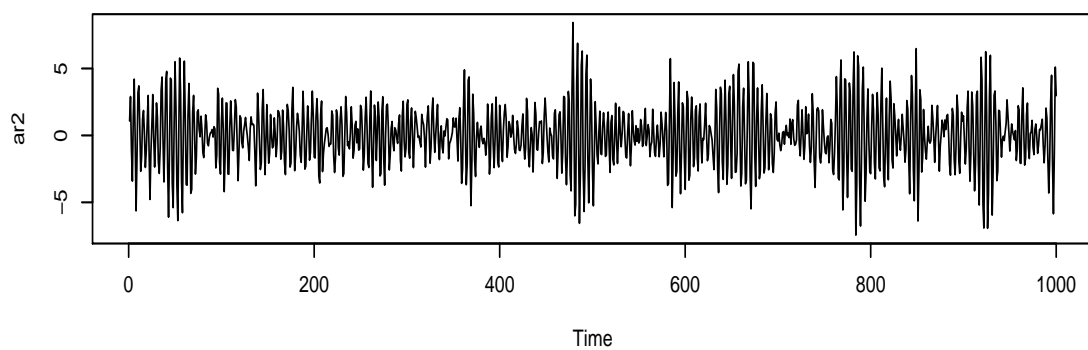
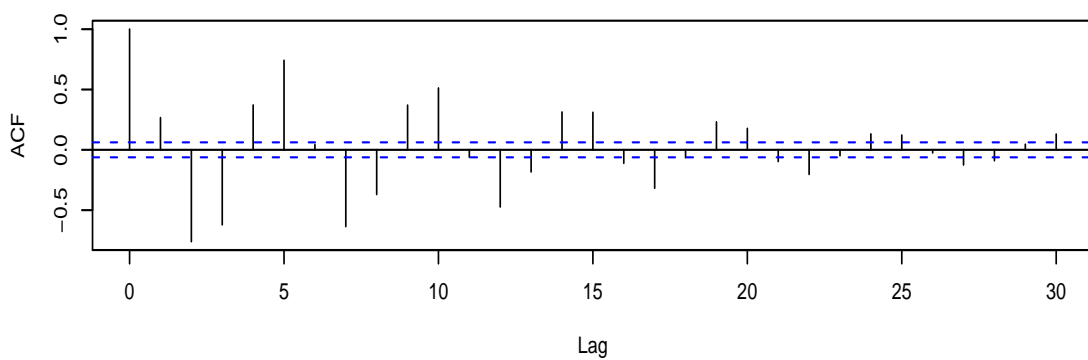


FIG. 19 – Simulation d'un AR_2 : $X_t = 0.9X_{t-2} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar2



Series ar2

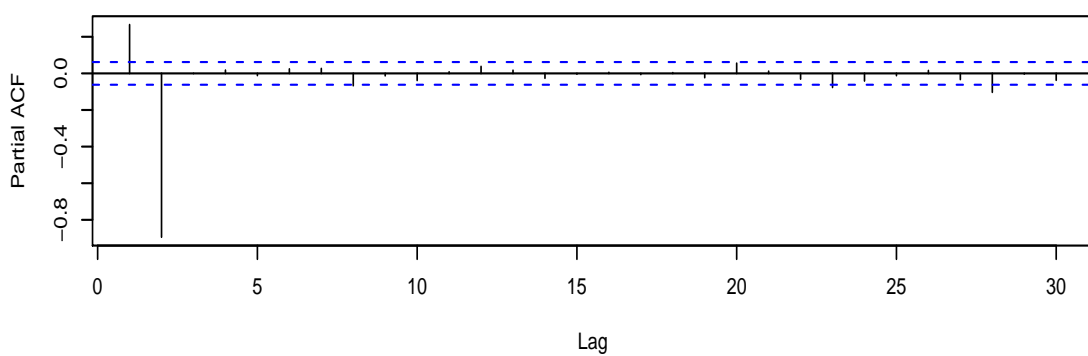
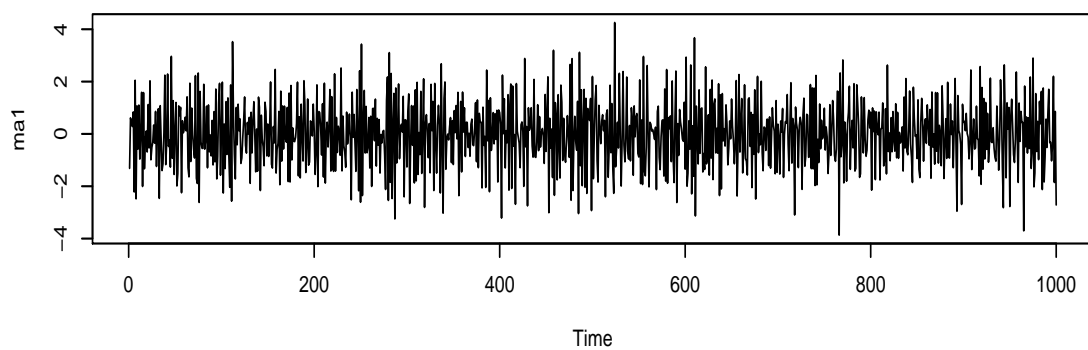
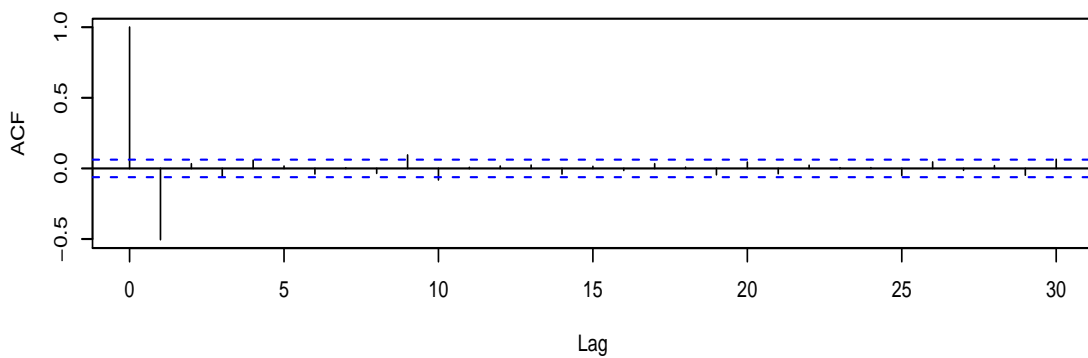


FIG. 20 – Simulation d'un AR_2 : $X_t = -0.5X_{t-1} - 0.9X_{t-2} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ma1



Series ma1

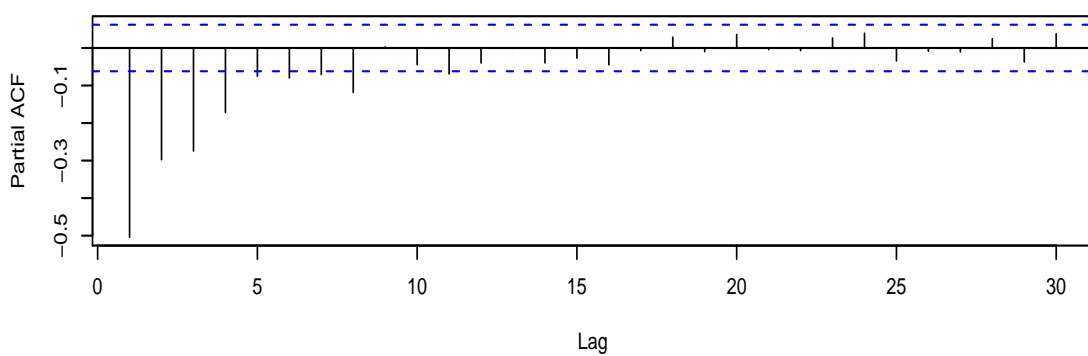
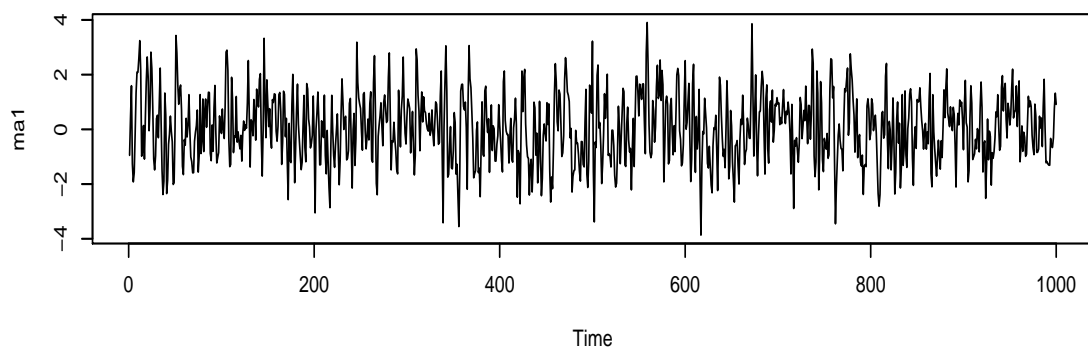
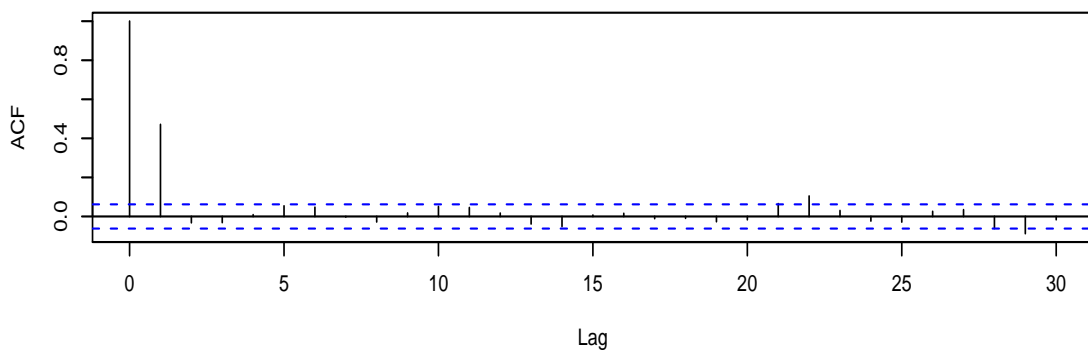


FIG. 21 – Simulation d'un MA_1 : $X_t = \epsilon_t - 0.8\epsilon_{t-1}$, auto-corrélation et auto-corrélation partielle.



Series ma1



Series ma1

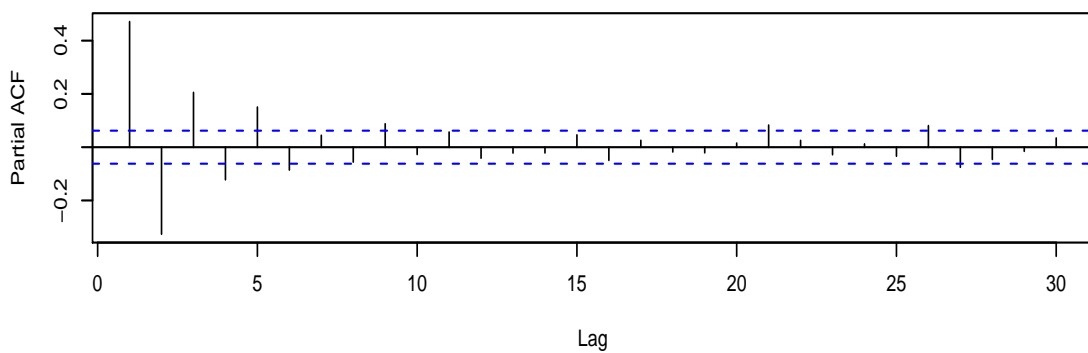
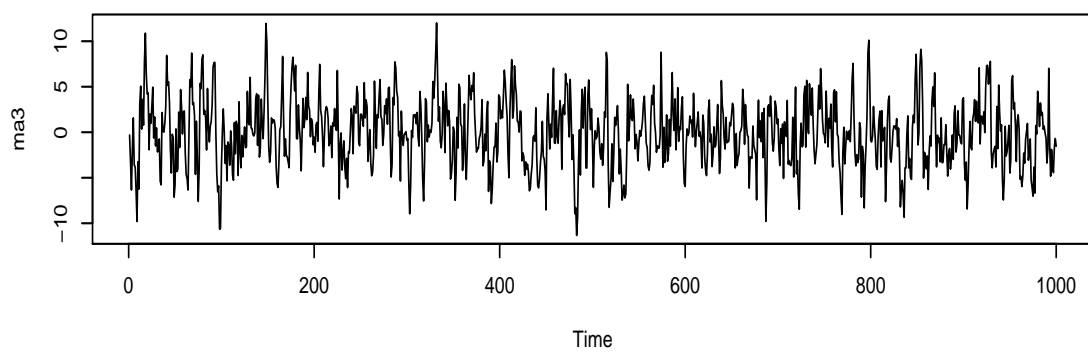
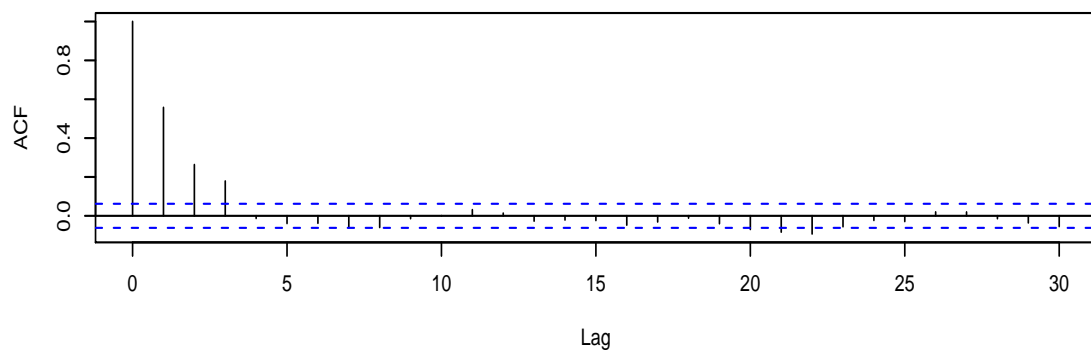


FIG. 22 – Simulation d'un MA_1 : $X_t = \epsilon_t + 0.8\epsilon_{t-1}$, auto-corrélation et auto-corrélation partielle.



Series ma3



Series ma3

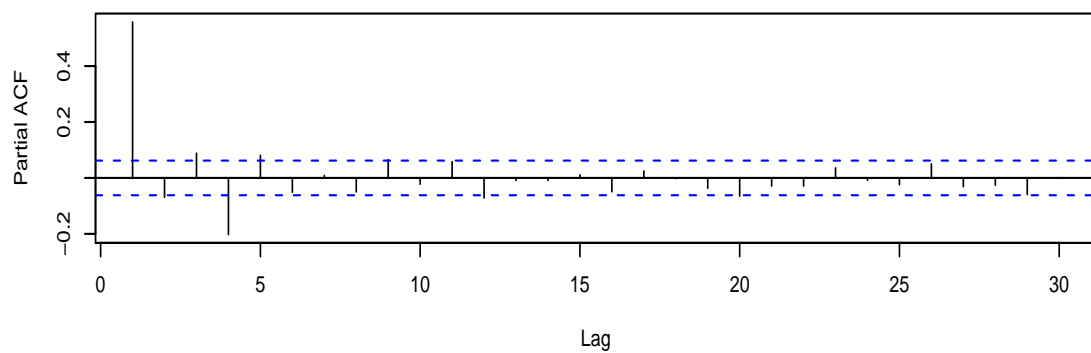


FIG. 23 – Simulation d'un MA_3 , auto-corrélation et auto-corrélation partielle.

4.4 Les processus mixtes $ARMA_{p,q}$

Cette classe plus générale de modèles définit des processus sous la forme d'une récurrence auto-régressive avec un second membre de type moyenne mobile.

Définition 4. Un processus auto-régressif moyenne mobile d'ordres p et q est de la forme :

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=0}^q b_j \epsilon_{t-j}.$$

où les ϵ_j pour $t - q \leq j \leq t$ sont des bruits blancs centrés de variance σ^2 .

La stationnarité d'un $ARMA_{p,q}$ est assurée lorsque toutes les racines du polynôme $A(z) = 1 - a_1 z - \dots - a_p z^p$ sont de module strictement supérieur à 1. Ce polynôme forme avec $B(z) = 1 + b_1 z + \dots + b_q z^q$ les deux polynômes caractéristiques du processus. On supposera également que les polynômes A et B n'ont pas de racine commune, afin de s'assurer qu'il n'y a pas de représentation plus courte.

Exercice 15. Soit le processus $ARMA_{1,1}$ défini par $X_t + aX_{t-1} = \epsilon_t + a\epsilon_{t-1}$. Est-ce un processus stationnaire ? Existe-t-il une écriture plus simple de ce processus ?

On peut écrire le processus $ARMA_{p,q}$ sous la forme suivante

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

Ainsi,

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

qui est nulle dès que $h > q$. L'auto-covariance d'un processus $ARMA_{p,q}$ évolue selon la même récurrence qu'un AR_p (cf. équation (3)). Ainsi, l'auto-covariance (et l'auto-corrélation) d'un processus $ARMA_{p,q}$ vont tendre exponentiellement vers 0 lorsque h tend vers l'infini, à partir de l'ordre $q + 1$.

4.4.1 Exercice : le processus $ARMA_{1,1}$

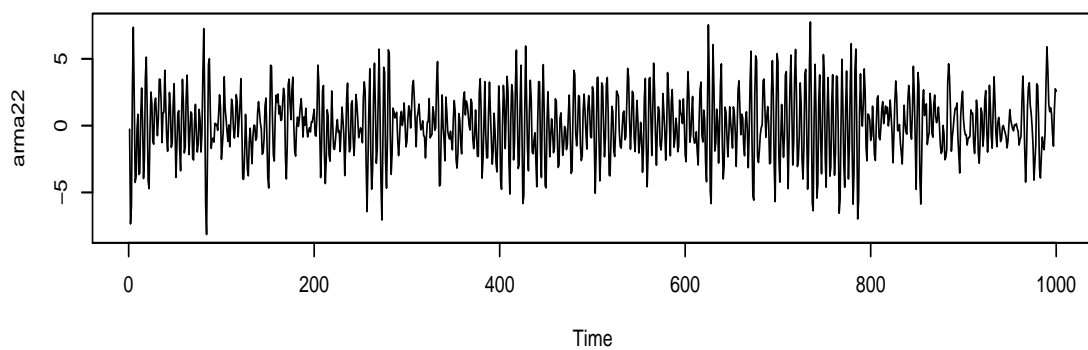
Considérons le processus

$$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}.$$

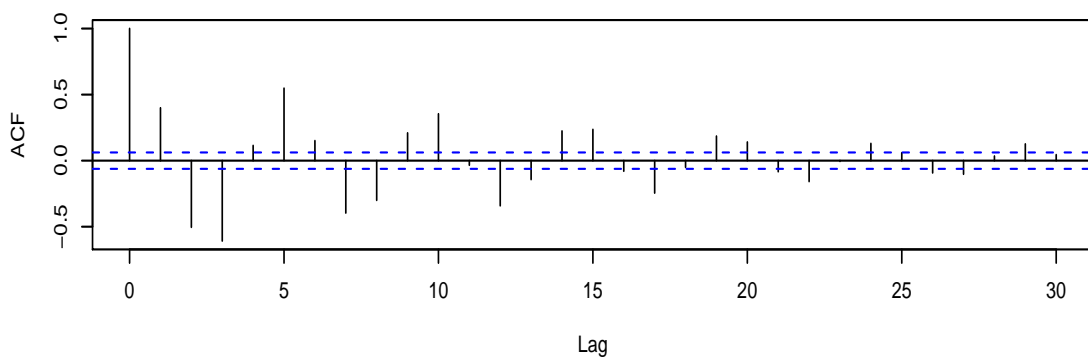
1. Montrer que la variance du processus est $\sigma(0) = \sigma^2 \frac{1+b^2+2ab}{1-a^2}$.
2. Montrer que l'auto-covariance d'ordre 1 est $\sigma(1) = \sigma^2 \frac{a+b+ab^2+a^2b}{1-a^2}$.
3. En déduire les auto-corrélations $\rho(1) = \frac{(a+b)(1+ab)}{b^2+2ab+1}$ et $\rho(h) = a^{h-1} \rho(1)$.

4.4.2 Illustrations

Voici une simulation d'un processus $ARMA_{2,2}$ (figure 24) ainsi que ses auto-corrélation et auto-corrélation partielle empiriques.



Series arma22



Series arma22

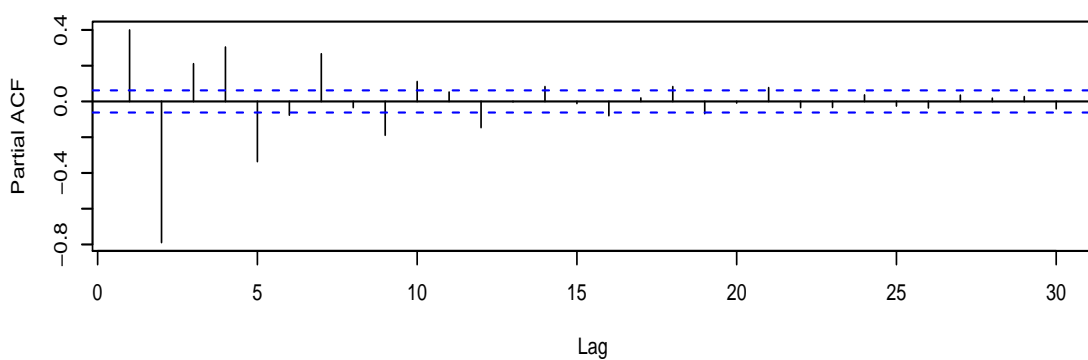


FIG. 24 – Simulation d'un $ARMA_{2,2}$, auto-corrélation et auto-corrélation partielle.

4.5 Récapitulatif des propriétés des processus MA_q , AR_p et $ARMA_{p,q}$

La tableau 4.5 récapitule les principales propriétés des processus MA_q , AR_p et $ARMA_{p,q}$.

modèle	MA_q	AR_p	$ARMA_{p,q}$
auto-covariance	$\sigma(h) = 0 \ \forall h > q$	$\lim_{h \rightarrow \infty} \sigma(h) = 0$	$\forall h > q, \lim_{h \rightarrow \infty} \sigma(h) = 0$
auto-corrélation	$\rho(h) = 0 \ \forall h > q$	$\lim_{h \rightarrow \infty} \rho(h) = 0$	$\forall h > q, \lim_{h \rightarrow \infty} \rho(h) = 0$
auto-corrélation partielle	$\lim_{h \rightarrow \infty} r(h) = 0$	$r(h) = 0 \ \forall h > p \text{ et } r(p) = a_p$	

TAB. 1 – Récapitulatif des propriétés des processus MA_q , AR_p et $ARMA_{p,q}$.

4.6 Estimation, choix de modèle et prévision

Les principaux modèles de séries temporelles ont été définis. A partir d'une série observée, il faut maintenant choisir un modèle, éventuellement plusieurs, estimer ses paramètres et enfin faire des prévisions pour les réalisations futures. Dans le cas où l'on hésite entre plusieurs modèles, des critères de choix de modèles seront utilisés pour sélectionner le meilleur d'entre eux.

4.6.1 Estimation

Exemple Considérons le processus AR_2 suivant :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t.$$

On peut montrer que les auto-covariances d'ordre 1 et 2 sont les suivantes :

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

Exercice 16. Montrez-le !

Ainsi, les paramètres du modèle a_1 et a_2 peuvent être exprimés en fonction de la variance et des 2 premières auto-covariance du processus :

$$a_1 = \frac{\sigma(1)}{\sigma(0)} \frac{\sigma(0)^2 - \sigma(0)\sigma(1)}{\sigma(0)^2 - \sigma(1)^2} \quad \text{et} \quad a_2 = \frac{\sigma(0)\sigma(2) - \sigma(1)^2}{\sigma(0)^2 - \sigma(1)^2}$$

Exercice 17. Montrez-le également !

En estimant les auto-covariances du processus à l'aide des auto-covariances empiriques, on en déduit des estimateurs des paramètres du modèle AR_2 .

Comme de plus $\sigma(0) = \sigma^2 + a_1 \sigma(1) + a_2 \sigma(2)$, il est possible d'obtenir également un estimateur de la variance du bruit d'innovation.

Cas général En réalité, les estimateurs précédents peuvent être améliorés : ainsi, dans le cas général, l'estimation des paramètres des modèles ARMA (AR, MA) est faite par maximum de vraisemblance. L'expression de la vraisemblance étant généralement trop complexe pour que l'on puisse obtenir un maximum explicite, des algorithmes numériques (type Newton) sont utilisés.

4.6.2 Choix de modèle

Comme on l'a vu précédemment (cf. tableau récapitulatif 4.5) l'étude des auto-covariances, auto-corrélations et auto-corrélations partielles peut conduire à certaines hypothèses sur la nature du modèle. Une fois quelques modèles choisis, et leur paramètres estimés, des critères vont être utilisés pour choisir le modèle qui effectue le meilleur compromis entre :

- ajustement à la série de données,
- complexité du modèle.

Il est en effet très important de prendre en compte ce compromis, car si on ne s'intéressait qu'à coller au mieux aux données, on serait tenté de choisir un modèle *ARMA* avec un très grand nombre de paramètres. Or, plus il y a de paramètres, plus il faut de données pour les estimer. Et donc pour un nombre d'observations fixé de la série, plus le modèle sera complexe, moins bien seront estimés les paramètres.

Les critères de choix de modèles les plus courants sont :

- le critère AIC (*Akaike Information Criterion*), qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision, et qui est défini par :

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

où $L(\cdot)$ est la vraisemblance du modèle, θ représente les paramètres du modèle et ν le nombre de ces paramètres.

- le critère BIC (*Bayesian Information Criterion*) sera quant à lui généralement préféré si l'objectif de l'étude est de s'ajuster à la série observée, et est défini par :

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

où n est le nombre d'observations de la série.

Les modèles ayant la plus petite valeur du critère devront être choisis.

Ces deux critères conduisent donc à sélectionner des modèles dont la vraisemblance est grande, en la pénalisant par la complexité du modèle.

4.6.3 Prévision

L'objectif est de prévoir la valeur que va prendre la variable aléatoire X_{n+h} , h étant appelé l'horizon de la prévision, ayant observé la réalisation des variables aléatoires X_1, \dots, X_n .

Dans le cadre d'une modélisation ARMA, on choisit d'estimer X_{n+h} par une combinaison linéaire des X_j précédents ($1 \leq j \leq n$) :

$$\hat{X}_{n,h} = c_{1,h}X_1 + \dots + c_{n,h}X_n.$$

Les coefficients $c_{j,h}$ sont estimés de sorte à ce qu'il minimise :

$$E[(X_{n+h} - c_{1,h}X_1 - \dots - c_{n,h}X_n)^2].$$

L'estimateur $\hat{X}_{n,h}$ ainsi défini n'est autre que la projection de X_{n+h} sur l'espace vectoriel engendré par les variables X_1, \dots, X_n .

Exercice 18. Montrer que la prévision au rang 1 d'un processus AR_p défini par l'équation 2 n'est autre que $\hat{X}_{n+1} = a_1X_n + \dots + a_pX_{n+1-p}$. En déduire que l'erreur de prévision à horizon 1 est le bruit d'innovation.

Nous ne détaillerons pas plus en détail cette partie dans ce cours, mais nous citons néanmoins, sans démonstrations, deux propriétés intéressantes.

Proposition 8. – L'erreur de prévision à horizon 1 pour un processus ARMA est le bruit d'innovation ϵ_{n+1} .
 – La variance de l'erreur de prévision à horizon h dans un processus ARMA croît depuis la variance du bruit d'innovation (valeur prise pour $h = 1$) jusqu'à la variance du processus lui-même.

Intervalle de confiance sur la prévision Enfin, si on suppose que le bruit d'innovation ϵ_t est gaussien, les variables aléatoires X_t sont elles aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, il sera possible de construire des intervalles de confiances sur la prédiction, ce que permettent les fonctions du logiciel R décrites plus loin.

5 Processus non stationnaire : ARIMA et SARIMA

Les modèles ARMA sont destinés à modéliser des processus stationnaires. En pratique, les séries temporelles sont généralement non stationnaires, et un pré-traitement est nécessaire pour supprimer les tendances et saisonnalités. Une fois la série *stationnarisée* analysée, et les valeurs futures prédites, il est ensuite nécessaire de revenir à la série initiale.

5.1 Exemple

Considérons une série temporelle $(x_t)_{1 \leq t \leq n}$ qui présente une saisonnalité de période 12. Afin de supprimer cette saisonnalité, c'est la série suivante que nous étudions :

$$y_t = x_t - x_{t-12} \quad \text{pour } t = 13, \dots, n \quad (4)$$

L'ajustement d'un modèle ARMA et les prévisions sont réalisées sur la série (y_t) . Il est donc nécessaire de revenir à la série initiale, car ce sont les prévisions de (x_t) qui nous intéressent.

De l'équation (4) on obtient :

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

où $r(t)$ est le reste de la division euclidienne de t par 12.

Puisque l'on connaît les x_t pour $t \leq n$ ainsi que les prévisions y_{n+h} , on peut en déduire les prévisions de x_{n+h} .

Exercice 19. Ecrire pour cet exemple la prévision à l'horizon 1 de la série $(x_t)_{1 \leq t \leq n}$.

5.2 Les processus ARIMA et SARIMA

Les processus ARIMA et SARIMA sont la généralisation des modèles ARMA pour des processus non stationnaires, admettant une tendance (ARIMA), ou encore une tendance et une saisonnalité (SARIMA). En pratique, et dans le logiciel R notamment, ce sont ces processus qui sont directement utilisés.

Soit Δ l'opération de différenciation précédemment défini (section 3), qui associe au processus $(X_t)_{t \in \mathbb{N}}$ le processus $(X_t - X_{t-1})_{t \in \mathbb{N}}$. Nous rappelons que le processus obtenu en différenciant deux fois de suite, $(X_t - 2X_{t-1} + X_{t-2})_{t \in \mathbb{N}}$, est noté Δ^2 . Et ainsi de suite.

Définition 5. Un processus (X_t) est un processus $ARIMA_{p,d,q}$ si le processus

$$Y_t = \Delta^d X_t$$

est un processus $ARMA_{p,q}$.

Les processus $ARIMA_{p,d,q}$ sont donc bien adaptés aux séries temporelles présentant une tendance polynômiale de degré $d - 1$.

Soit Δ_T l'opérateur précédemment défini (section 3), qui fait passer de (X_t) à $(X_t - X_{t-T})$.

Définition 6. Un processus (X_t) est un processus $SARIMA_{p,d,q,T}$ si le processus

$$Y_t = \Delta_T \circ \Delta^d X_t$$

est un processus $ARMA_{p,q}$.

Les processus $SARIMA_{p,d,q,T}$ sont bien adaptés aux séries temporelles présentant une période de longueur T et une tendance polynômiale de degré d .

Remarque : en réalité, les processus SARIMA sont plus complexes et comportent d'autres paramètres. Nous ne considérerons que cette version $SARIMA_{p,d,q,T}$ dans ce cours, se référer à [4] pour des compléments sur ces processus.

5.3 Mise en oeuvre sous R : processus ARMA, ARIMA, SARIMA

La fonction `arima.sim(modele,n)` permet de simuler un processus $ARMA_{p,q}$ défini par

$$X_t - \sum_{k=1}^p a_k X_{t-k} = \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}.$$

Les paramètres a_k et b_j du processus sont précisés dans le paramètre `modele` de la fonction :

`modele<-list(ar=c(a1,...,ap),ma=c(b1,...,bq)).`

Pour simuler un modèle $ARIMA_{p,d,q}$ il faut ajouter le composant `order=c(p,d,q)` dans le paramètre `modele` de la fonction `arima.sim`.

La fonction `ar` permet d'estimer les paramètres d'un processus AR_p :

`out<-ar(data,aic=TRUE,order.max=NULL)`

L'ordre p du processus auto-régressif est choisi (inférieur à `order.max`) à l'aide du critère *AIC* (si l'option *aic* est validée).

La fonction `arima` permet d'estimer les paramètres :

- d'un $ARMA_{p,q}$: `out<-arima(serie,order=c(p,0,q))`
- d'un $ARIMA_{p,d,q}$: `out<-arima(serie,order=c(p,d,q))`
- d'un $SARIMA_{p,d,q,T}$:
`out<-arima(serie,order=c(p,d,q),seasonal=list(order=c(P,D,Q),period=T))`

Les paramètres P, D, Q du modèle SARIMA ne sont pas abordés dans ce cours, nous leur donnerons par défaut la valeur des paramètres p, d, q (pas trop grands).

Parmi les sorties de cette fonction, on peut obtenir :

- `out$coef` : estimation des coefficients,
- `out$aic` : valeur du critère AIC,
- `out$resid` : estimation des résidus.

La fonction `p=predict(out,h)` permet d'effectuer une prévision à l'horizon h . Parmi les sorties de cette fonction, `p$pred` contient les prévisions, et `p$se` contient l'écart-type de l'erreur de prévision. Il n'existe pas de fonction prédéfinie pour calculer un intervalle de confiance sur les prévisions, mais cela peut être fait manuellement grâce à ces deux sorties de la fonction `predict`.

6 Processus ARCH et GARCH

Les séries précédemment étudiées étaient supposées stationnaires. Si besoin, tendances et saisonnalités étaient supprimées pour obtenir une série résiduelle stationnaire. Néanmoins, toutes les séries résiduelles obtenues de la sorte ne sont pas nécessairement stationnaires.

C'est le cas par exemple de la série représentée par la figure 25, qui contient les évolutions journalières de la bourse des valeurs de New-York (NYSE) du 19 octobre 1984 au 31 décembre 1991.

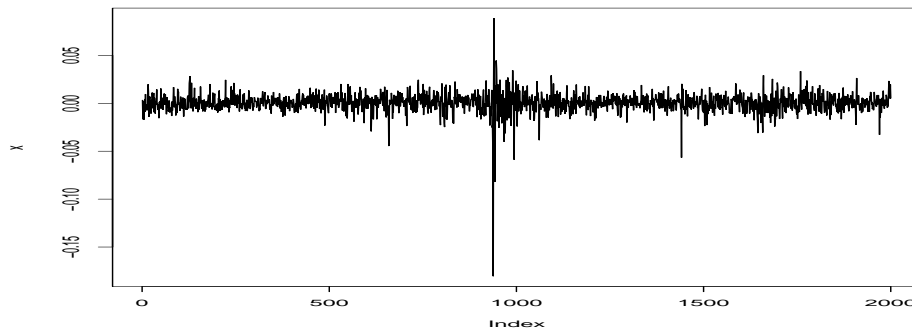


FIG. 25 – Evolution journalière de la bourse des valeurs de New-York (1984-1991)

La figure 26 représente la simulation d'un processus $ARCH_2$.

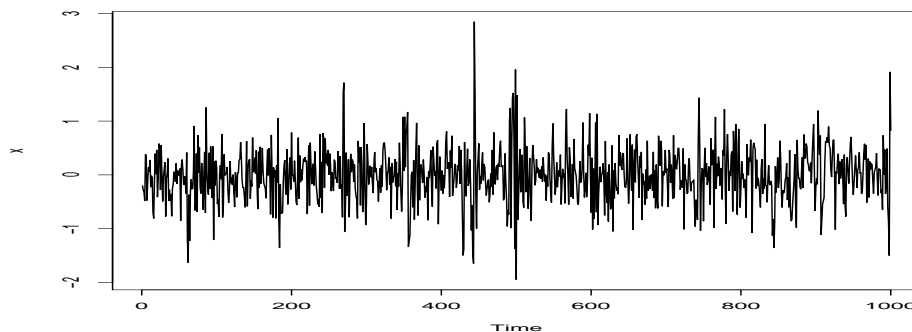


FIG. 26 – Simulation d'un $ARCH_2$

Comme on peut le voir, la moyenne semble constante alors que la variance change au cours du temps (on qualifie ce comportement d'« hétéroscédastique »). De plus, les moments de grande variabilité semblent regroupés. Les modèles de type ARIMA qui supposent un comportement « homoscédastique » (variance constante), ne sont pas adaptés à ce type de série.

Nous présentons dans cette section des modèles adaptés à ce type de série : les processus ARCH (*AutoRegressive Conditionally Heteroscedastic*) introduits par Engle vers 1982, ainsi que leur généralisation, les processus GARCH.

6.1 Définitions

Un processus $ARCH_p$ est défini par :

$$\begin{aligned} X_t &= \epsilon_t \\ \epsilon_t | X_{t-1}, X_{t-2}, \dots &\sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 \end{aligned}$$

La variance conditionnelle dépend du temps : si les valeurs précédentes sont grandes (en valeur absolue), la variance sera grande, et inversement. Ainsi, si on observe un choc dans la série (valeur anormalement grande), elle sera suivie d'une période de haute volatilité, dont la durée dépend de l'ordre p du modèle ARCH.

Exercice 20. Pensez-vous que cela correspond à la série NYSE (figure 25) ?

6.2 Quelques rappels de probabilités

Soit un couple de variables aléatoires continues (X, Y) , de densité $f(., .)$. Soit $f_Y(.)$ la densité marginale de Y . La densité conditionnelle de X sachant que $Y = y$ est définie par :

$$f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)} \quad \text{si } f_Y(y) > 0$$

L'espérance conditionnelle est donc l'espérance par rapport à cette loi :

$$E[X|Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x, y) dx$$

La variance conditionnelle est

$$\sigma_t^2 = V(X_t | \mathcal{I}_{t-1}) = E[X_t^2 | \mathcal{I}_{t-1}] - E[X_t | \mathcal{I}_{t-1}]^2$$

Soit $\mathcal{I}_{s-1} = \{X_{s-1}, X_{s-2}, \dots\}$ l'ensemble des observations précédant l'instant s .

On a :

$$\begin{aligned} E[X_t | \mathcal{I}_s] &= X_t \quad \forall t \leq s \\ E[X_t | \mathcal{I}_r] &= E[E[X_t | \mathcal{I}_s] | \mathcal{I}_r] \quad \forall r \leq s \end{aligned}$$

et en particulier $E[X_t] = E[E[X_t | \mathcal{I}_s]]$.

6.3 Propriétés des processus ARCH

Si X_t est un processus ARCH, alors :

$$\begin{aligned} E[X_t] &= 0 \\ E[X_t | \mathcal{I}_{t-1}] &= 0 \quad \text{où } \mathcal{I}_{t-1} = X_{t-1}, X_{t-2}, \dots \\ V(X_t) &= \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i} \quad \text{si } \sum_{i=1}^p \alpha_i < 1 \\ V(X_t | \mathcal{I}_{t-1}) &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 \\ \text{Cov}(X_t, X_{t+h}) = \sigma_h &= 0 \quad \forall h > 0 \\ \text{Cov}(X_t, X_{t+h} | \mathcal{I}_{t-1}) &= 0 \end{aligned}$$

Remarque : un processus ARCH est conditionnellement hétéroscédastique mais inconditionnellement homoscedastique !

Condition suffisante de stationnarité : $\sum_{i=1}^p \alpha_i < 1$.

On peut montrer également, ce qui peut être intéressant pour détecter un ARCH en pratique, que la distribution d'un processus ARCH a

- un *skewness* nul (moment centré d'ordre 3) : la distribution est donc symétrique,
- un *kurtosis* (moment centré d'ordre 4) supérieur à 3 : la distribution est donc plus aplatie qu'une gaussienne.

6.4 Processus GARCH et propriétés

Un processus $GARCH_{p,q}$ est défini par :

$$\begin{aligned}X_t &= \epsilon_t \\ \epsilon_t | X_{t-1}, X_{t-2}, \dots &\sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2\end{aligned}$$

avec $\alpha_0 > 0$, $\alpha_i \geq 0$ pour $i = 1, \dots, p$ et $\beta_j \geq 0$ pour $j = 1, \dots, q$.

Un processus GARCH peut être vu comme un processus ARCH d'ordre infini, et peut ainsi représenter formellement de façon plus parcimonieuse un processus ARCH comprenant un nombre élevé de paramètres.

Remarque évidente : un $GARCH_{p,0}$ est un $ARCH_p$.

Si X_t est un processus $GARCH_{p,q}$, alors :

$$\begin{aligned}E[X_t] &= 0 \\ E[X_t | \mathcal{I}_{t-1}] &= 0 \\ \text{Cov}(X_t, X_{t+h}) &= \sigma_h = 0 \quad \forall h > 0 \\ \text{Cov}(X_t, X_{t+h} | \mathcal{I}_{t-1}) &= 0\end{aligned}$$

Propriété : soit X_t un processus $GARCH_{p,q}$, et soit $m = \max(p, q)$. Le processus X_t^2 admet une représentation $ARMA(m, q)$.

Ainsi, pour identifier un $GARCH_{p,q}$, on identifiera tout d'abord le processus $ARMA(m, q)$ qui modélise X_t^2 . Pour identifier p dans le cas où $m = q$ ($p \leq q$), il faut effectuer des tests de significativité des coefficients $\alpha_q, \dots, \alpha_1$ du processus $ARMA(m, q)$ (sont-ils significativement non nuls ?).

6.5 Mise en oeuvre sous R

Pour utiliser les fonctions spécifiques à l'étude des modèles ARCH et GRACH, il faut avant tout charger le package `tseries` à l'aide de la commande `library(tseries)`.

La fonction `garch` permet d'estimer un $GARCH_{p,q}$: `serie<-garch(data, order=c(q,p))`

Parmi les sorties de cette fonction : `coef`, `residuals`, `fitted.values`.

La prédiction se fait de la même façon que pour les modèles de type ARMA.

Références

- [1] Bensaber A. et Bleuse-Trillon B.. *Pratique des chroniques et de la prévision à court terme*, Masson, Paris, 1989.
- [2] Bosq D. et Lecoutre J-P. *Analyse et prévision des séries chronologiques*, Masson, Paris, 1992.
- [3] Brockwell P.J. et Davis R.A. *Introduction to Time Series and Forecasting*, Springer, 2001.
- [4] Gouriéroux C. et Montfort A. *Cours de séries temporelles*, Economica, Paris, 1983.
- [5] Shumway R.H. et Stoffer D.S. *Times Series Analysis and Its Applications, With R Example*, Springer, 2006.

Ce cours est en partie inspiré du cours de M.-Cl. Viano disponible à l'adresse suivante :

<http://math.univ-lille1.fr/~viano/economcours.pdf>