

Séries chronologiques (avec R)
(Cours et exercices)
M1 IM, 2018-2019



Sylvain Rubenthaler

Table des matières

Préface	iii
Chapitre 1. Introduction	1
1.1. Tendances et composantes saisonnières	2
1.2. Indices descriptifs d'une série temporelle	2
1.3. Feuille d'exercices numéro 1 (durée : 3h)	4
1.4. Corrigé de la feuille d'exercices numéro 1 (qui constitue un exemple de ce qui est attendu aux contrôles sur machine)	5
Chapitre 2. Lissages exponentiels	15
2.1. Lissage exponentiel simple	15
2.2. Lissage exponentiel double	16
2.3. Méthode de Holt-Winters	18
2.4. Feuille d'exercices numéro 2 (durée : 3h)	19
Chapitre 3. Estimation et élimination de la tendance et de la saisonnalité	23
3.1. Bruit blanc	23
3.2. Processus stationnaire	23
3.3. Estimation paramétrique de la tendance	23
3.4. Estimation non paramétrique : moyenne mobile	26
3.5. Élimination de la tendance et de la saisonnalité par la méthode des différences	27
3.6. Test sur la série résiduelle	29
3.7. Exemple : un système proies-prédateurs	30
3.8. Feuille d'exercices numéro 3 (durée : 3h)	31
Chapitre 4. Modélisation des séries stationnaires	35
4.1. Auto-corrélation partielle	35
4.2. Les processus auto-régressifs	35
4.3. Les processus en moyenne mobile	39
4.4. Les processus mixtes ARMA(p,q).	39
4.5. Tableau des propriétés	42
4.6. Estimation, choix de modèle et prévisions	43
4.7. Processus non stationnaires : ARIMA et SARIMA	44
4.8. Feuille d'exercices numéro 4 (durée : 6h)	53
4.9. Corrigé de la feuille d'exercices numéro 4	54
4.10. Feuille d'exercices numéro 5 (durée : 6h)	60
Chapitre 5. Analyse spectrale	63
5.1. Densité spectrale	63
5.2. Le périodogramme	66
5.3. Récupération des composantes périodiques	67
5.4. Feuille d'exercices numéro 6 (durée : 3h)	67
Chapitre 6. Processus ARCH et GARCH	71
6.1. Processus ARCH	71
6.2. Processus GARCH	73

6.3. Feuille d'exercices numéro 7 (durée : 3h)	74
6.4. Feuille d'exercices numéro 8 (révisions)	74
6.5. Corrigé de la feuille d'exercices numéro 8	76
Table de la loi normale	83
Bibliographie	85
Liste des symboles	87
Index	89

Préface

Ce polycopié s'inspire fortement de [**Jac, OPV**]. Les TP se feront en **R**, les exemples de programmes seront aussi donnés en **R**. Les corrigés des exercices sur tables sont inclus dans ce polycopié. Pour les corrigés des exercices sur ordinateur : voir sur internet.

Prérequis : cours de L3 MASS d'introduction aux séries chronologiques et cours de L3 MASS de probabilités.

Important : les fichiers sources sont disponibles sur : <http://www.math.unice.fr/~rubentha/enseignement>. J'encourage toute personne enseignant ce cours à utiliser ces fichiers et à ajouter son nom à la liste d'auteurs de ce polycopié.

La première utilisation en cours de ce polycopié est prévue pour 2016-2017. Il va de soi que le nombre de coquilles ira en décroissant avec les années.

Introduction

DÉFINITION 1.1. Une série temporelle (ou série chronologique) est une suite réelle finie $(x_t)_{1 \leq t \leq n}$ ($n \in \mathbb{N}^*$). L'indice t représente une unité de temps (qui peut être le mois, l'année ...).

EXEMPLE 1.2. La figure 1.0.1 représente le total mondial des passagers aériens par mois entre 1949 et 1960. Noter que les points sont reliés par des traits (qui sont là pour faire joli et n'ont pas de signification particulière). Les données (**AirPassengers**) sont disponibles dans R.

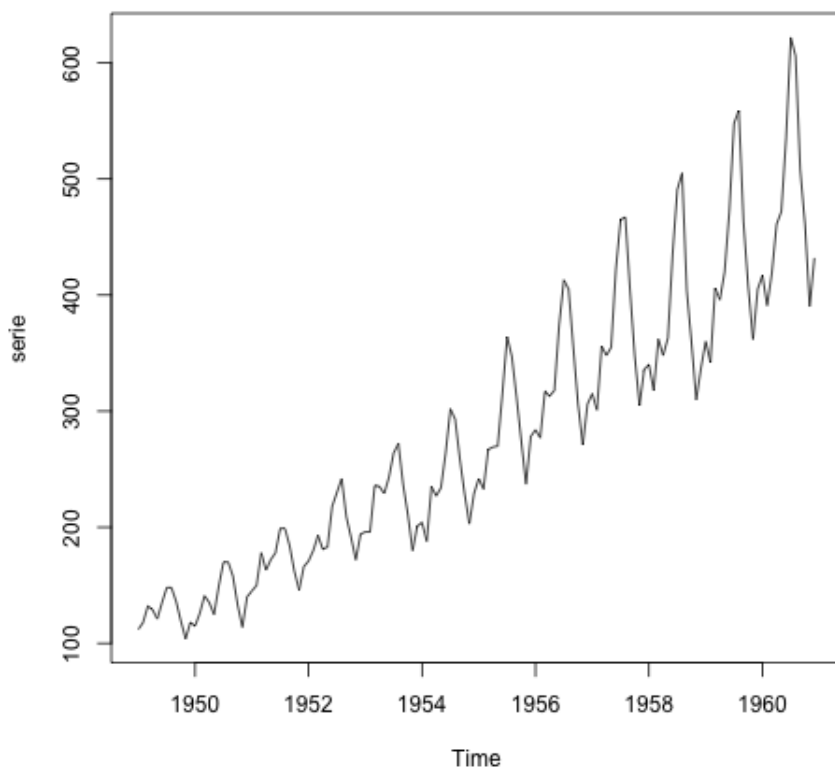


FIGURE 1.0.1. **AirPassengers**

L'objectif de l'étude des séries temporelles est de faire des prédictions sur l'évolution de la série. Voici une liste non-exhaustive des modèles mathématiques que l'on pourra utiliser :

- Régression. On suppose que x_t est polynomial en t , par exemple $x_t = \alpha_2 t^2 + \alpha_1 t + \alpha_0 + \epsilon_t$ (avec ϵ_t un bruit aléatoire). On estime les coefficients par $\hat{\alpha}_2, \hat{\alpha}_1, \hat{\alpha}_0$ (à partir des valeurs x_1, \dots, x_n). Ainsi, avec la donnée de x_1, \dots, x_n , on fera la prédiction $\hat{x}_{n+1} = \hat{\alpha}_2(n+1)^2 + \hat{\alpha}_1(n+1) + \hat{\alpha}_0$ de la valeur x_{n+1} .

- Lissages exponentiels (voir chapitre suivant).
- Modèles ARMA, qui consistent à enlever de la série les tendances et la saisonnalité (=périodicité). Ces modèles sont plus lourds numériquement, mais plus performants.

Les défis à relever (dans l'ordre) :

- Définir un modèle avec un nombre fini de paramètres.
- Estimer les paramètres du modèle.
- Vérifier la qualité de l'ajustement du modèle, comparer différents modèles (on pourra découper les données en un échantillon d'apprentissage et un échantillon de test).
- Effectuer des prédictions.

1.1. Tendances et composantes saisonnières

DÉFINITION 1.3. On dit que la série admet une tendance si on peut écrire $x_t = f(t) + \epsilon_t$ avec f une fonction fixée et (ϵ_t) des bruits aléatoires.

- Si $f(t) = \alpha t + \beta$, on dit que la tendance est linéaire. Plus généralement, si $x_t = \sum_{i=0}^p \alpha_i t^i$, on dit que la tendance est polynomiale.
- Si $f(t)$ est périodique, on dit que la tendance est périodique.
- Si $f(t) = s(t) + \alpha t + \beta$ avec s une fonction périodique on dit que la série a une tendance linéaire et une composante périodique (/saisonnière). (On remarque que ces définitions ne sont pas très cohérentes.)

1.2. Indices descriptifs d'une série temporelle

1.2.1. Indice de tendance centrale. Moyenne empirique : $\bar{x}_n = \frac{1}{n} \sum_{t=1}^n x_t$.

1.2.2. Indices de dispersion. Variance empirique : $\hat{\sigma}_n(0) = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}_n)^2$ (sa racine carrée est l'écart-type empirique).

1.2.3. Indices de dépendance. (qui renseignent sur la dépendance entre les données x_t)

Auto-covariance empirique d'ordre h (h dans \mathbb{N}^*) : $\hat{\sigma}_n(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (x_t - \bar{x}_n)(x_{t+h} - \bar{x}_n)$ ($h < n$ pour que la formule ait un sens).

Fonction d'auto-covariance empirique : $h \mapsto \hat{\sigma}_n(h)$.

Auto-corrélation empirique : $\hat{\rho}_n(h) = \frac{\hat{\sigma}_n(h)}{\hat{\sigma}_n(0)}$ (prend ses valeurs dans $[0; 1]$).

Fonction d'auto-corrélation empirique : $h \mapsto \hat{\rho}_n(h)$.

REMARQUE 1.4. Les quantités empiriques ci-dessus sont des estimateurs consistants de certaines grandeurs (c'est à dire qu'elles convergent vers certaines grandeurs quand $n \rightarrow +\infty$). Les convergences sont basées sur des applications de la loi des grands nombres. En particulier, pour h proche de n (disons $|n - h| < 50$), la quantité $\hat{\sigma}_n(h)$ n'a pas beaucoup d'intérêt.

La représentation graphique d'un nuage de points $(x_t, x_{t+1})_{1 \leq t \leq n-1}$ illustre la valeur de $\hat{\rho}_n(1)$ (voir figure 1.2.1). Plus le nuage est arrondi, plus $\hat{\rho}_n(1)$ est proche de 0. Plus le nuage est allongé, plus $\hat{\rho}_n(1)$ est proche de 1. Cette remarque est aussi valable pour les $\hat{\rho}_n(h)$ avec $h \geq 2$.

PROPOSITION 1.5. Supposons $x_t = a + bt + \epsilon_t$, avec $(\epsilon_t)_{t \geq 1}$ une suite de variable aléatoires i.i.d. (indépendantes et identiquement distribuées) et $a \neq 0$. Supposons que $\mathbb{E}(\epsilon_1^2) < \infty$. Alors, pour h fixé dans \mathbb{N}^* ,

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

DÉMONSTRATION. Notons $\bar{\epsilon}_n = \frac{1}{n}(\epsilon_1 + \epsilon_2 + \dots + \epsilon_n)$. Nous avons

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^n (at + b + \epsilon_t) = \frac{a(n+1)}{2} + b + \bar{\epsilon}_n.$$

Fixons $h \in \{1, 2, \dots, n-1\}$. Nous avons

$$\hat{\sigma}_n(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} \left(a \left(t - \frac{n+1}{2} \right) + \epsilon_t - \bar{\epsilon}_n \right) \left(a \left(t+h - \frac{n+1}{2} \right) + \epsilon_{t+h} - \bar{\epsilon}_n \right)$$

$$\begin{aligned}
&= \frac{1}{n-h} \sum_{t=1}^{n-h} \left[a^2 \left(t - \frac{n+1}{2} \right) \left(t+h - \frac{n+1}{2} \right) + (\epsilon_t - \bar{\epsilon}_n)(\epsilon_{t+h} - \bar{\epsilon}_n) \right. \\
&\quad \left. + (\epsilon_t - \bar{\epsilon}_n)a \left(t+h - \frac{n+1}{2} \right) + a \left(t - \frac{n+1}{2} \right) (\epsilon_{t+h} - \bar{\epsilon}_n) \right].
\end{aligned}$$

Nous avons

$$\frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n)(\epsilon_{t+h} - \bar{\epsilon}_n) = \frac{1}{n-h} \left(\sum_{t=1}^n \epsilon_t \epsilon_{t+h} + \bar{\epsilon}_n^2 - \bar{\epsilon}_n \epsilon_{t+h} - \epsilon_t \bar{\epsilon}_n \right),$$

et (par application de la loi de grands nombres),

$$\begin{aligned}
&\frac{1}{n-h} \sum_{t=1}^{n-h} \epsilon_t \epsilon_{t+h} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(\epsilon_1 \epsilon_{1+h}), \\
&\frac{n \bar{\epsilon}_n^2}{n-h} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(\epsilon_1^2), \\
&\frac{1}{n-h} \sum_{t=1}^{n-h} \bar{\epsilon}_n \epsilon_{t+h} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(\epsilon_1)^2, \\
&\frac{1}{n-h} \sum_{t=1}^{n-h} \bar{\epsilon}_n \epsilon_t \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(\epsilon_1)^2.
\end{aligned}$$

De plus, par Cauchy-Schwartz,

$$\begin{aligned}
\left| \frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n) a \left(t+h - \frac{n+1}{2} \right) \right| &\leq a \left(\frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n)^2 \right)^{1/2} \\
&\quad \times \left(\frac{1}{n-h} \sum_{t=1}^{n-h} \left(t+h - \frac{n+1}{2} \right)^2 \right)^{1/2} \\
&\leq a \left(\frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n)^2 \right)^{1/2} \\
&\quad \times \left(\frac{1}{n-h} \left(n+h - \frac{n+1}{2} \right)^2 \right)^{1/2}.
\end{aligned}$$

Nous avons (par application de la loi de grands nombres)

$$\frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n)^2 \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \text{Var}(\epsilon_1).$$

Donc, p.s.,

$$\frac{1}{n-h} \sum_{t=1}^{n-h} (\epsilon_t - \bar{\epsilon}_n) a \left(t+h - \frac{n+1}{2} \right) = O(n).$$

De même

$$\frac{1}{n-h} \sum_{t=1}^{n-h} a \left(t - \frac{n+1}{2} \right) (\epsilon_{t+h} - \bar{\epsilon}_n) = O(n).$$

Nous avons

$$\begin{aligned}
&\frac{1}{n-h} \sum_{t=1}^{n-h} a^2 \left(t - \frac{n+1}{2} \right) \left(t+h - \frac{n+1}{2} \right) \\
&= \frac{a^2}{n-h} \sum_{t=1}^{n-h} \left[t^2 + \frac{(n+1)^2}{4} + (h - (n+1))t + h \times \frac{n+1}{2} \right]
\end{aligned}$$

$$\begin{aligned}
 (\text{formule pour la somme des carrés}) &= \frac{a^2}{n-h} \left(\frac{(2(n-h)+1)(n-h+1)(n-h)}{6} \right) + a^2 \frac{(n+1)^2}{4} \\
 &\quad - \frac{a^2(h-n-1)}{n-h} \frac{n(n+1)}{2} + a^2 \frac{(n+1)}{2} = \frac{a^2}{3} n^2 + \frac{a^2}{4} n^2 + o(n^2).
 \end{aligned}$$

Donc, p.s.,

$$\hat{\sigma}_n(h) \underset{n \rightarrow +\infty}{\sim} \frac{7a^2}{12} n^2.$$

Donc

$$\hat{\rho}_n(h) \underset{n \rightarrow +\infty}{\xrightarrow{\text{p.s.}}} 1.$$

□

PROPOSITION 1.6. *Sous les mêmes hypothèse que dans la proposition 1.5, si $x_t = a \cos(2\pi t/T)$, avec $a \neq 0$ et $T \in \mathbb{N}^*$, alors*

$$\hat{\rho}_n(h) \underset{n \rightarrow +\infty}{\longrightarrow} \cos(2\pi h/T)$$

REMARQUE 1.7. On admet ensuite que si l'auto-corrélation d'une série est constante, c'est qu'elle a une tendance linéaire et que si cette auto-corrélation est périodique de période T , alors la série a une composante périodique de période T . Les convergences décrites dans les propositions ci-dessus sont plus lentes quand h est grand. On se limitera donc à $h \leq 20$ (tant que cela permet d'observer quelques périodes dans le cas avec saisonnalité).

1.3. Feuille d'exercices numéro 1 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf (c'est ce qui sera demandé au partiel).

1.3.1. Données de varicelle. Récupérer le fichier contenant le nombre de cas de varicelle relevés à New-York de janvier 1931 à juin 1972 (<http://math.unice.fr/~rubentha/enseignement/varicelle.dat>).

- (1) Créer un objet de type série temporelle contenant cette série. Représenter graphiquement la série. (Voir appendice pour les instructions utiles en R.)
- (2) Analyser qualitativement cette série, c'est-à-dire repérer d'éventuelles tendances et/ou saisonnalités (changer d'échelle si besoin).
- (3) Quel est le nombre de cas de varicelle mensuel moyen ?
- (4) Tracer les 25 premières auto-corrélations. Interpréter ces résultats.
- (5) Tracer sur un même graphique, les évolutions mensuelles du nombre de cas de varicelle pour chaque année (une courbe pour chaque année, ce qui nous donnera un certain nombre de courbes superposées).
- (6) Tracer sur un graphique l'évolution annuelle du nombre de cas de varicelle.
- (7) Ces deux dernières questions vous permettent-elles d'améliorer vos conclusions de la question 2 ?

1.3.2. Simulations de séries temporelles. On appelle bruit blanc gaussien une suite de variables aléatoires indépendantes et identiquement distribuées $(\epsilon_t)_{t \in \mathbb{N}}$ de loi normale centrée réduite.

- (1) Quelle est la fonction d'auto-corrélation d'un bruit blanc ?
- (2) Simuler un bruit blanc gaussien de taille 100, et représenter le graphiquement.
- (3) Tracer la fonction d'auto-corrélation.
- (4) Recommencer les deux questions précédentes et observer la variabilité des résultats. Jouer sur la longueur de la série.

- (5) Simuler maintenant la série temporelle $X(t) = 0.5t + 2\epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$ (taille 100).
- (6) Représenter graphiquement la série et interpréter-la qualitativement.
- (7) Faites de même pour $X(t) = 0.5t + \epsilon_t + 3 \cos(t\pi)$ avec $\epsilon_t \sim \mathcal{N}(0, 1)$.

1.3.3. Appendice : mise en œuvre sous R. Quelques fonctions R utiles à l'étude des séries temporelles :

- Avant de charger le fichier dans R, taper l'adresse dans le navigateur pour le visualiser. Charger un fichier de données numériques en sautant les k premières lignes (s'il y a du texte au début) :
`data=scan(file="donnee.dat",skip=k)`. Définir le répertoire courant :
`setwd("~/Documents/")` (ici on veut être dans ~/Documents), savoir quel est le répertoire courant : `getwd()`.
- 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`). Découper une série temporelle de `c(i,j)` à `c(k,1)` : `xd<-window(x,c(i,j),c(k,1))`. Transformer une série temporelle en vecteur de nombres : `v=as.numeric(cp)`.
- 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 ou auto-covariances (choisir corrélation OU covariance).

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.

Simulation de variables normales indépendantes : `x<-rnorm(n,mean=0,sd=1)` renvoie une suite de variables normales indépendantes de moyenne 0 et d'écart-type 1.

1.4. Corrigé de la feuille d'exercices numéro 1 (qui constitue un exemple de ce qui est attendu aux contrôles sur machine)

1.4.1. Données de varicelle.

- (1) Regarder le fichier sur internet pour voir combien de lignes il faut sauter (taper l'adresse `http://...` dans la barre d'adresse). On peut ensuite le sauvegarder sur l'ordinateur et le charger dans R avec l'instruction `scan` ou le charger directement dans R (avec l'instruction `scan` aussi). Instruction pour charger le fichier (en sautant la première ligne) :
`data=scan(file="../../../series-temp/varicelle.dat",skip=1),`
`serie<-ts(data,frequency=12)`

(bien mettre le chemin vers le fichier en entier). On choisit `frequency=12` pour avoir les années en abscisses (les données sont mensuelles). Pour dessiner le graphique : `plot.ts(serie)`. Voir le graphique dans la figure 1.4.1. Pour exporter un ou plusieurs graphiques en jpeg : `jpeg (filename='~/.../tp1_graphique%d.jpeg')` (bien mettre le chemin en entier), `plot.ts(serie), plot ..., plot ..., dev.off()`.

- (2) On ne voit pas de tendance mais on voit une saisonnalité. Pour mieux voir la saisonnalité, on peut zoomer sur une durée plus petite (les 50 premiers mois) (`serie_z<-ts(data,2,50),plot.ts(serie_z)`). Voir le graphique : figure 1.4.2.
- (3) On fait `mean(serie)` et on trouve 732,4076.
- (4) On rentre l'instruction `acf(serie,lag.max=25,type=c('correlation'))` et obtient le graphique de la figure 1.4.3. La fonction d'autocorrélation est périodique, ce qui indique une périodicité dans la série temporelle. La ligne pointillée bleue indique le niveau en-dessous duquel la corrélation n'est plus statistiquement significative.

- (5) On utilise le code :

```
png (filename='~/.../tp1_graphe-multiple.png')
cp<-window(serie,start=c(1,1),end=c(1,12))
v=as.numeric(cp)
plot(c(1:12),v,'l',xlim=c(0,13),ylim=c(0,3100))
for(i in 2:41)
{
  cp<-window(serie,start=c(i,1),end=c(i,12))
  v=as.numeric(cp)
  lines(v)
}
dev.off()
```

Remarquer la boucle, l'extraction des lignes avec la fonction `window`, la conversion en vecteur avec `as.numeric`. Voir le graphique dans la figure 1.4.4.

- (6) Code pour l'évolution annuelle du nombre de cas de varicelle :

```
v=c();
for (i in 1:41)
{
  cp<-window(serie,start=c(i,1),end=c(i,12))
  v=c(v,sum(cp))
}
plot(1:41,v)
```

Et on obtient le graphique : voir figure 1.4.5.

- (7) Le dernier graphique semble indiquer une tendance décroissante.

1.4.2. Simulation de série temporelle.

- (1) Si on prend $\epsilon_1, \epsilon_2, \epsilon_3 \dots$ des variables i.i.d. de loi $\mathcal{N}(0,1)$ alors, pour tout i , ϵ_i est de variance 1 et pour tout $i \neq j$, $\mathbb{E}(\epsilon_i \epsilon_j) - \mathbb{E}(\epsilon_i) \mathbb{E}(\epsilon_j) = 0$. Donc la fonction d'autocorrélation est la suivante

$$R(i,j) = \frac{\mathbb{E}(\epsilon_i \epsilon_j) - \mathbb{E}(\epsilon_i) \mathbb{E}(\epsilon_j)}{\mathbb{E}(\epsilon_i^2) - \mathbb{E}(\epsilon_i)^2} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon.} \end{cases}$$

- (2) `v=rnorm(100, mean = 0, sd = 1), plot(1:100,v)` (voir le graphique dans la figure 1.4.6)

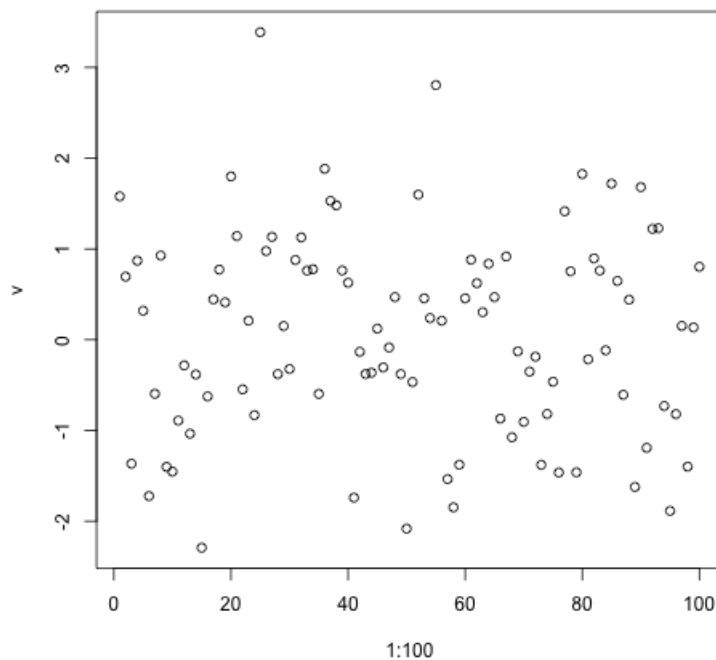


FIGURE 1.4.6. Bruit blanc gaussien.

- (3) `acf(v, lag.max=24, type=c("correlation"))` (voir le graphique dans la figure 1.4.7)
- (4) On recommence avec une série de longueur 1000, les corrélations sont plus petites, ce qui est attendu puisque les corrélations empiriques convergent vers la corrélation quand la taille de la série tend vers l'infini. En 0, on n'a pas convergence vers 0 puisque la corrélation vaut 1. Voir le graphique dans la figure 1.4.8.
- (5) `v=rnorm(100, mean = 0, sd = 1)`, `t=1:100`, `x=0.5*t+2*v` (Nous utilisons ici des manipulations vectorielles pour construire la suite (X_t) .)
- (6) `xs<-ts(x)` (Nous convertissons `x` en série temporelle par commodité.), `plot(xs)`, `acf(xs, lag.max=25, type=c("correlation"))`. Voir le graphique dans la figure 1.4.9. On repère une tendance croissante mais pas de saisonnalité (graphe de la série à gauche). Le graphe des corrélations (à droite) montre que les termes successifs de la suite sont très corrélés (et que donc il y a une tendance).
- (7) `t=1:100`, `v=rnorm(100, mean = 0, sd = 1)`, `x=0.5*t+v+3*cos(pi*t)`, `acf(xs, lag.max=24, type=c("correlation"))` (voir le graphique dans la figure 1.4.10) On observe une tendance et une saisonnalité sur le graphe de gauche. On remarque que les deux caractéristiques (tendance et saisonnalité) sont mélangées dans le graphe de droite.

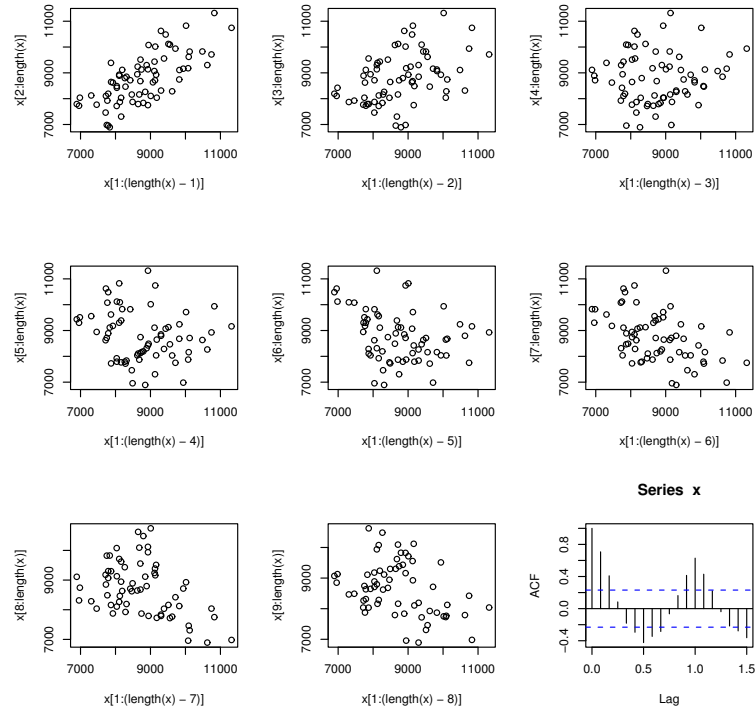


FIG. 6 – Nuages de points (x_t, x_{t+k}) pour $k = 1, \dots, 8$ et auto-corrélation pour la série temporelle du nombre de morts accidentelles aux Etats-Unis de 1973 à 1978

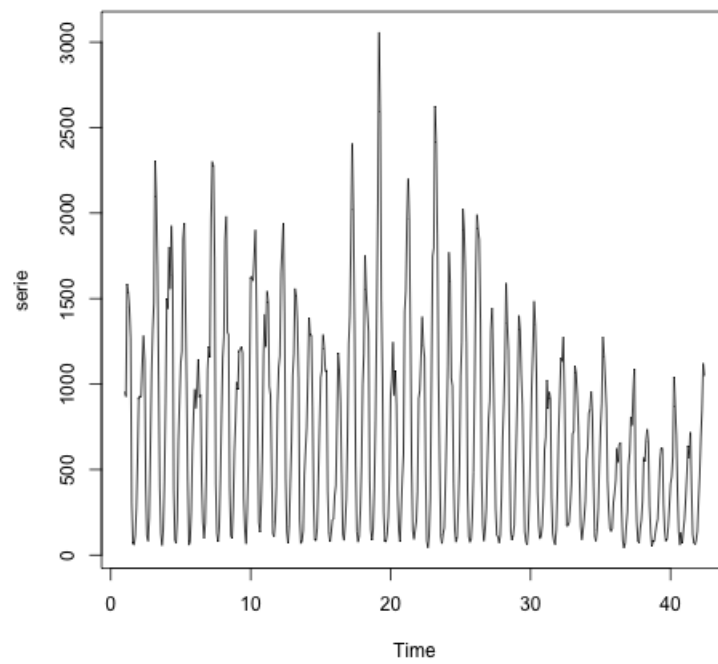


FIGURE 1.4.1. varicelle

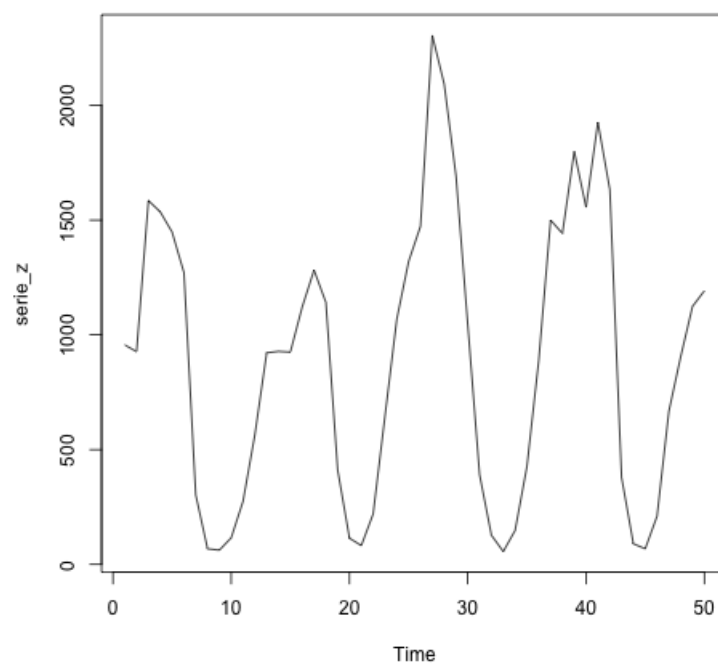


FIGURE 1.4.2. zoom

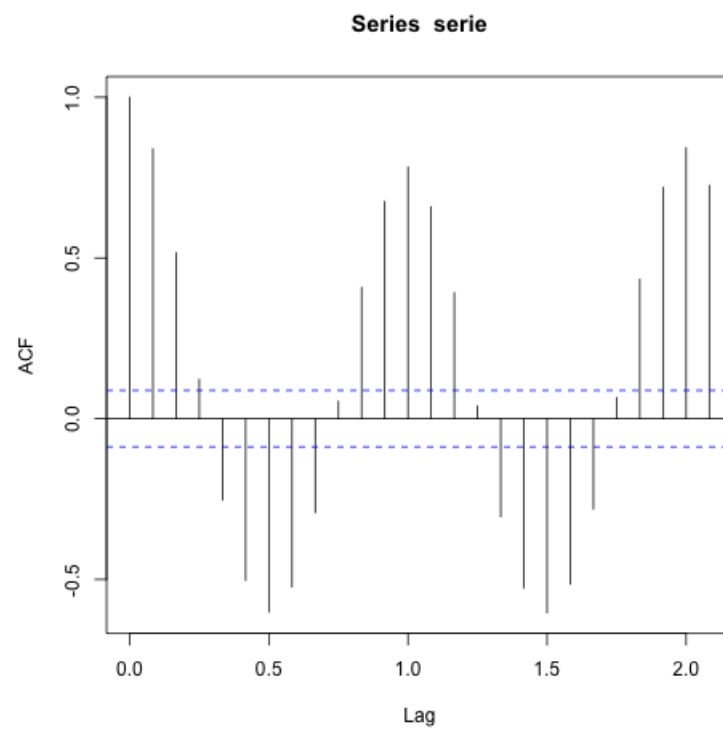


FIGURE 1.4.3. 25 premières autocorrélation

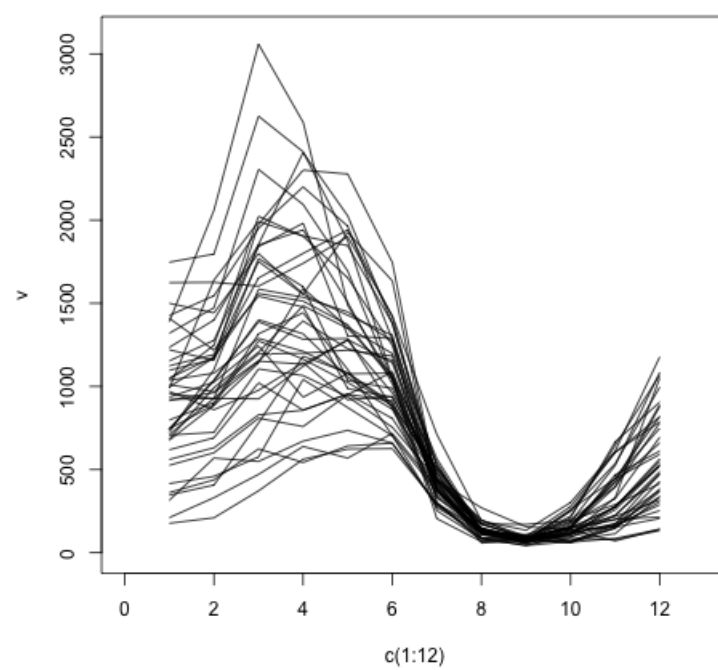


FIGURE 1.4.4. Graphes superposés

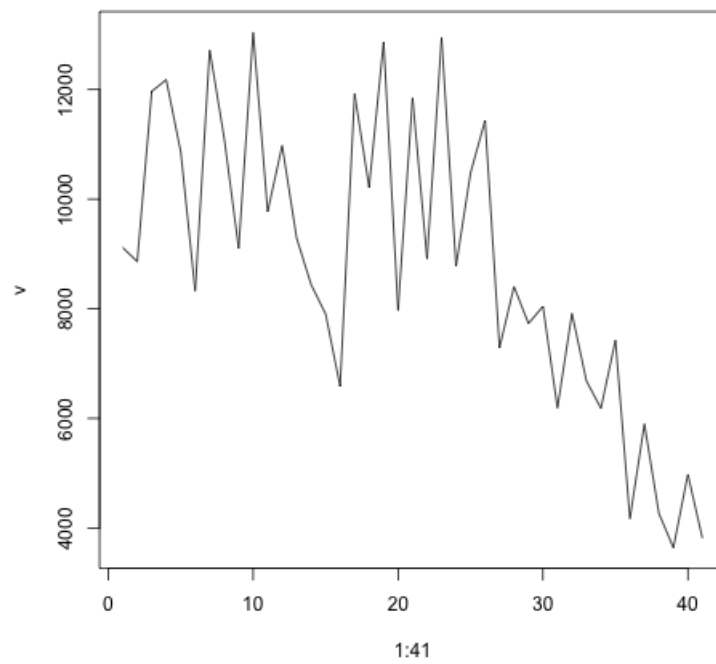


FIGURE 1.4.5. Par année

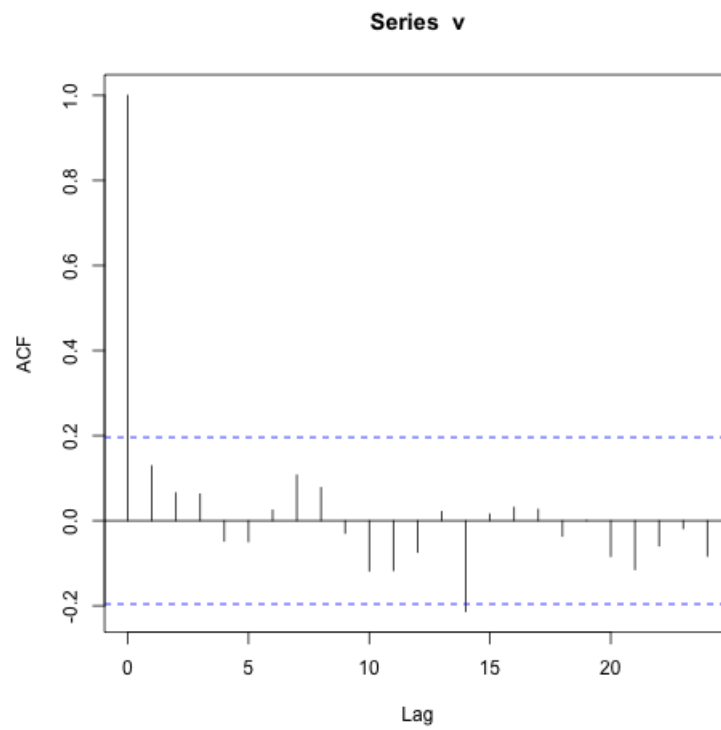


FIGURE 1.4.7. Autocorrélation du bruit blanc.

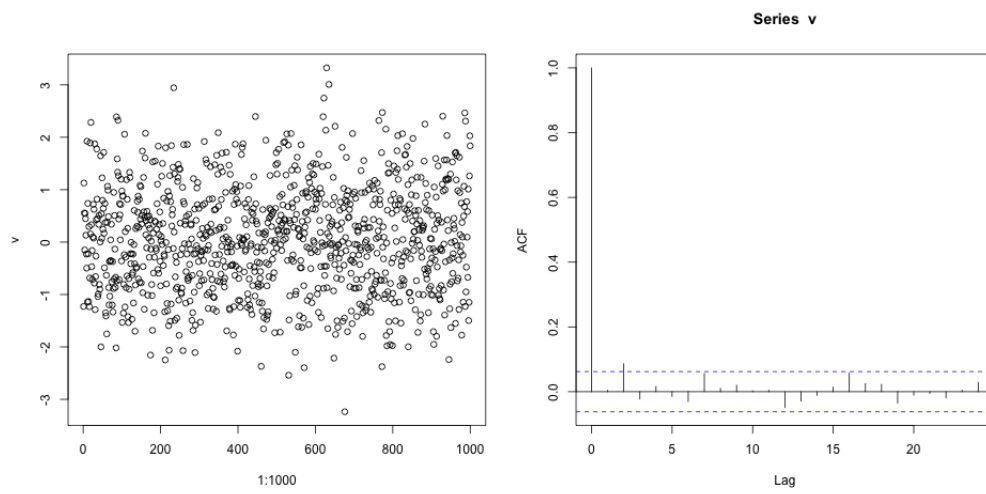


FIGURE 1.4.8. Longueur=1000.

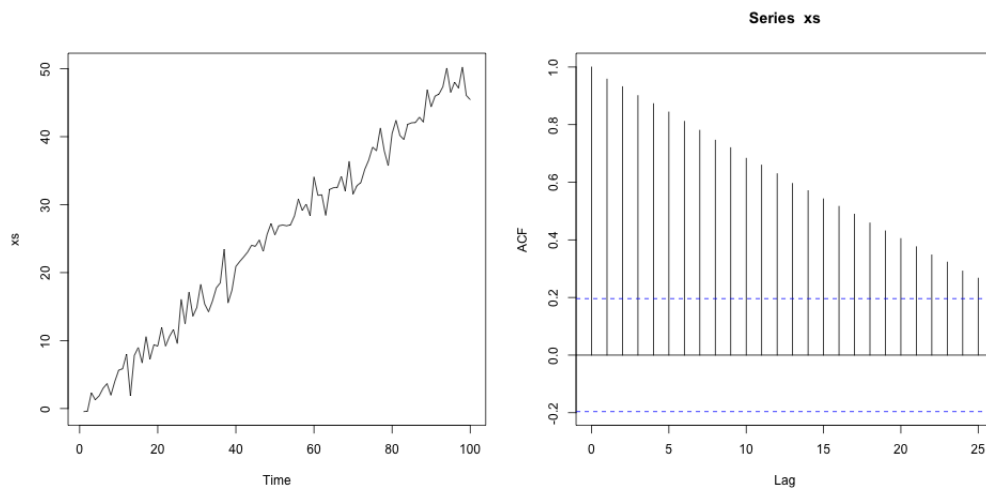


FIGURE 1.4.9. Suite à tendance croissante.

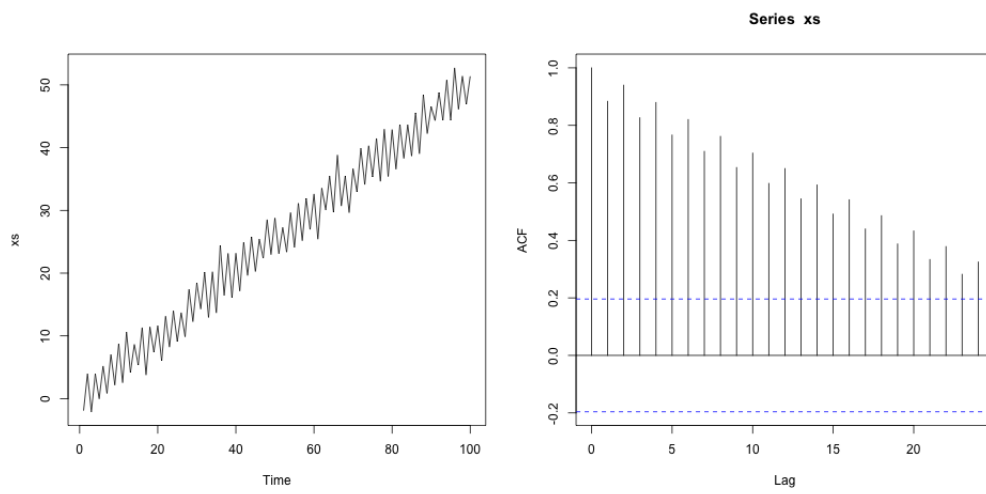


FIGURE 1.4.10. Longueur=1000.

Lissages exponentiels

2.1. Lissage exponentiel simple

On dispose de x_1, \dots, x_n et on veut estimer x_{n+h} ($h \in \mathbb{N}^*$). Pour $\alpha \in]0; 1[$, on définit la prévision par lissage exponentiel simple

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

(on fait en fait une moyenne pondérée des observations passées).

REMARQUE 2.1. Plus α est petit, plus on donne d'importance aux observations anciennes. On remarque que $\hat{x}_{n,h}$ ne dépend pas de h .

On peut calculer par récurrence à l'aide de la formule de mise à jour :

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

Ces formules permettent de calculer rapidement (en limitant la complexité).

LEMME 2.2. La prédiction $\hat{x}_{n,h}$ est (asymptotiquement) la solution de

$$\hat{x}_{n,h} = \arg \min_x \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x)^2.$$

C'est un problème de moindres carrés.

DÉMONSTRATION. Soit

$$f : x \mapsto \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x)^2.$$

On commence par chercher les points critiques. Nous avons :

$$f'(x) = -2 \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x).$$

Donc f' s'annule en

$$x_0 = \frac{\sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}}{\sum_{j=0}^{n-1} (1-\alpha)^j} = \frac{\alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}}{1 - (1-\alpha)^n} \underset{n \rightarrow +\infty}{\sim} \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}.$$

Une étude rapide du tableau de variation de f nous dit que x_0 est le minimum absolu. C'est à cause de l'équivalent que l'on dit que $\hat{x}_{n,h}$ est la solution « asymptotique » du problème des moindres carrés. \square

REMARQUE 2.3. Pour choisir α , on peut se servir de la remarque 2.1 ou calculer, pour tout α , les estimateurs calculés avec le paramètre $\alpha : \hat{x}_{t,h}(\alpha)$. On regarde ensuite l'erreur quadratique

$$E_2(\alpha) = \sum_{t=1}^{n-h} (x_{t+h} - \hat{x}_{t,h}(\alpha))^2.$$

Si cette erreur est petite, c'est que le paramètre α produit des prédictions performantes, au vu des données x_1, \dots, x_n . On peut choisir entre plusieurs paramètres $\alpha_1, \dots, \alpha_p$ en prenant

$$\alpha = \arg \min_{\alpha_1, \dots, \alpha_p} E_2(\alpha_i).$$

REMARQUE 2.4. Notons x_1, x_2, \dots, x_n la température à midi sur la promenade des Anglais (mesurée sous abri) pour n jour. La série x_1, \dots, x_n est supposée prendre ses valeurs dans les entiers. Pour chaque t dans $\{1, 2, \dots, n\}$, on peut comparer x_{t+1} avec une prédiction basée sur x_1, \dots, x_t . Nous nous intéressons à deux prédictions : $\hat{x}_{t,1}^{(1)}$ est la prédiction de Météo France, $\hat{x}_{t,1}^{(2)} = x_t$. Pour $i = 1, 2$, nous considérons deux erreurs :

$$E_1(i) = \sum_{t=1}^{n-1} \mathbb{1}_{\hat{x}_{t,1} = x_{t+1}},$$

$$E_2(i) = \sum_{t=1}^{n-1} (\hat{x}_{t,1} - x_{t+1})^2.$$

En général, $E_1(2) < E_1(1)$, donc si on s'intéresse au critère E_1 , pas besoin de prévision météo. Toujours en général, $E_2(1) < E_2(2)$.

2.2. Lissage exponentiel double

On cherche à ajuster, à l'instant t , une droite d'équation : $y_t = a_1 + a_2(t - n)$. La prévision par lissage exponentiel double est la suivante (où $\alpha \in]0; 1[$)

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

où (\hat{a}_1, \hat{a}_2) solution de

$$(\hat{a}_1, \hat{a}_2) = \arg \min_{(a_1, a_2) \in \mathbb{R}} \sum_{j=0}^{n-1} (1 - \alpha)^j (x_{n-j} - (a_1 + a_2 j))^2.$$

LEMME 2.5. *Les solutions du problème de minimisation ci-dessus sont (asymptotiquement)*

$$\begin{cases} \hat{a}_1 &= -L_2(n) + 2L_1(n) \\ \hat{a}_2 &= \frac{\alpha}{1-\alpha} (-L_2(n) + L_1(n)), \end{cases}$$

où

$$\begin{cases} L_1(n) &= \alpha \sum_{j=0}^{n-1} (1 - \alpha)^j x_{n-j} \\ L_2(n) &= \alpha \sum_{j=0}^{n-1} (1 - \alpha)^j L_1(n - j). \end{cases}$$

(On remarque qu'on a ici deux lissages exponentiels successifs.)

REMARQUE 2.6. C'est parce que L_2 est un lissage exponentiel de L_1 qu'on parle de lissage exponentiel double.

DÉMONSTRATION. Notons

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

On commence par chercher les points critiques de C . Nous avons

$$\begin{aligned} \frac{\partial C}{\partial a_1}(a_1, a_2) &= -2 \sum_{j=0}^{n-1} (1 - \alpha)^j (x_{n-j} - (a_1 + a_2 j)), \\ \frac{\partial C}{\partial a_2}(a_1, a_2) &= -2 \sum_{j=0}^{n-1} (1 - \alpha)^j j (x_{n-j} - (a_1 + a_2 j)). \end{aligned}$$

Nous cherchons donc (a_1, a_2) solutions du système

$$(2.1) \quad \begin{cases} -\sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} & + & \sum_{j=0}^{n-1} (1-\alpha)^j a_1 & + & \sum_{j=0}^{n-1} (1-\alpha)^j j a_2 & = & 0 \\ -\sum_{j=0}^{n-1} (1-\alpha)^j j x_{n-j} & + & \sum_{j=0}^{n-1} (1-\alpha)^j j a_1 & + & \sum_{j=0}^{n-1} (1-\alpha)^j j^2 a_2 & = & 0. \end{cases}$$

Rappelons les formules

$$\begin{aligned} \sum_{j=0}^{+\infty} (1-\alpha)^j &= \frac{1}{\alpha}, \\ \sum_{j=0}^{+\infty} j(1-\alpha)^j &= \frac{1-\alpha}{\alpha^2}, \\ \sum_{j=0}^{+\infty} j^2(1-\alpha)^j &= \frac{(1-\alpha)(2-\alpha)}{\alpha^3} \end{aligned}$$

(que l'on peut retrouver facilement en manipulant des séries entières). On obtient donc, en remplaçant certaines sommes partielles de séries par leurs limites,

$$(2.2) \quad \begin{cases} -\alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} & + & a_1 & + & \frac{1-\alpha}{\alpha} a_2 & = & 0 \\ -\alpha^2 \sum_{j=0}^{n-1} (1-\alpha)^j j x_{n-j} & + & (1-\alpha)a_1 & + & \frac{(1-\alpha)(2-\alpha)}{\alpha} a_2 & = & 0. \end{cases}$$

Attention, les systèmes (2.1) et (2.2) ne sont pas équivalents. Plus n est grand, plus ils se ressemblent. Nous allons continuer le calcul à partir du système (2.2) et c'est pour cela que l'on parle de solution « asymptotique » dans l'énoncé du lemme. Posons (pour tout n)

$$\begin{cases} L_1(n) = & \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} \\ L_2(n) = & \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j). \end{cases}$$

Nous remarquons que

$$\begin{aligned} L_2(n) &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j) \\ &= \alpha^2 \sum_{j=0}^{n-1} (1-\alpha)^j \sum_{i=0}^{n-j-1} (1-\alpha)^i x_{n-j-i} \\ (k=i+j) &= \alpha^2 \sum_{k=0}^{n-1} x_{n-k} (1-\alpha)^k (k+1). \end{aligned}$$

Le système devient alors

$$\begin{cases} -L_1(n) + a_1 + \frac{(1-\alpha)}{\alpha} a_2 & = & 0 \\ -L_2(n) + \alpha L_1(n) + (1-\alpha)a_1 + \frac{(1-\alpha)(2-\alpha)}{\alpha} a_2 & = & 0. \end{cases}$$

D'où

$$\begin{aligned} -L_1(n)(2-\alpha) + a_1(2-\alpha) + L_2(n) - \alpha L_1(n) &- (1-\alpha)a_1 = & 0 \\ -L_2(n) + 2L_1(n) &= & a_1, \end{aligned}$$

et

$$\begin{aligned} a_2 &= \frac{\alpha}{1-\alpha} (L_1(n) - a_1) \\ &= \frac{\alpha}{1-\alpha} (L_2(n) - L_1(n)). \end{aligned}$$

Montrons que la fonction C est convexe. Pour (a_1, a_2) et (b_1, b_2) dans \mathbb{R}^2 et $\lambda \in [0, 1]$, nous avons

$$C(\lambda(a_1, a_2) + (1-\lambda)(b_1, b_2)) = \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - (\lambda a_1 + (1-\lambda)b_1 + (\lambda a_2 + (1-\lambda)b_2)j))^2$$

$$\begin{aligned}
&= \sum_{j=0}^{n-1} (1-\alpha)^j (\lambda(x_{n-j} - (a_1 + ja_2)) + (1-\lambda)(x_{n-j} - (b_1 + jb_2)))^2 \\
(\text{convexité de la fonction carré}) &\leq \sum_{j=0}^{n-1} (1-\alpha)^j [\lambda(x_{n-j} - (a_1 + ja_2))^2 \\
&\quad + (1-\lambda)(x_{n-j} - (b_1 + jb_2))^2] \\
&= \lambda C(a_1, a_2) + (1-\lambda)C(b_1, b_2).
\end{aligned}$$

Ce qui montre que la fonction C est convexe. L'unique point critique est donc le minimum absolu. \square

Les formules de mise à jour sont

$$(2.3) \quad \begin{cases} \hat{a}_1(n) = \hat{a}_1(n-1) + \hat{a}_2(n-1) + (2\alpha - \alpha^2)(x_n - \hat{x}_{n-1,1}) \\ \hat{a}_2(n) = \hat{a}_2(n-1) + \alpha^2(x_n - \hat{x}_{n-1,1}), \end{cases}$$

avec l'initialisation

$$\begin{cases} \hat{a}_1(0) = x_1 \\ \hat{a}_2(0) = x_2 - x_1. \end{cases}$$

Ces formules permettent de comprendre comment calculer récursivement les coefficients (ce qui limite la complexité du calcul).

2.3. Méthode de Holt-Winters

2.3.1. Méthode non saisonnière. On cherche, au voisinage de l'instant n , à ajuster une droite d'équation

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

La prévision en $n + h$ sera

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

On choisit deux constantes de lissage β et α dans $]0; 1[$. Les \hat{a}_1, \hat{a}_2 sont calculés récursivement par

$$(2.4) \quad \begin{cases} \hat{a}_1(n) = \alpha x_n + (1-\alpha)(\hat{a}_1(n-1) + \hat{a}_2(n-1)) \\ \hat{a}_2(n) = \beta(\hat{a}_1(n) - \hat{a}_1(n-1)) + (1-\beta)\hat{a}_2(n-1). \end{cases}$$

REMARQUE 2.7. Cette méthode est plus souple que la précédente. Le α joue un rôle dans l'estimée de l'ordonnée en n et le β joue un rôle dans l'estimée de la pente.

Plus α et β sont petits, plus on tient compte du passé lointain.

LEMME 2.8. *Pour obtenir un lissage exponentiel double de paramètre $1 - \alpha'$ (qui n'a pas d'instruction dédiée dans R), il suffit de faire un lissage de Holt-Winters sans composante saisonnière avec*

$$(2.5) \quad \begin{cases} \alpha &= 1 - (\alpha')^2 \\ \beta &= \frac{1 - \alpha'}{1 + \alpha'}. \end{cases}$$

On retrouve le lissage exponentiel simple de paramètre α si $\beta = 0$ (et $\alpha \neq 0$).

DÉMONSTRATION. Pour un lissage exponentiel double de paramètre $1 - \alpha'$ dans $]0; 1[$, on peut écrire les formules de mise à jour (2.3) sous la forme (on remarque au préalable que $(1 - \alpha')(2 - (1 - \alpha')) = 1 - (\alpha')^2$,

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

$$\begin{aligned}
& +(\hat{a}_1(n-1) + \hat{a}_2(n-1)) \frac{(1-\alpha')}{1+\alpha'} (-(\alpha')^2 - (1-\alpha')(1+\alpha')) \\
= & \hat{a}_2(n-1) + \left(\frac{1-\alpha'}{1+\alpha'} \right) \hat{a}_1(n) - (\hat{a}_1(n-1) + \hat{a}_2(n-1)) \left(\frac{1-\alpha'}{1+\alpha'} \right) \\
= & (1-\beta)\hat{a}_2(n-1) + \beta(\hat{a}_1(n) - \hat{a}_1(n-1))
\end{aligned}$$

On retrouve bien la formule de mise à jour (2.4) avec $\alpha = 1 - (\alpha')^2$, $\beta = \frac{1-\alpha'}{1+\alpha'}$. \square

2.3.2. Méthode saisonnière additive. On cherche, au voisinage de n , à ajuster une courbe d'équation

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

(avec s_t périodique, de période T). Il faut choisir α, β, δ dans $]0; 1[$ (et aussi T). Les formules de récursion sont

$$\begin{cases} \hat{a}_1(n) = & \alpha(x_n - \hat{s}_{n-T}) + (1-\alpha)(\hat{a}_1(n-1) + \hat{a}_2(n-1)) \\ \hat{a}_2(n) = & \beta(\hat{a}_1(n) - \hat{a}_1(n-1)) + (1-\beta)\hat{a}_2(n-1) \\ \hat{s}_n = & \delta(x_n - \hat{a}_1(n)) + (1-\delta)\hat{s}_{n-T}. \end{cases}$$

La prévision prend la forme

$$\begin{cases} \hat{x}_{n,h} = & \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-T}, 1 \leq h \leq T \\ \hat{x}_{n,h} = & \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-2T}, T+1 \leq h \leq 2T \\ \dots & \dots \end{cases}$$

Les valeurs initiales suivantes permettent de commencer le calcul à $T+2$

$$\begin{cases} \hat{a}_1(T+1) = & x_{T+1} \\ \hat{a}_2(T+1) = & \frac{x_{T+1} - x_1}{T} \\ \hat{s}_j = & x_j - (x_1 + (T-1)\hat{a}_2(T+1)), 1 \leq j \leq T. \end{cases}$$

2.3.3. Méthode saisonnière multiplicative. On cherche, au voisinage de n , à ajuster une courbe d'équation

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

(avec s_t périodique, de période T). Il faut choisir α, β, δ dans $]0; 1[$. Les formules de récursion sont

$$\begin{cases} \hat{a}_1(n) = & (1-\alpha) \frac{x_n}{\hat{s}_{n-T}} + \alpha(\hat{a}_1(n-1) + \hat{a}_2(n-1)) \\ \hat{a}_2(n) = & (1-\beta)(\hat{a}_1(n) - \hat{a}_1(n-1)) + \beta\hat{a}_2(n-1) \\ \hat{s}_n = & (1-\delta) \frac{x_n}{\hat{a}_1(n)} + \delta\hat{s}_{n-T}. \end{cases}$$

La prévision prend la forme

$$\begin{cases} \hat{x}_{n,h} = & (\hat{a}_1(n) + h\hat{a}_2(n)) \times \hat{s}_{n+h-T}, 1 \leq h \leq T \\ \hat{x}_{n,h} = & (\hat{a}_1(n) + h\hat{a}_2(n)) \times \hat{s}_{n+h-2T}, T+1 \leq h \leq 2T \\ \dots & \dots \end{cases}$$

Initialisation : c'est très dur mais on peut se contenter de mettre les premiers \hat{a}_2 à 0 et les premiers \hat{a}_1 à 1 (et le système se corrige par la suite).

REMARQUE 2.9. Les deux méthodes ci-dessus (saisonnière additive et saisonnière multiplicative) ne peuvent être mises en place que si nous disposons d'un certain nombre de période pour estimer la partie périodique. Si le nombre de données est insuffisant, R renvoie un message d'erreur.

On choisit la méthode saisonnière additive ou saisonnière multiplicative en fonction de l'aspect graphique des données. En gros : amplitudes constantes \rightarrow méthode saisonnière additive, amplitudes non constantes \rightarrow méthode saisonnière multiplicative.

2.4. Feuille d'exercices numéro 2 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf (c'est ce qui sera demandé au partiel).

2.4.1. Lissage et prévision de données simulées. Simuler les trois séries temporelles suivantes de taille 100 :

- $X_1(t) = \epsilon_t$,
- $X_2(t) = 0.5t + 2\epsilon_t$,
- $X_3(t) = 0.5t + \epsilon_t + 3\cos(t\pi/6)$,

où les ϵ_t sont i.i.d. de loi $\mathcal{N}(0, 1)$.

- (1) Pour chaque série, effectuer la prévision $\hat{x}_{t,1}$ ($1 \leq t \leq 99$) par lissage exponentiel simple et double. Tester différentes valeurs du paramètre de lissage α (4 à 5 valeurs), et représenter graphiquement la série ainsi que la prévision. Commenter chaque résultat, et essayer de déterminer graphiquement le lissage le plus adaptée pour chaque série.
- (2) Calculer pour chaque prévision effectuée la somme des carrés des erreurs, et sélectionner le meilleur paramètre à l'aide de cette quantité.
- (3) Tester maintenant le lissage exponentiel de Holt-Winters avec composante saisonnière additive puis multiplicative (seulement pour les séries avec une composante périodique).
- (4) Les prédictions obtenues sont-elles meilleures ?

2.4.2. Lissage et prévision de la concentration en co2. Le fichier de données `co2` contenu dans `R1` contient les concentrations en CO2 à proximité du volcan Mauna Loa (Hawaï) de 1959 à 1997.

Après avoir représenté graphiquement ces données, quel modèle de lissage exponentiel vous semble le mieux approprié ?

Afin de valider ce modèle, tester la prédiction des données de 1990 à 1997 en utilisant celles de 1959 à 1989. Si cela vous semble graphiquement correct, utilisez cette méthode pour prédire les concentrations en CO2 de 1997 à 2007. Sinon, tester d'autres méthodes de lissage exponentiel.

2.4.3. Lissage et prévision du CAC40. Récupérer le fichier contenant les valeurs de clôture journalière du CAC40 de 1991 à 1998 (données `R EuStockMarkets`). Essayer de prédire par lissage exponentiel les valeurs de clôture de 1998 en utilisant les valeurs de clôture de 1991 à 1997.

Appendice : mise en œuvre sous R. La commande `par(mfrow=c(n,p))` permet d'afficher `n*p` graphique sur la même page.

Obtenir un vecteur aléatoire de longueur `n` de composantes des variables i.i.d. $\sim \mathcal{N}(0, 1)$: `e=rnorm(n)`. Multiplier/additionner deux vecteurs composante par composante : `x*y`, `x+y`. Prendre la fonction `f` d'un vecteur `x` composante par composante : `f(x)`. Créer un vecteur par concaténation : `v=c()`; `v=c(v,1,2,3)`. Trouver le type d'une variable : `typeof(x)`.

Découper une série temporelle de `c(i,j)` à `c(k,1)` : `xd<-window(x,c(i,j),c(k,1))`.

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=FALSE,gamma=FALSE)`,
- un lissage exponentiel double paramètre $1 - \alpha'$:
`xlisse <- HoltWinters(x,alpha=α, beta=β, gamma=FALSE)` avec $\alpha = 1 - (\alpha')^2$, $\beta = \frac{1-\alpha'}{1+\alpha'}$,
- un lissage de Holt-Winters sans composante saisonnière :
`xlisse <- HoltWinters(x,alpha=α,beta=β,gamma=FALSE)`,
- 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')`.

À 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.

Les prévisions à l'horizon `h` sont réalisées à l'aide de la fonction `predict` :
`p<-predict(xlisse,n.ahead=h)` (attention le `n` fait partie de la commande et le `h` est une variable

1. Il est dans l'espace de travail dès que vous lancez R.

qui peut prendre les valeurs 1, 2, ...). 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`. Tracer la prédiction et la série sur le même graphique : `xlisse=HoltWinters(x,...); p=predict(xlisse,n.ahead=50); plot(xlisse,p)`.

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.

Estimation et élimination de la tendance et de la saisonnalité

On observe un processus (X_t) et on cherche à le modéliser. La démarche est la suivante :

- Représenter graphiquement (x_1, x_2, \dots, x_n) pour repérer tendance et saisonnalité.
- Choisir un modèle pour les résidus (ce qui reste et qui est aléatoire) et valider le modèle (ou invalider le modèle).
- Prédire les réalisations futures à l'aide du modèle.

3.1. Bruit blanc

DÉFINITION 3.1. Un processus de bruit blanc est une suite de variables aléatoires (X_t) , d'espérance et de variance constante (c'est à dire $\mathbb{E}(X_t)$ et $\mathbb{V}(X_t)$ ne dépendent pas de t) et tel que $\text{Cov}(X_t, X_{t+h}) = 0$, si $h \neq 0$. Si l'espérance est nulle, le bruit blanc est dit centré. Si les variables sont gaussiennes, le bruit blanc est dit gaussien.

3.2. Processus stationnaire

DÉFINITION 3.2. Un processus aléatoire $(X_t)_{t \geq 0}$ est stationnaire s'il est d'espérance constante ($\mathbb{E}(X_t)$ ne dépend pas de t) et si les covariances sont stables par translation dans le temps ($\text{Cov}(X_t, X_{t+h})$ ne dépend pas de t , quel que soit $h \geq 0$). Attention, ce n'est pas la définition usuelle des ouvrages de probabilités.

Pour un processus $(X_t)_{t \geq 0}$ stationnaire, on définit la fonction d'auto-covariance par

$$h \in \mathbb{Z} \mapsto \sigma(h) = \text{Cov}(X_t, X_{t+h}) \text{ (ne dépend pas de } t),$$

et la fonction d'auto-corrélation par

$$h \in \mathbb{N}^* \mapsto \rho(h) = \frac{\sigma(h)}{\sigma(0)}.$$

On remarque que la fonction σ est paire.

EXEMPLE 3.3. Le processus de la figure 3.2.1 a l'air stationnaire. Le processus de la figure 3.2.2 n'a pas l'air stationnaire car l'espérance des points paraît non constante. Le processus de la figure 3.2.3 n'a pas l'air stationnaire car les X_{2k-1} , X_{2k} ont l'air d'être corrélés positivement, ce qui n'est pas le cas des X_{2k} , X_{2k+1} . Le processus de la figure 3.2.4 n'a pas l'air stationnaire car la variance des points ne paraît pas constante. Le processus de la figure 3.2.5 a l'air stationnaire (mais c'est encore plus difficile à dire avec ce type de graphique). (Plus de commentaires en cours.)

3.3. Estimation paramétrique de la tendance

On suppose que $X_t = m_t + \epsilon_t$, avec m_t déterministe et ϵ_t aléatoire. Dans le cas $m_t = at + b$, on peut estimer (a, b) par

$$(\hat{a}, \hat{b}) = \arg \min_{a, b} \sum_{t=1}^n (x_t - a - bt)^2$$

(méthode des moindres carrés). Comme d'habitude, on suppose que l'on dispose des observations x_1, \dots, x_n .

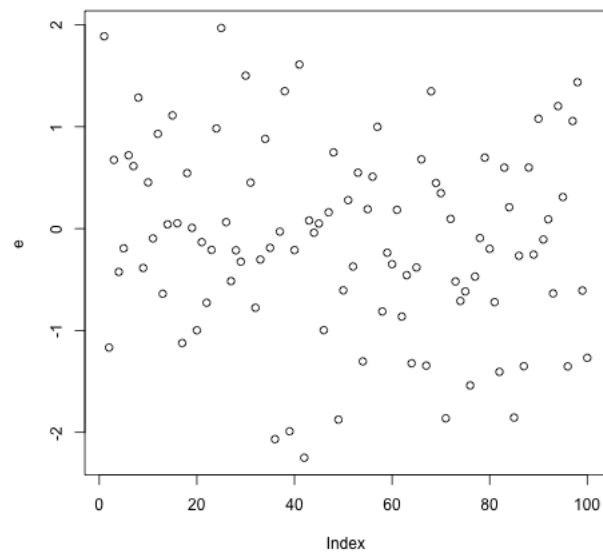


FIGURE 3.2.1.

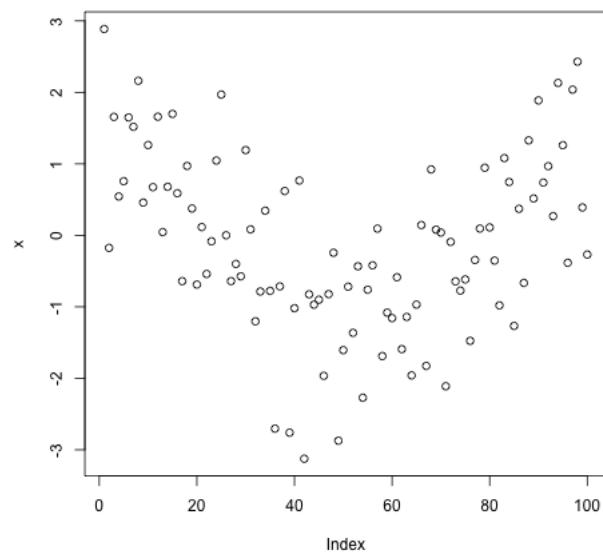


FIGURE 3.2.2.

LEMME 3.4. *Les solutions au problème de minimisation ci-dessus sont*

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

$$(\bar{x}_n = (x_1 + \cdots + x_n)/n).$$

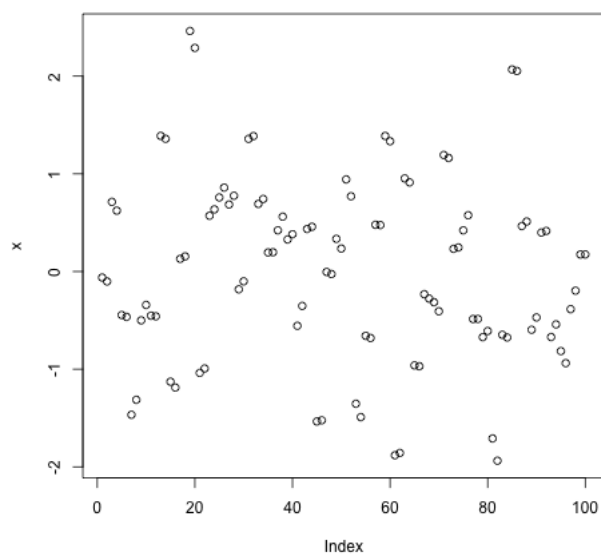


FIGURE 3.2.3.

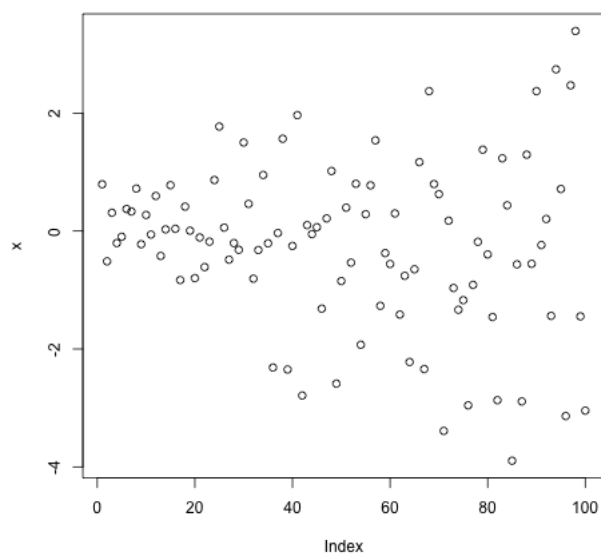


FIGURE 3.2.4.

(La démonstration de ce lemme fait partie d'une feuille d'exercices.) On pourrait aussi chercher une tendance polynomiale de plus haut degré (par exemple $m_t = a + bt + ct^2$), les calculs sont alors plus compliqués.

FAIT 3.5. (*À savoir par cœur.*) Pour tout n dans \mathbb{N}^* ,

$$\begin{cases} 1 + 2 + \cdots + n &= \frac{n(n+1)}{2} \\ 1^2 + 2^2 + \cdots + n^2 &= \frac{n(n+1)(2n+1)}{6} \end{cases}.$$

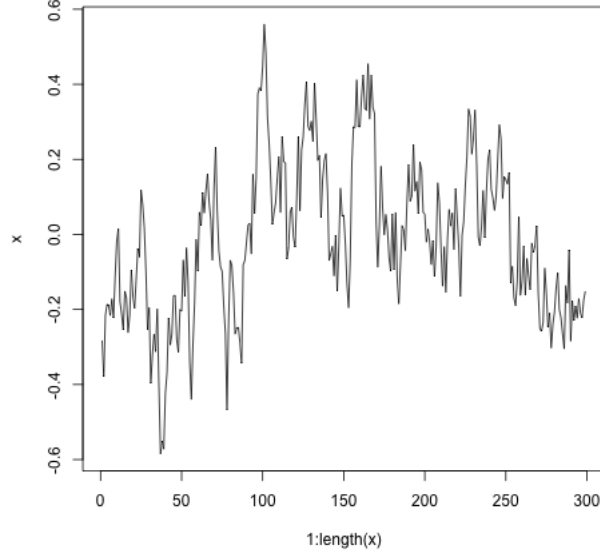


FIGURE 3.2.5.

3.4. Estimation non paramétrique : moyenne mobile

3.4.1. Tendance linéaire. On suppose que $X_t = m_t + \epsilon_t$, avec m_t déterministe et ϵ_t aléatoire. Si m_t est une fonction affine (donc à croissance linéaire) dans un intervalle autour de t , on peut estimer la moyenne par

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

C'est une estimation par moyenne mobile. Par convention,

$$\begin{cases} x_t = x_1 & \text{si } t \leq 0, \\ x_t = x_n & \text{si } t \geq n. \end{cases}$$

3.4.2. Tendance linéaire et composante périodique. On suppose que $X_t = m_t + s_t + \epsilon_t$, avec m_t et s_t déterministes et ϵ_t aléatoire. La fonction m_t est supposée affine et la fonction s_t est supposée T -périodique. On peut toujours supposer que

$$\sum_{t=1}^T s_t = 0$$

(quitte à translater s et m d'une constante). On peut estimer m_t par la formule

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

(exercice : expliquer pourquoi il est avantageux de choisir $2q+1$ le plus proche possible de T). Puis on estime la composante périodique (on dit aussi : saisonnière) en faisant la moyenne sur toutes les périodes :

$$\text{pour } t \in \{1, 2, \dots, T\}, \hat{s}_t = \frac{(x_t - \hat{m}_t) + (x_{t+T} - \hat{m}_{t+T}) + \dots + (x_{t+kT} - \hat{m}_{t+kT})}{k+1},$$

avec $k = \sup\{i : t + iT \leq n\}$,

puis, pour $t \in \{T+1, T+2, \dots, n\}$, $\hat{s}_t = \hat{s}_{t[T]}$,

où $t[T]$ est le reste de la division euclidienne de t par T . Cette dernière formule fournit un estimateur raisonnable puisque s_t est T -périodique.

3.5. Élimination de la tendance et de la saisonnalité par la méthode des différences

On peut éliminer la partie déterministe sans l'estimer. Fixons T dans \mathbb{N}^* et notons

$$\Delta_T : (X_t)_{t \in \mathbb{Z}} \mapsto (X_t - X_{t-T})_{t \in \mathbb{Z}}$$

(on notera parfois $\Delta = \Delta_T$).

PROPOSITION 3.6. *Si (X_t) est une série temporelle admettant une tendance polynômiale de degré $k \geq 1$ alors (ΔX_t) admet une tendance polynômiale de degré $\leq k - 1$. Si (X_t) est une série temporelle admettant une tendance polynômiale de degré 0 (c'est à dire une constante) alors (ΔX_t) a une tendance égale à 0.*

DÉMONSTRATION. On écrit

$$X_t = P(t) + \epsilon_t$$

avec des ϵ_t aléatoires et $P(t) = a_0 + a_1 t + \dots + a_k t^k$. Nous avons

$$(\Delta_T X)_t = P(t) - P(t - T) + \epsilon_t - \epsilon_{t-T}.$$

La série $(\epsilon_t - \epsilon_{t-T})_t$ est la partie aléatoire de $\Delta_T X$, la partie déterministe est le polynôme suivant en Y :

$$P(Y - T) - P(Y) = a_k(Y - T)^k - a_k Y^k + \sum_{i=0}^{k-1} a_i(Y - T)^i - a_i Y^i.$$

On voit que le degré de $\sum_{i=0}^{k-1} a_i(Y - T)^i - a_i Y^i$ est $\leq k - 1$. Regardons

$$\begin{aligned} a_k(Y - T)^k - a_k Y^k &= \sum_{i=0}^k [C_k^i Y^{k-i} T^i] - a_k Y^k \\ &= \sum_{i=1}^k C_k^i Y^{k-i} T^i, \end{aligned}$$

qui est bien de degré $\leq k - 1$ □

REMARQUE 3.7. Si $\mathbb{E}(\epsilon_t) = 0$ alors $\mathbb{E}(\epsilon_t - \epsilon_{t-T})$ donc on garde un bruit centré si on applique Δ (et en fait, si les (ϵ_t) forment un bruit blanc alors les $(\epsilon_t - \epsilon_{t-T})$ aussi)

Si on applique $k + 1$ fois Δ une série dont la tendance est un polynôme de degré k , on tombe sur un processus dont la tendance est égale à 0 et dont les bruits sont centrés. Donc on ne tombe pas sur un processus nul mais sur un processus qui ne contient que du bruit (qui est en général d'un ordre de grandeur inférieur au processus de départ).

Dans la pratique, on travaille avec la donnée de (x_1, \dots, x_n) . Pour appliquer Δ_T , \mathbb{R} utilise la convention : $x_k = x_1$ si $k \leq 0$ et $x_k = x_n$ si $k \geq n$. On a ainsi prolongé la série par des constantes. En particulier, $(\Delta_T^i x)_j$ avec $j \leq iT$ ou $|j - n| \leq iT$ est calculé avec des différences qui contiennent qui viennent du prolongement par des constantes. Donc, pour ces termes, la proposition ci-dessus n'est pas valable. Quand on regarde l'allure du graphe de $(\Delta_T^i x)$, il faut donc s'assurer que iT est petit devant n .

Quand $(\Delta_T^i x)$ et $(\Delta_T^{i+1} x)$ sont « petits » et du même ordre, on considère qu'ils sont constitués de bruit. On en déduit que (x_t) a une tendance polynômiale de degré $i - 1$. Pour comparer les ordres de grandeur de deux séries temporelles, on peut regarder les graphes ou les moyennes empiriques.

PROPOSITION 3.8. *Dans le cas $X_t = m_t + s_t + \epsilon_t$ avec m_t et s_t déterministe, ϵ_t un bruit et s_t qui est T -périodique, le processus $\Delta_T X$ est un processus qui ne contient plus de partie périodique. De plus, si la tendance est linéaire, elle est également supprimée.*

DÉMONSTRATION. Nous avons

$$X_t - X_{t-T} = m_t - m_{t-T} + \epsilon_t - \epsilon_{t-T}.$$

Si, de plus, $m_t = at + b$, alors

$$\begin{aligned} X_t - X_{t-T} &= at + b - a(t-T) - b + \epsilon_t - \epsilon_{t-T} \\ &= aT + \epsilon_t - \epsilon_{t-T}. \end{aligned}$$

□

EXEMPLE 3.9. On considère $x_t = \frac{t}{2} + 3 \cos\left(\frac{t\pi}{6}\right)$ avec $\epsilon_t \sim \mathcal{N}(0; 1)$. Voir la figure 3.5.1 pour les dessins.

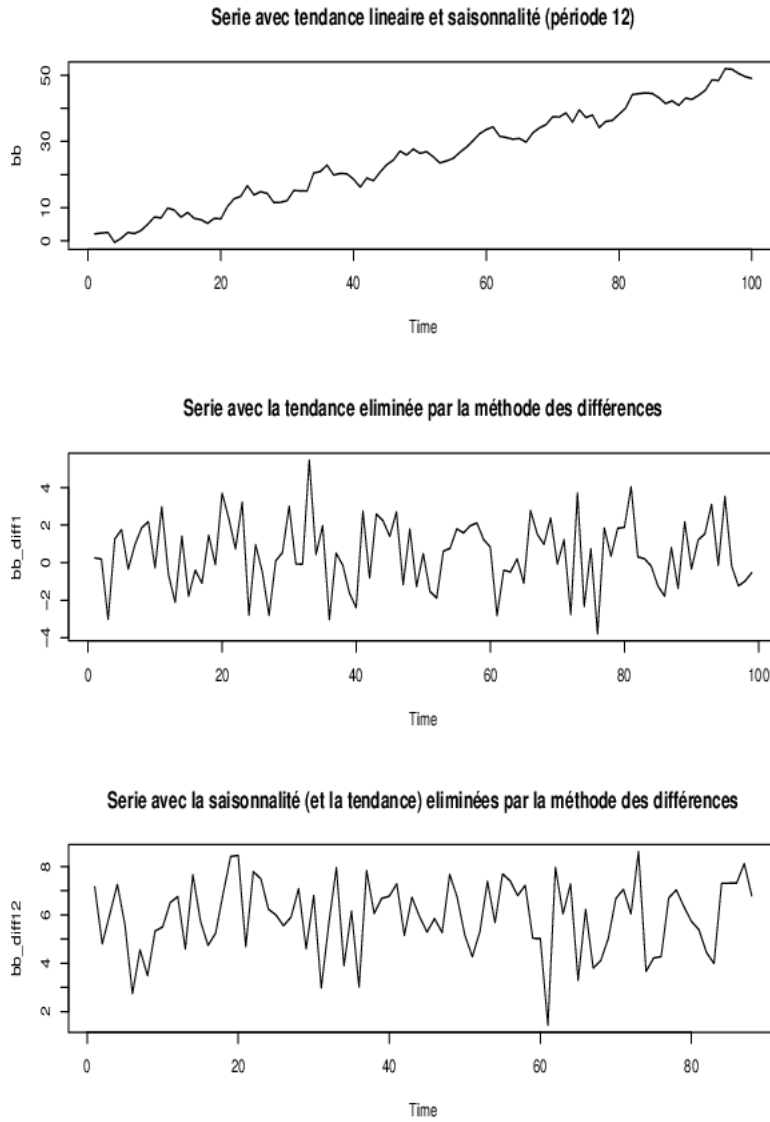


FIGURE 3.5.1. En haut : x_t , au milieu : $x_t - x_{t-1}$, en bas : $x_t - x_{t-12}$ (exemple 3.9).

3.6. Test sur la série résiduelle

On aimerait savoir si on est en présence d'un bruit blanc.

LEMME 3.10. Si $(\epsilon_k)_{k \geq 1}$ est un bruit blanc, alors ses fonctions d'auto-corrélation empiriques vérifient

$$\sqrt{n}\hat{\rho}_n(h) \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0; 1),$$

pour tout h dans \mathbb{N}^* . De plus, les $(\hat{\rho}_n(h))_{h \geq 1}$ sont asymptotiquement indépendants (quand $n \rightarrow +\infty$).

3.6.1. Test d'hypothèses simple. Toute convergence en loi nous donne une idée pour construire un test. On veut tester

(H0) Les (ϵ_k) forment un bruit blanc.

contre

(H1) Les (ϵ_k) ne forment pas un bruit blanc.

On commence par fixer un niveau α . La quantité $\hat{\rho}_n(1)$ s'appelle une statistique (c'est une quantité que l'on peut calculer à partir des données). Disons $\alpha = 0,05$. Le lemme ci-dessus nous dit que, si **(H0)** est vraie,

$$\mathbb{P}(\sqrt{n}\hat{\rho}_n(1) \in [-1,96; 1,96]) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(Z \in [-1,96; 1,96]) = 0,05,$$

avec $Z \sim \mathcal{N}(0; 1)$ (la dernière égalité se lisant dans la table de la loi normale). La procédure de test est donc la suivante

- si $\hat{\rho}_n(1) \in \left[-\frac{1,96}{\sqrt{n}}; \frac{1,96}{\sqrt{n}}\right]$ alors on considère que **(H0)** est vraie,
- dans le cas contraire, on considère que **(H0)** n'est pas vraie.

Si **(H0)** est vraie, la probabilité de rejeter **(H0)** à tort est

$$\begin{aligned} \mathbb{P}(\text{rejeter } \mathbf{(H0)} \mid \mathbf{(H0)}) &\approx \mathbb{P}(Z \notin [-1,96; 1,96]) \\ &= 0,05. \end{aligned}$$

La probabilité de rejeter à tort **(H0)** est donc inférieure ou égale à α . Donc nous avons construit un test de valeur α . La probabilité calculée ci-dessus est la probabilité « d'erreur de première espèce ». Il serait intéressant de borner aussi la probabilité de l'erreur de seconde espèce

$$\mathbb{P}(\text{rejeter } \mathbf{(H1)} \mid \mathbf{(H1)}),$$

mais c'est plus compliqué.

REMARQUE 3.11. Dans le test ci-dessus, **(H0)** et **(H1)** n'ont pas le même statut. L'esprit est qu'on suppose **(H0)**. Sous **(H0)**, une certaine statistique doit suivre une loi $\mathcal{N}(0; 1)$, donc ne doit pas être trop loin de 0. Si cette statistique est trop grande en valeur absolue, il n'est pas crédible qu'elle soit de loi $\mathcal{N}(0; 1)$ et on doit donc rejeter **(H0)** (à contre-cœur).

Si nous voulons faire un test de niveau (ou valeur α) (on dit aussi : de niveau de confiance $1 - \alpha$), il faut trouver M tel que $\mathbb{P}(Z \in [-M; M]) \geq 1 - \alpha$ et rejeter **(H0)** si $|\sqrt{n}\hat{\rho}_n(1)| \geq M$. Ceci assure que la probabilité de rejeter **(H0)** à tort est plus petite que α . Dans la pratique, on regarde dans une table de la loi gaussienne et on prend le M le plus petit possible tel que $\mathbb{P}(Z \in]-\infty; M]) \geq 1 - \alpha/2$ (il n'y a pas forcément de M dans la table tel que nous ayons l'égalité).

3.6.2. Test d'hypothèses plus sophistiqués. Dans la pratique, on utilise des statistiques plus compliquées (qui ont des limites en loi remarquables). Par exemple,

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

converge en loi (quand $n \rightarrow +\infty$) vers une loi $\chi^2(h)$ (une loi du chi-deux à h degrés de liberté). On peut tester les hypothèses ci-dessus en disant que l'on rejette **(H0)** quand $Q_n \geq \chi_{h,1-\alpha}^2$ (qui est

le α -quantile de la loi limite). Voir le cours de statistiques pour la définition du mot « quantile ». Le langage R utilise la statistique

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

(qui est aussi, asymptotiquement, de loi $\chi^2(h)$ (voir les feuilles de TP et la fonction `Box.test`)).

REMARQUE 3.12. Pour des raisons de vitesse de convergence, on se limitera à un `lag` inférieur à 10 dans l'utilisation de `Box.test`. Soit X de loi $\chi^2(h)$. La fonction `Box.test` avec un `lag` égal à h renvoie la valeur numérique de $\mathbb{P}(X > Q_{n, LB})$. Notre test (de niveau α) consiste à rejeter (**H0**) quand $Q_{n, LB} > \chi_{h, 1-\alpha}$. Nous remarquons que

$$Q_{n, LB} > \chi_{h, 1-\alpha} \iff \mathbb{P}(X > Q_{n, LB}) < \alpha.$$

Donc notre test revient à rejeter (**H0**) quand $\mathbb{P}(X > Q_{n, LB}) < \alpha$.

3.7. Exemple : un système proies-prédateurs

On s'intéresse au nombre de lynx capturés par la compagnie de la Baie d'Hudson, données annuelles de 1821 à 1934¹. Nous noterons x la série (voir le graphique des données dans la figure 3.7.1). Nous traçons les auto-corrélations dans la figure 3.7.2. Au vu des auto-corrélations, la série

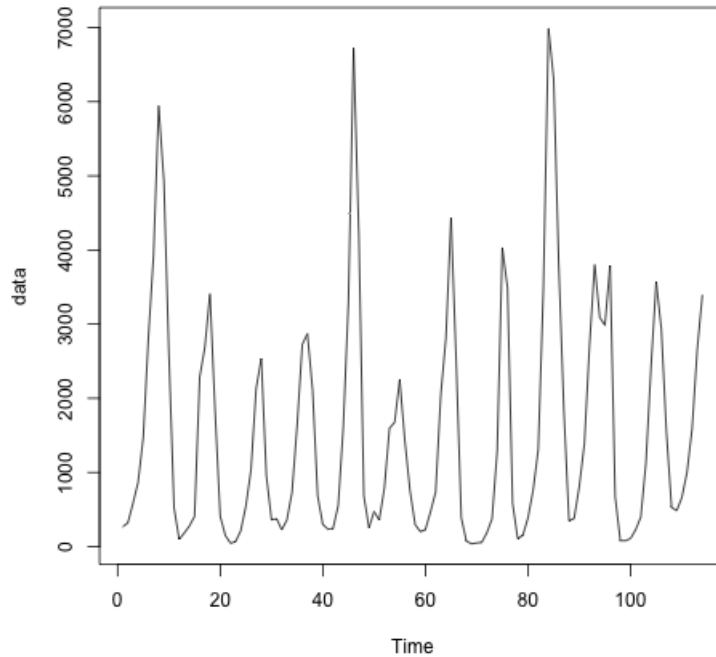


FIGURE 3.7.1. Lynx capturés

semble être périodique de période 10. Nous utilisons ensuite la méthode des moyennes mobiles pour faire apparaître la composante saisonnière, la tendance et la partie aléatoire de la série x (voir figure 3.7.3). Testons maintenant la blancheur des résidus au niveau $\alpha = 0,05$. Nous obtenons une p -valeur inférieure à $2,2 \cdot 10^{-16}$ donc inférieure à α . Il y a donc quelque chose qui ne va pas dans

1. pour télécharger le fichier :
<http://math.unice.fr/~rubentha/enseignement/annual-numer-of-lynx-trapped-ma.txt>

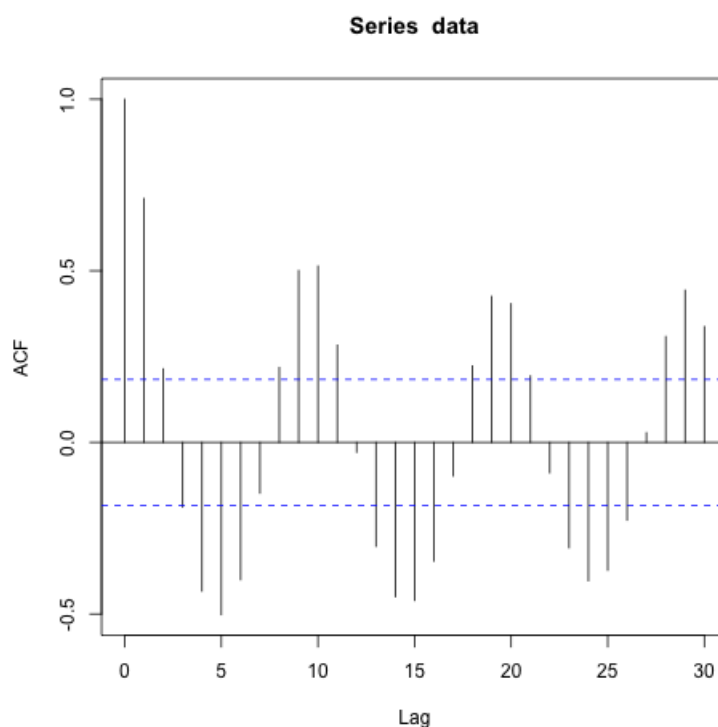


FIGURE 3.7.2. Auto-corrélations

notre modèle. C'est a priori surprenant puisque nous avons fait peu d'hypothèses. En fait, la seule hypothèse que nous faisons c'est qu'il y a quelque chose de périodique dans x . Et c'est justement cette assertion qui est fausse.

La taille de la population de lynx suit une équation de Lotka-Volterra (dans un système proies-prédateurs). Le nombre de prises est proportionnel à la taille de la population. Les lynx se nourrissent essentiellement de lièvres. En gros, quand il y a beaucoup de lièvres à manger, la population de lynx augmente, jusqu'à ce que les lynx mangent une grande partie des lièvres et se retrouvent affamés. La population de lynx baisse alors. Les lièvres étant moins chassés, leur population se met à augmenter et le cycle recommence. Voir les livres [Ber17, HSD13, WH99] pour plus de détail sur ce système proies-prédateurs (en temps continu, mais c'est pareil en temps discret).

La taille de la population a un comportement quasi-périodique mais n'est pas périodique, ce qui fait que notre méthode de moyenne mobile fonctionne mal.

3.8. Feuille d'exercices numéro 3 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf. C'est ce qui sera demandé au partiel.

3.8.1. Données AirPassengers. Nous étudions la série chronologique du nombre de passagers par mois (en milliers) dans les transports aériens, de 1949 à 1960. Cette série est disponible sous R (`AirPassengers`).

- (1) Estimation paramétrique de la tendance
 - (a) Représenter graphiquement la série. Ce processus vous semble-t-il stationnaire ? Présente-t-il des tendances et saisonnalités ?

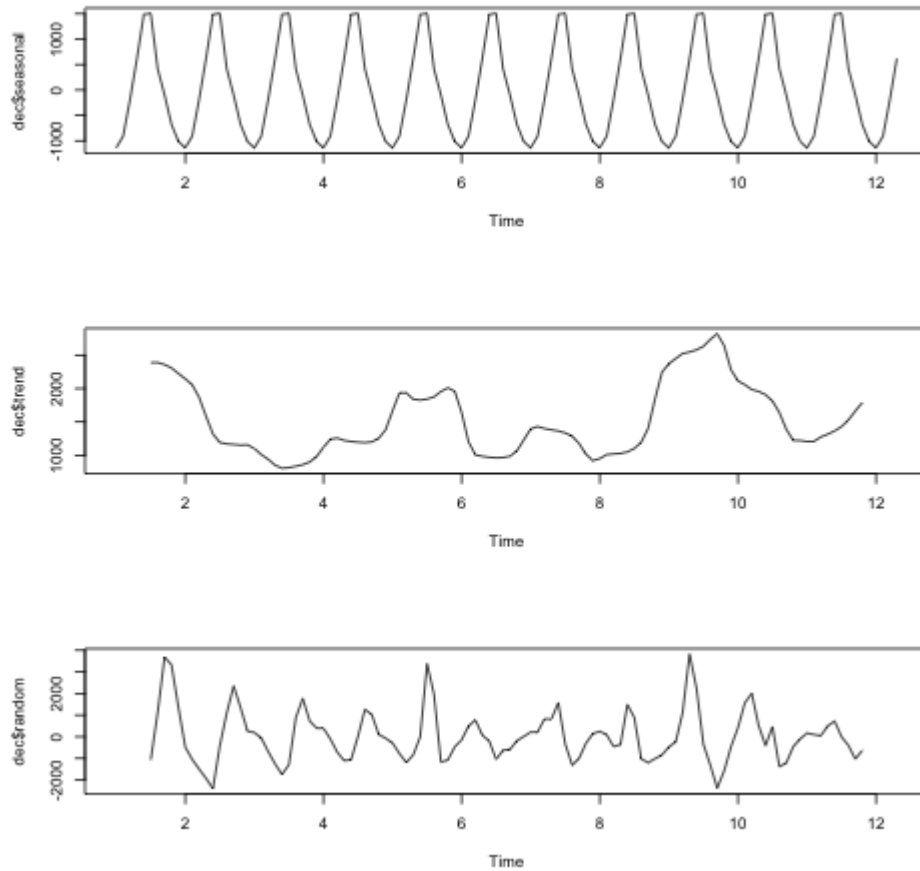


FIGURE 3.7.3. Décomposition par moyenne mobile

- (b) Estimer les paramètres d'une tendance linéaire $at + b$ par la méthode des moindres carrés.
- (c) Supprimer cette tendance et représenter graphiquement la série ainsi obtenue. Vérifier que la série des résidus est de moyenne nulle.
- (d) Calculer et représenter l'auto-corrélation de la série des résidus.
- (2) Méthode des différences
 - (a) Appliquer la méthode des différences pour enlever la tendance et la saisonnalité. Préciser la période de la saisonnalité, le degré du polynôme de tendance.
 - (b) La série obtenue semble-t-elle stationnaire ?
- (3) Méthode des moyennes mobiles
 - (a) Appliquer la méthode des moyennes mobiles pour enlever la tendance et la saisonnalité.
 - (b) Faire un test pour savoir si la série obtenue est un bruit blanc (utiliser fonction `Box.test`). On prendra (H_0) =bruit blanc, (H_1) =pas bruit blanc, probabilité de rejeter (H_0) à tort $=\alpha$ avec $\alpha = 0,05$.

3.8.2. Données simulées. Récupérer la série temporelle contenue dans le fichier <http://math.unice.fr/~rubentha/enseignement/simulation.dat>. Cette série a été simulée à partir d'un processus relativement simple. Essayer de le retrouver !

- (1) Analyser qualitativement cette série temporelle.
- (2) Pouvez-vous proposer une modélisation pour cette série (c'est-à-dire définir le processus X_t qui a généré cette série)? Indication : la partie saisonnière est de la forme $a \cos(t\pi/b)$ ou $a \sin(t\pi/b)$... à vous de deviner!

Appendice : mise en œuvre sous R. Les indications concernent le premier exercice (on réutilisera les mêmes commandes pour l'exercice suivant). Nous mentionnons ici des commandes utiles dans les questions où on en a besoin pour la première fois.

La commande `as.numeric` transforme une série temporelle en vecteur (ce qui est utile pour certaines opérations).

Question 1d. Utiliser la commande `acf(...)` pour calculer les autocorrélations empiriques (voir corrigé du premier TP).

Question 2a. La fonction `diff.ts(serie,lag=T,difference=k)` (ou `diff` tout court) permet d'appliquer l'opérateur de différenciation Δ_T^k .

Question 3a. 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"),`
 choisir "additive","multiplicative" mais pas les deux en même temps)

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

Question 3b. 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). La première valeur affichée est la statistique du test. La p -value est la probabilité d'obtenir une valeur aussi élevée sous l'hypothèse nulle. Si nous voulons faire un test au niveau de confiance 0,95 : dans le cas où cette valeur est $< 0,05$, nous rejetons l'hypothèse nulle.

Modélisation des séries stationnaires

4.1. Auto-corrélation partielle

DÉFINITION 4.1. Pour un processus stationnaire (X_t) , le coefficient de corrélation partielle (on dit aussi corrélation partielle) entre X_n et X_1 est

$$r_{X_2, \dots, X_n}(X_1, X_n) = \frac{1}{\sigma(0)} \text{Cov}(X_1 - \mathbb{E}(X_1|X_2, \dots, X_{n-1}), X_n - \mathbb{E}(X_n|X_2, \dots, X_{n-1})),$$

le coefficient d'auto-corrélation partielle en h est défini par

$$\begin{cases} r(h) = r_{X_2, \dots, X_{h-1}}(X_1, X_h) & \text{pour } h \geq 2, \\ r(1) = \rho(1), \\ r(0) = 1, \\ r(-h) = r(h) & \text{pour } h \geq 1. \end{cases}$$

Pour un processus stationnaire,

$$r(h) = r_{X_{n+1}, \dots, X_{n+h-1}}(X_n, X_{n+h}) \text{ pour tout } n.$$

Donc on va pouvoir estimer ces coefficients en faisant des moyennes empiriques (c'est ce que fait la fonction `pacf` de R)

4.2. Les processus auto-régressifs

DÉFINITION 4.2. Un processus (X_t) est dit auto-régressif d'ordre p centré s'il vérifie

$$(4.1) \quad X_t = \epsilon_t + \sum_{j=1}^p a_j X_{t-j}, \text{ pour tout } t \geq 0$$

(avec $p \in \mathbb{N}^*$, $a_p \neq 0$) avec des ϵ_t qui forment un bruit blanc centré de variance σ^2 ($p \in \mathbb{N}^*$, $a_1, \dots, a_p \in \mathbb{R}$), tels que ϵ_t est indépendant de X_{t-1}, X_{t-2}, \dots pour tout t . Par convention : $X_{-k} = 0$ pour tout k dans \mathbb{N}^* . On dira aussi que (X_t) est un processus $AR(p)$.

On dit que le X_t « s'explique » par les p observations précédentes $(X_{t-1}, X_{t-2}, \dots, X_{t-p})$. Dans cette définition, le processus (ϵ_t) s'appelle processus des innovations.

PROPOSITION 4.3. On associe le polynôme de $\mathbb{R}[X]$ suivant à l'équation (4.1)

$$A(X) = 1 - a_1 X - \dots - a_p X^p.$$

Si les racines (dans \mathbb{C}) de ce polynôme sont toutes de module strictement supérieur à 1 alors il existe un processus stationnaire (X_t) vérifiant (4.1) et tel que ϵ_t est le bruit d'innovation pour ce processus (c'est à dire que, pour tout t , ϵ_t est indépendant de X_{t-1}, X_{t-2}, \dots).

REMARQUE 4.4. La condition « racines de A de module > 1 » n'est pas nécessaire à l'existence d'une suite stationnaire vérifiant (4.1). Voir le chapitre 5 de l'édition en anglais de [GM97] pour plus de détails.

DÉMONSTRATION. On peut toujours développer la fraction rationnelle $1/A(z)$ ($z \in \mathbb{C}$) dans un voisinage de 0

$$\frac{1}{A(z)} = \frac{1}{1 - a_1 z - \dots - a_p z^p} = 1 + (a_1 z + \dots + a_p z^p) + (a_1 z + \dots + a_p z^p)^2 - \dots$$

$$= 1 + \alpha_1 z + \alpha_2 z^2 + \dots$$

pour certains coefficients $\alpha_1, \alpha_2, \dots$ (on remarque que ces coefficients sont dans \mathbb{R}). Plus précisément, nous avons pour tout z ,

$$\frac{1}{A(z)} = \prod_{i=1}^p \left(\frac{1}{1 - u_i z} \right),$$

où l'ensemble des racines de A est $\{1/u_1, \dots, 1/u_p\}$ (les racines apparaissent avec leur multiplicité). Donc, pour z tel que $|z| < \inf_i (1/|u_i|)$, nous pouvons développer en produit de séries entières

$$\frac{1}{A(z)} = \prod_{i=1}^p \left(\sum_{k=0}^{+\infty} u_i^k z^k \right).$$

Quand on développe le produit ci-dessus, le coefficient de z^n est

$$\alpha_n = \sum_{\substack{j_1, \dots, j_p \geq 0 \\ j_1 + \dots + j_p = n}} u_1^{j_1} \dots u_p^{j_p}.$$

Donc

$$(4.2) \quad |\alpha_n| \leq \left(\sup_i |u_i| \right)^n \frac{(n+1)^p}{p!}$$

(nous utilisons : $\#\{(j_1, \dots, j_p) \in (\mathbb{N})^p : j_1 + \dots + j_p = n\} = \frac{(n+1)^p}{p!}$, qui nécessite un petit calcul).

On suppose que l'on dispose d'un bruit blanc $(\epsilon_k)_{k \in \mathbb{Z}}$. On fixe t dans \mathbb{N} et on s'intéresse à la suite $(Y_k = \epsilon_t + \alpha_1 \epsilon_{t-1} + \dots + \alpha_k \epsilon_{t-k})_{k \geq 1}$. La série $\sum_{n \geq 1} |\alpha_n|$ est convergente (petit exercice sur la convergence des séries, c'est ici que l'on utilise que $\sup_i |u_i| < 1$) donc la série $\sum_{n \geq 1} \alpha_n^2$ aussi. En particulier, pour tout $\delta > 0$, il existe N dans \mathbb{N} tel que

$$n \geq N \Rightarrow \forall k \geq 0, \alpha_n^2 + \alpha_{n+1}^2 + \dots + \alpha_{n+k}^2 < \delta.$$

Pour tout $n > N$ et $k \geq 0$, nous avons alors

$$\begin{aligned} \mathbb{E}((Y_n - Y_{n+k})^2) &= \mathbb{E}((\alpha_{n+1} \epsilon_{t-n-1} + \dots + \alpha_{n+k} \epsilon_{t-n-k})^2) \\ &= (\alpha_{n+1}^2 + \dots + \alpha_{n+k}^2) \sigma^2 \\ &< \delta \end{aligned}$$

Donc la suite $(Y_n)_{n \geq 1}$ est une suite de Cauchy dans $L^2(\mathbb{P})$ (l'espace des variables aléatoires réelles de carré intégrable, muni de la norme L^2). Comme cette espace est complet, la suite $(Y_n)_{n \geq 1}$ converge dans L^2 vers une limite que nous noterons X_t . Nous pouvons écrire

$$(4.3) \quad X_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots$$

(en considérant le terme de droite comme une limite dans L^2).

Nous aimerions montrer que cette limite a aussi lieu p.s. Soit $\delta > 0$. For all $n \geq 1$,

$$\begin{aligned} \mathbb{P}(|Y_n - X_t| \geq \delta) &= \mathbb{P}((\alpha_{n+1} \epsilon_{t-n-1} + \alpha_{n+2} \epsilon_{t-n-2} + \dots)^2 \geq \delta^2) \\ (\text{inégalité de Markov}) &\leq \frac{\mathbb{E}((\alpha_{n+1} \epsilon_{t-n-1} + \alpha_{n+2} \epsilon_{t-n-2} + \dots)^2)}{\delta^2} \\ &= \frac{\sigma^2(\alpha_{n+1}^2 + \alpha_{n+2}^2 + \dots)}{\delta^2}. \end{aligned}$$

Et donc

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}(|Y_n - X_t| \geq \delta) &\leq \sum_{n \geq 1} \frac{\sigma^2(\alpha_{n+1}^2 + \alpha_{n+2}^2 + \dots)}{\delta^2} \\ &= \frac{\sigma^2}{\delta^2} \sum_{n \geq 2} (n-1) \alpha_n^2 \\ &< \infty \end{aligned}$$

(petit exercices sur les séries à partir de l'équation (4.2)). Le lemme de Borel-Cantelli nous dit que, p.s., il existe un N tel que, pour $n > N$, $|Y_n - X_t| < \delta$. Ceci est valable pour tout $\delta > 0$, donc (attention, nous sautons une étape de la démonstration) $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X_t$.

Montrons que les X_t que nous venons de définir vérifient la relation de récurrence (4.1). Nous avons, pour tout t ,

$$\begin{aligned} X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} &= \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots \\ &\quad - a_1 (\epsilon_{t-1} + \alpha_1 \epsilon_{t-2} + \alpha_2 \epsilon_{t-3} + \dots) \\ &\quad - \dots \\ &\quad - a_p (\epsilon_{t-p} + \alpha_1 \epsilon_{t-p-1} + \alpha_2 \epsilon_{t-p-2} + \dots) \\ (\text{convention } \alpha_0 = 1) &= \epsilon_t + \epsilon_{t-1}(\alpha_1 - a_1) + \epsilon_{t-2}(\alpha_2 - a_1 \alpha_1 - a_3) \\ &\quad + \dots \\ &\quad + \epsilon_{t-k}(\alpha_k - \sum_{1 \leq i \leq k \wedge p} \alpha_{k-i} a_i) \\ &\quad + \dots \end{aligned}$$

Nous avons, pour tout z dans un voisinage adéquat de 0,

$$\begin{aligned} 1 &= A(z) \times \frac{1}{A(z)} \\ &= (1 - a_1 z - \dots - a_p z^p) \times (1 + \alpha_1 z + \alpha_2 z^2 + \dots). \end{aligned}$$

Si nous développons ce dernier produit, nous trouvons une série entière dans laquelle le coefficient de z^n ($n \geq 1$) est

$$\alpha_n - \sum_{1 \leq i \leq n \wedge p} \alpha_{n-i} a_i,$$

qui doit être nul puisque la série vaut 1. Donc

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \epsilon_t.$$

On montre facilement que la suite $(X_t)_{t \geq 1}$ est stationnaire (exercice) (voir la définition 3.2). \square

PROPOSITION 4.5. *S'il existe un processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant la relation de récurrence de l'équation (4.1) alors sa fonction d'auto-covariance vérifie :*

$$(4.4) \quad \begin{aligned} \text{pour } h > 0, \sigma(h) &= a_1 \sigma(h-1) + a_2 \sigma(h-2) + \dots + a_p \sigma(h-p), \\ \text{et } \sigma(0) &= \sigma^2 + a_1 \sigma(1) + a_2 \sigma(2) + \dots + a_p \sigma(p), \end{aligned}$$

et sa fonction d'auto-corrélation vérifie

$$\text{pour } h > 1, \rho(h) = a_1 \rho(h-1) + a_2 \rho(h-2) + \dots + a_p \rho(h-p).$$

DÉMONSTRATION. Nous avons pour tout t et tout $h \geq 1$,

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}(X_t X_{t+h}) \\ \sigma(h) &= \mathbb{E}(X_t (\epsilon_{t+h} + \sum_{i=1}^p a_i X_{t+h-i})) \\ \sigma(h) &= 0 + \sum_{i=1}^p a_i \sigma(h-i) \end{aligned}$$

et

$$\begin{aligned} \sigma(0) = \mathbb{V}(X_t) &= \mathbb{E}(X_t (\epsilon_t + \sum_{i=1}^p a_i X_{t-i})) \\ &= \mathbb{E}(X_t \epsilon_t) + \sum_{i=1}^p a_i \sigma(i) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}((\epsilon_t + \sum_{i=1}^p a_i X_{t-i}) \epsilon_t) + \sum_{i=1}^p a_i \sigma(i) \\
(\text{car } \epsilon_t \text{ ind. de } X_j \text{ pour } j < t) &= \sigma^2 + \sum_{i=1}^p a_i \sigma(i).
\end{aligned}$$

□

Le polynôme caractéristique de la relation de récurrence de l'équation (4.4) est

$$\begin{aligned}
B(X) = X^p - a_1 X^{p-1} - \dots - a_p &= X^p \left(1 - \frac{a_1}{X} - \dots - \frac{a_p}{X^p} \right) \\
&= X^p A \left(\frac{1}{X} \right).
\end{aligned}$$

Donc, si les racines de A sont de modules > 1 , alors celle de B sont de module < 1 . Si les racines de B sont distinctes égales à $\lambda_1, \lambda_2, \dots, \lambda_p$, les solutions de (4.4) sont de la forme

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

(voir [Gou08], p. 196 pour plus de détails). On en déduit le résultat suivant.

LEMME 4.6. *S'il existe un processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant la relation de récurrence de l'équation (4.1) et si le polynôme de la relation de récurrence a des racines qui sont toutes de module > 1 , alors la fonction d'auto-covariance du processus décroît exponentiellement quand $h \rightarrow +\infty$ (et donc sa fonction d'auto-corrélation a le même comportement).*

PROPOSITION 4.7. *S'il existe un processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant la relation de récurrence de l'équation (4.1) alors sa fonction d'auto-corrélation partielle vérifie :*

$$r(h) = 0 \text{ si } h \geq p + 1.$$

DÉMONSTRATION. Soit $h \geq p + 1$. Nous calculons pour t quelconque :

$$\begin{aligned}
\mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}) &= \mathbb{E}(\epsilon_{t+h} + a_1 X_{t+h-1} + \dots + a_p X_{t+h-p} | X_{t+1}, \dots, X_{t+h-1}) \\
&= a_1 X_{t+h-1} + \dots + a_p X_{t+h-p},
\end{aligned}$$

$$X_{t+h} - \mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}) = \epsilon_{t+h},$$

$$\begin{aligned}
\mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1}) &= \\
\mathbb{E} \left(\frac{1}{a_p} (X_{t+p} - \epsilon_{t+p} - a_1 X_{t+p-1} - \dots - a_{p-1} X_{t+1}) | X_{t+1}, \dots, X_{t+h-1} \right) &= \\
\frac{1}{a_p} (X_{t+p} - a_1 X_{t+p-1} - \dots - a_{p-1} X_{t+1}) - \frac{1}{a_p} \mathbb{E}(\epsilon_{t+p} | X_{t+1}, \dots, X_{t+h-1}), &
\end{aligned}$$

$$X_t - \mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1}) = \frac{\epsilon_{t+p} + \mathbb{E}(\epsilon_{t+p} | X_{t+1}, \dots, X_{t+h-1})}{a_p}.$$

Donc

$$\begin{aligned}
\text{Cov}(X_{t+h} - \mathbb{E}(X_{t+h} | X_{t+1}, \dots, X_{t+h-1}), X_t - \mathbb{E}(X_t | X_{t+1}, \dots, X_{t+h-1})) \\
= \mathbb{E}(\epsilon_{t+h} \mathbb{E}(\epsilon_{t+p} | X_{t+1}, \dots, X_{t+h-1})) = 0,
\end{aligned}$$

car ϵ_{t+h} est indépendant de $X_{t+1}, \dots, X_{t+h-1}$.

□

4.3. Les processus en moyenne mobile

DÉFINITION 4.8. Un processus en moyenne mobile d'ordre q est un processus de la forme

$$(4.5) \quad X_t = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}, \text{ pour tout } t \geq 0$$

(avec $q \in \mathbb{N}^*$, $b_q \neq 0$) avec des $(\epsilon_t)_{t \in \mathbb{Z}}$ qui forment un bruit blanc centré de variance σ^2 . On dira aussi que (X_t) est un processus $MA(q)$.

Pour un bruit blanc centré (ϵ_t) , un tel processus existe toujours et est toujours stationnaire.

PROPOSITION 4.9. L'auto-covariance d'un processus $MA(q)$ vérifiant l'équation (4.5) vérifie, pour $h \geq 0$,

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

sous la convention $b_0 = 1$.

DÉMONSTRATION. Nous calculons

$$\begin{aligned} \mathbb{E}(X_t X_{t+h}) &= \mathbb{E} \left(\left(\sum_{k=0}^q \epsilon_{t-k} b_k \right) \times \left(\sum_{k=0}^q \epsilon_{t+h-k} b_k \right) \right) \\ &= \begin{cases} 0 & \text{si } h > q, \\ \sum_{k=q}^{q-h} \mathbb{E}(\epsilon_{t-k} b_k \epsilon_{t-k} b_{k+h}) & \text{sinon.} \end{cases} \end{aligned}$$

□

De la proposition 4.3 et de l'équation (4.3), nous déduisons le corollaire suivant.

COROLLAIRE 4.10. Sous les hypothèses de la proposition 4.3, le processus $AR(p)$ stationnaire vérifiant l'équation de récurrence (4.1) peut s'écrire comme un processus AM d'ordre infini (au sens de l'équation (4.3) (dans laquelle la limite a lieu p.s. et dans L^2).

La démonstration du résultat suivant étant similaire à la démonstration de la proposition 4.3, nous le citons comme un simple corollaire.

COROLLAIRE 4.11. Si $(X_t)_{t \geq 0}$ est un processus $AM(q)$ satisfaisant l'équation (4.5) et si le polynôme $B(X) = 1 + b_1 X + \cdots + b_q X^q$ n'admet que des racines de module > 1 , alors (X_t) peut s'écrire comme un processus AR d'ordre infini et tel que ϵ_t est le bruit d'innovation pour ce processus.

En particulier, sa fonction d'auto-corrélation partielle vérifie :

$$r(h) \xrightarrow{h \rightarrow +\infty} 0.$$

En effet, en partant de l'équation (4.5), on peut exprimer ϵ_t en fonction de X_t, X_{t-1}, \dots pour tout t (de même qu'en partant de l'équation (4.1), on peut exprimer X_t en fonction de $\epsilon_t, \epsilon_{t-1}, \dots$ pour tout t). Pour plus de détail voir [GM97] (ces questions sont développées au chapitre 5 de l'édition en anglais de ce livre).

4.4. Les processus mixtes ARMA(p,q).

DÉFINITION 4.12. Un processus auto-régressif en moyenne mobile d'ordres p, q (tels que $p \geq 0, q \geq 0, p+q \geq 1$) est un processus qui peut s'écrire

$$(4.6) \quad X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=0}^q b_j \epsilon_{t-j}, \text{ pour tout } t \geq 0,$$

où les (ϵ_j) sont des bruits blancs centrés de variance σ^2 et $b_0 = 1$. Par convention : $X_{-k} = \epsilon_{-k}$ pour tout k dans \mathbb{N}^* . On dira aussi que (X_t) est un processus $ARMA(p, q)$.

On peut toujours écrire

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

d'où la relation

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

Cette dernière quantité est nulle dès que $h > q$. Les auto-covariances d'un processus $ARMA(p, q)$ vérifie donc la même relation de récurrence que celles d'un processus $AR(p)$ à partir d'un certain rang. Donc ces auto-covariances (ainsi que les auto-corrélations correspondantes) convergent exponentiellement vite vers 0 à l'infini.

PROPOSITION 4.13. *On s'intéresse à la relation de récurrence (4.6) ci-dessus. On définit les polynômes*

$$A(X) = 1 - a_1 X - \cdots - a_p X^p, \quad B(X) = 1 + b_1 X + \cdots + b_q X^q.$$

Si les racines de A et B sont de modules > 1, alors il existe un processus stationnaire vérifiant la relation (4.6) et tel que ϵ_t est le bruit d'innovation pour ce processus.

Si les racines de A et B sont de module > 1 alors : « (X_t) ne vérifie pas de relation de récurrence plus courte que (4.6) » implique « A et B n'ont pas de racine commune ». Si on suppose, de plus, que les (ϵ_t) on des densités par rapport à la mesure de Lebesgue, alors les deux propositions sont équivalentes.

REMARQUE 4.14. Pour plus de détail voir [GM97] (ces questions sont développées au chapitre 5 de l'édition en anglais de ce livre).

Avant de nous lancer dans la démonstration de la proposition, nous avons besoin de définir quelques notions. On définit l'opérateur de décalage sur les suites stationnaires :

$$L : (Y_t)_{t \geq 0} \mapsto \left(L(Y)_t = \begin{cases} Y_{t-1} & \text{si } t \geq 1 \\ 0 & \text{si } t = 0 \end{cases} \right)_{t \geq 0}$$

(ici, (Y_t) est une suite stationnaire). Pour un polynôme $P \in \mathbb{C}[X]$ ($P(X) = \alpha_0 + \alpha_1 X + \cdots + \alpha_n X^n$, n étant le degré de P), nous définissons l'opérateur $P(L)$ par :

$$P(L) : (Y_t)_{t \geq 0} \mapsto (P(L)(Y)_t) \text{ avec } P(L)(Y)_t = \alpha_0 Y_t + \alpha_1 L(Y)_t + \cdots + \alpha_n L^n(Y)_t.$$

Nous remarquons que pour deux polynômes P et Q :

$$P(L)Q(L) = (PQ)(L),$$

où le produit de gauche est la composition.

LEMME 4.15. *Pour λ un nombre complexe de module < 1, l'opérateur $1 - \lambda L$ admet pour inverse*

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

DÉMONSTRATION. Pour toute suite stationnaire $(Y_t)_{t \geq 0}$, nous regardons à t fixé :

$$\sum_{k=0}^{+\infty} ((\lambda L)^k Y)_t = \sum_{k=0}^t \lambda^k Y_{t-k},$$

qui est une somme finie. Donc la suite

$$\left(\sum_{k=0}^{+\infty} (\lambda L)^k Y \right)_{t \geq 0}$$

est bien définie. Nous avons

$$\begin{aligned} Y &= (Y_0, Y_1, \dots) \xrightarrow{\sum_{k=0}^{+\infty} (\lambda L)^k} (Y_0, Y_1 + \lambda Y_0, Y_2 + \lambda Y_1 + \lambda^2 Y_0, \dots) \\ &\xrightarrow{1 - \lambda L} (Y_0, (Y_1 + \lambda Y_0) - \lambda Y_0, (Y_2 + \lambda Y_1 + \lambda^2 Y_0) - \lambda(Y_1 + \lambda Y_0) - \lambda^2 Y_0, \dots) \\ &= (Y_0, Y_1, Y_2, \dots). \end{aligned}$$

Donc $(1 - \lambda L) \sum_{k=0}^{+\infty} (\lambda L)^k = \text{Id}$. On montre facilement de même que $\sum_{k=0}^{+\infty} (\lambda L)^k (1 - \lambda L) = \text{Id}$. \square

DÉMONSTRATION PARTIELLE DE LA PROPOSITION 4.13. Supposons que A et B n'ont que des racines de module > 1 . Supposons, de plus, que A et B ont une racine commune, nous allons montrer que (X_t) et (ϵ_t) vérifie une relation plus courte que (4.6). Notons $1/\lambda$ une racine commune à A et B . Nous avons donc $p \geq 1$ et $q \geq 1$. Nous pouvons écrire

$$A(X) = (1 - \lambda X)A_1(X), \quad B(X) = (1 - \lambda X)B_1(X)$$

avec A_1 polynôme de degré $p - 1$ et B_1 polynôme de degré $q - 1$. Nous avons

$$\begin{aligned} A(L)X &= B(L)\epsilon \\ (1 - \lambda L)A_1(L)X &= (1 - \lambda L)B_1(L)\epsilon \\ (1 - \lambda L)^{-1}(1 - \lambda L)A_1(L)X &= (1 - \lambda L)^{-1}(1 - \lambda L)B_1(L)\epsilon \\ A_1(L)(X) &= B_1(L)\epsilon. \end{aligned}$$

□

EXEMPLE 4.16. Soient

$$A(X) = \left(1 - \frac{X}{2}\right) \left(1 - \frac{X}{3}\right), \quad B(X) = \left(1 - \frac{X}{2}\right) \left(1 + \frac{X}{4}\right).$$

Nous avons

$$\begin{aligned} A(X) &= 1 - \frac{5}{6}X + \frac{X^2}{6}, \\ B(X) &= 1 - \frac{X}{4} - \frac{X^2}{8}. \end{aligned}$$

On s'intéresse au processus $ARMA(2, 2)$ (stationnaire) vérifiant la relation de récurrence :

$$(4.7) \quad X_t - \frac{5}{6}X_{t-1} + \frac{1}{6}X_{t-2} = \epsilon_t - \frac{1}{4}\epsilon_{t-1} - \frac{1}{8}\epsilon_{t-2}.$$

Les polynômes caractéristiques de cette relation sont les A, B ci-dessus. Ils ont une racine commune (2). Nous allons montrer que $(X_t), (\epsilon_t)$ vérifient une relation de récurrence plus courte que (4.7). Soit la suite

$$\begin{aligned} (Y_0, Y_1, Y_2, \dots, Y_t, \dots) &= (X_0, X_1 - \frac{5}{6}X_0, X_2 - \frac{5}{6}X_1 + \frac{1}{6}X_0, \dots, X_t - \frac{5}{6}X_{t-1} + \frac{1}{6}X_{t-2}, \dots) \\ &= (\epsilon_0, \epsilon_1 - \frac{1}{4}\epsilon_0, \epsilon_2 - \frac{1}{4}\epsilon_1 - \frac{1}{8}\epsilon_0, \dots, \epsilon_t - \frac{1}{4}\epsilon_{t-1} - \frac{1}{8}\epsilon_{t-2}, \dots). \end{aligned}$$

Par convention, $X_{-k} = \epsilon_{-k} = 0$ pour tout k dans \mathbb{N}^* . Et donc :

$$X_0 = X_0 - \frac{5}{6}X_{-1} + \frac{1}{6}X_{-2}, \quad X_1 = X_1 - \frac{5}{6}X_0 + \frac{1}{6}X_{-1}.$$

Calculons maintenant

$$(Y_0, Y_1 + \frac{1}{2}Y_0, Y_2 + \frac{1}{2}Y_1 + \frac{1}{2^2}Y_2, \dots, Y_t + \frac{1}{2}Y_{t-1} + \dots + \frac{1}{2^t}Y_0, \dots).$$

Le terme général de cette suite est

$$Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 = \sum_{k=0}^{+\infty} \frac{1}{2^k} Y_{t-k},$$

puisque les termes de la somme sont nuls à parti d'un certain rang. Donc, pour tout $t \geq 0$:

$$\begin{aligned} Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 &= \sum_{k=0}^{+\infty} \frac{1}{2^k} \left(X_{t-k} - \frac{5}{6}X_{t-k-1} + \frac{1}{6}X_{t-k-2} \right) \\ &= X_t + \left(-\frac{5}{6}X_{t-1} + \frac{1}{2}X_{t-1} \right) + \sum_{j=2}^{+\infty} X_{t-j} \left(\frac{1}{2^j} - \frac{1}{2^{j-1}} \frac{5}{6} + \frac{1}{2^{j-2}} \frac{1}{6} \right) \\ &= X_t - \frac{1}{3}X_{t-1} + \sum_{j=2}^{+\infty} X_{t-j} \frac{1}{2^j} A(2) \end{aligned}$$

$$= X_t - \frac{1}{3}X_{t-1}.$$

De même :

$$\begin{aligned} Y_t + \frac{1}{2}Y_{t-1} - \dots + \frac{1}{2^t}Y_0 &= \sum_{k=0}^{+\infty} \frac{1}{2^k} \left(\epsilon_{t-k} - \frac{1}{4}\epsilon_{t-k-1} - \frac{1}{8}\epsilon_{t-k-2} \right) \\ &= \epsilon_t + \left(-\frac{1}{4}\epsilon_{t-1} + \frac{1}{2}\epsilon_{t-1} \right) + \sum_{j=2}^{+\infty} \epsilon_{t-j} \left(\frac{1}{2^j} - \frac{1}{2^{j-1}}\frac{1}{4} - \frac{1}{2^{j-2}}\frac{1}{8} \right) \\ &= \epsilon_t + \frac{1}{4}\epsilon_{t-1} + \sum_{j=2}^{+\infty} \epsilon_{t-j} \frac{1}{2^j} B(2) \\ &= \epsilon_t + \frac{1}{4}\epsilon_{t-1}. \end{aligned}$$

Nous avons donc la relation :

$$X_t - \frac{1}{3}X_{t-1} = \epsilon_t + \frac{1}{4}\epsilon_{t-1}, \text{ pour tout } t \in \mathbb{N}^*.$$

4.5. Tableau des propriétés

Modèle	$MA(q)$	$AR(p)$	$ARMA(p, q)$
auto-corrélation	$\rho(h) = 0$ si $h > q$	$\rho(h) \xrightarrow{h \rightarrow +\infty} 0$	$\rho(h) \xrightarrow{h \rightarrow +\infty} 0$
auto-corrélation partielle	$r(h) \xrightarrow{h \rightarrow +\infty} 0$	$r(h) = 0$ si $h > p$	$r(h) \xrightarrow{h \rightarrow +\infty} 0$

TABLE 1. Tableau des propriétés

Voir à la fin du chapitre pour des illustrations de ce tableau (pages tirées du polycopié [Jac]).

Ces propriétés servent à identifier la nature des séries temporelles. Sous **R**, on utilisera les fonctions **acf**, **pacf** qui tracent, respectivement, les $\hat{\rho}(h)$ et les $\hat{r}(h)$ (les auto-corrélations empiriques et les auto-corrélations partielles empiriques). Le logiciel trace en plus un niveau bleu horizontal en $y = m_\alpha$ tel que pour tout h , $\mathbb{P}(|\hat{\rho}(h)| \geq m_\alpha | \rho(h) = 0) = \alpha$ (en général, le niveau α est fixé à 0,05) (la situation est la même pour les auto-corrélations partielles empiriques). On peut fixer α en ajoutant l'option : **acf(..., ci=alpha)**. Un $\hat{\rho}(h)$ sous la courbe bleue est donc non significatif (au niveau α). Le raisonnement est le suivant : on suppose $\rho(h) = 0$, si $|\hat{\rho}(h)| < m_\alpha$, on considère qu'il n'est pas nécessaire de revenir sur cette hypothèse de départ. La probabilité de rejeter à tort l'hypothèse « nulle » ($\rho(h) = 0$) est α .

Mais quand on veut utiliser le tableau ci-dessus, on cherche à savoir à partir de quel indice les $\rho(h)$ sont nuls (par exemple), pas si l'un d'eux est nul¹. Si on s'intéresse à la nullité de $(\rho(n), \rho(n+1), \dots, \rho(n+l-1))$, on pourrait vouloir trouver β tel que

$$\mathbb{P}(\exists h \in \{n, n+1, \dots, n+l-1\}, \hat{\rho}(h) \geq m_\beta | (\rho(n), \dots, \rho(n+l-1)) = (0, \dots, 0)) \leq \alpha.$$

Les $\hat{\rho}(h)$ sont supposés indépendants donc la probabilité ci-dessus est

$$\begin{aligned} 1 - \mathbb{P}(\forall h, |\hat{\rho}(h)| < m_\beta | \dots) &= 1 - (1 - \beta)^l \\ &= l\beta + o(\beta) \end{aligned}$$

(quand β est petit). Donc on prend, en général, $\beta = \alpha/l$.

Si on appelle les fonctions **acf** ou **pacf** sans préciser le **lag.max**, le logiciel le fixe par défaut à $10 \log_{10}(n)$ (où n est la longueur de la série). La raison est que l'on ne veut pas prendre le **lag.max** trop grand parce que les moyennes empiriques ne convergent pas bien pour h grand. On conseille en général de prendre $n \geq 50$ et $h \leq n/4$ (voir [BD02], p. 60 et [BJR08], p. 32).

1. Voir <https://xkcd.com/882/> sur le problème des tests multiples.

4.6. Estimation, choix de modèle et prévisions

La marche à suivre est la suivante.

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

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

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

D'où

$$\begin{aligned} a_1 &= \frac{\sigma(1)}{\sigma(0)} \times \frac{\sigma(0)^2 - \sigma(0)\sigma(1)}{\sigma(0)^2 - \sigma(1)^2}, \\ a_2 &= \frac{\sigma(0)\sigma(2) - \sigma(1)^2}{\sigma(0)^2 - \sigma(1)^2}. \end{aligned}$$

On peut donc estimer a_1, a_2 en remplaçant $\sigma(0), \sigma(1), \sigma(2)$ par leurs estimateurs empiriques dans les formules ci-dessus. Puisque $\sigma(0) = \sigma^2 + a_1 \sigma(1) + a_2 \sigma(2)$, on peut aussi estimer σ^2 .

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

4.6.2. Choix du modèle. On choisit entre plusieurs modèles en regardant :

- l'ajustement à la série de données,
- la complexité du modèle (il est plus facile d'estimer un nombre réduit de paramètres).

Pour concilier ces deux critères, on minimise une des deux quantités suivantes

$$\begin{aligned} AIC &= -2 \log(L(\Theta)) + 2\nu, \\ BIC &= -2 \log(L(\Theta)) + n\nu, \end{aligned}$$

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

EXEMPLE 4.17. Le calcul des quantités AIC et BIC se fait facilement en R. Calculons une vraisemblance dans un cas simple.

On suppose $X_t = at + b + \epsilon_t$ avec des ϵ_t i.i.d. de loi $\mathcal{N}(0; 1)$. Les paramètres du modèle sont a et b . On dispose d'observations x_1, \dots, x_n . La densité de (X_1, \dots, X_n) est la fonction

$$(u_1, \dots, u_n) \mapsto \prod_{t=1}^n \left\{ \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} (u_t - at - b)^2 \right) \right\}.$$

La vraisemblance est donc

$$(a', b') \mapsto L(a', b') = L((a', b'); (x_1, \dots, x_n)) = \prod_{t=1}^n \left\{ \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} (x_t - a't - b')^2 \right) \right\}.$$

Si on veut estimer (a, b) à partir de (x_1, \dots, x_n) , l'estimateur du maximum de vraisemblance est

$$(\hat{a}, \hat{b}) = \arg \max_{a', b'} L(a', b').$$

4.6.3. Prévision. Si le modèle est un processus $ARMA$, la prédiction pour X_{n+h} , sachant X_1, \dots, X_n est

$$\hat{X}_{n,h} = c_1 X_1 + \dots + c_n X_n,$$

où les coefficients sont choisis de manière à minimiser l'erreur quadratique

$$\mathbb{E}((X_{n+h} - c_1 X_1 - \dots - c_n X_n)^2).$$

PROPOSITION 4.18. *Ce choix de $\hat{X}_{n,h}$ entraîne l'égalité*

$$\hat{X}_{n,h} = \mathbb{E}(X_{n+h} | X_1, \dots, X_n).$$

Ce qui n'est pas surprenant si on se rappelle que l'espérance conditionnelle de X_{n+h} sachant X_1, \dots, X_n est la projection orthogonale de X_{n+h} sur $\sigma(X_1, \dots, X_n)$ dans l'espace des variables L^2 (muni de la norme L^2).

PROPOSITION 4.19. *L'erreur de prévision à l'horizon 1 ($X_{n+1} - \hat{X}_{n,1}$) est le bruit d'innovation ϵ_{n+1} .*

La variance de l'erreur de prévision ($\mathbb{E}((X_{n+h} - \hat{X}_{n,h})^2)$) est croissante avec h et tend vers $\mathbb{V}(X_1)$ quand $h \rightarrow +\infty$ (on rappelle que le processus (X_t) est supposé stationnaire).

Puisque les ϵ_t sont gaussiens, les X_t , $\hat{X}_{n,h}$ et $\hat{X}_{n,h} - X_{n+h}$ sont aussi gaussiens (nous sautons une petite démonstration). Ceci permet de construire facilement des intervalles de confiance (c'est inclus dans \mathbb{R}). Nous faisons ici un rappel dans un cas simple. Supposons $\hat{X}_{n,h} - X_{n+h} \sim \mathcal{N}(0; \sigma^2)$. Soit $\alpha = 0,01$. Nous cherchons Δ tel que

$$\mathbb{P}(X_{n+h} \in [\hat{X}_{n,h} - \Delta; \hat{X}_{n,h} + \Delta]) \geq 1 - \alpha$$

(ici, toutes les probabilités sont conditionnelles à X_1, \dots, X_n). Nous calculons

$$\begin{aligned} \mathbb{P}(X_{n+h} \in [\hat{X}_{n,h} - \Delta; \hat{X}_{n,h} + \Delta]) &= \mathbb{P}(|X_{n+h} - \hat{X}_{n,h}| \leq \Delta) \\ &= \mathbb{P}\left(\frac{|X_{n+h} - \hat{X}_{n,h}|}{\sigma} \leq \frac{\Delta}{\sigma}\right) \\ (\text{en utilisant les symétries de la gaussienne}) &= 1 - 2\mathbb{P}\left(\frac{X_{n+h} - \hat{X}_{n,h}}{\sigma} \geq \frac{\Delta}{\sigma}\right). \end{aligned}$$

Nous voulons donc Δ tel que

$$\mathbb{P}\left(\frac{X_{n+h} - \hat{X}_{n,h}}{\sigma} \geq \frac{\Delta}{\sigma}\right) \leq \frac{\alpha}{2} = 0,005.$$

Puisque $(X_{n,h} - \hat{X}_{n,h})/\sigma$ est de loi $\mathcal{N}(0; 1)$, on lit dans une table de la loi normale qu'il suffit de prendre $\Delta/\sigma = 2,58$ pour que l'inégalité ci-dessus soit vérifiée.

4.7. Processus non stationnaires : ARIMA et SARIMA

On veut revenir à la série d'origine une fois que l'on a étudié la partie stationnaire

EXEMPLE 4.20. Nous disposons d'une série temporelle (x_1, \dots, x_n) qui a une saisonnalité de période 12. On étudie $y_t = x_t - x_{t-12}$ (pour supprimer la saisonnalité). L'ajustement d'un modèle *ARMA* et les prévisions sont réalisées sur la série (y_t) . On écrit ensuite les (x_t) en fonction des (y_t) :

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

avec $r = t$ modulo 12 (le reste de la division euclidienne de t par 12).

On connaît les x_1, \dots, x_n (n supposé plus grand que 12), et donc aussi y_1, \dots, y_n . On peut calculer les prévisions $\hat{y}_{n,h}$ ($h \geq 1$). On en déduit les prévisions $\hat{x}_{n,h}$. Par exemple, si $h \in \{1, 2, \dots, 11\}$, nous utilisons l'égalité ci-dessus avec $t = n + h$ pour calculer la prévision

$$\hat{x}_{n,h} = \hat{y}_{n,h} + y_{t-12} + \dots + y_{r+12} + x_r.$$

Les processus. Ce sont des généralisation des processus *ARMA* aux cas non stationnaires, avec tendance polynômiale (*ARIMA*) ou avec une saisonnalité (*SARIMA*). Ce sont les processus directement utilisés par \mathbf{R} .

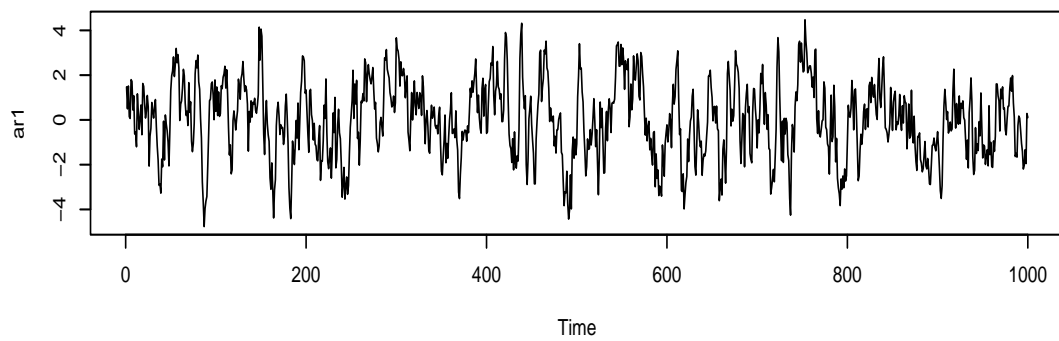
DÉFINITION 4.21. Le processus $(X_t)_{t \geq 0}$ est un processus *ARIMA*(p, d, q) si le processus $Y_t = \Delta_1^d X_t$ est un processus *ARMA*(p, q).

Les processus *ARIMA*(p, d, q) sont donc bien adaptés à l'étude des séries temporelles présentant une tendance polynômiale de degré $d - 1$.

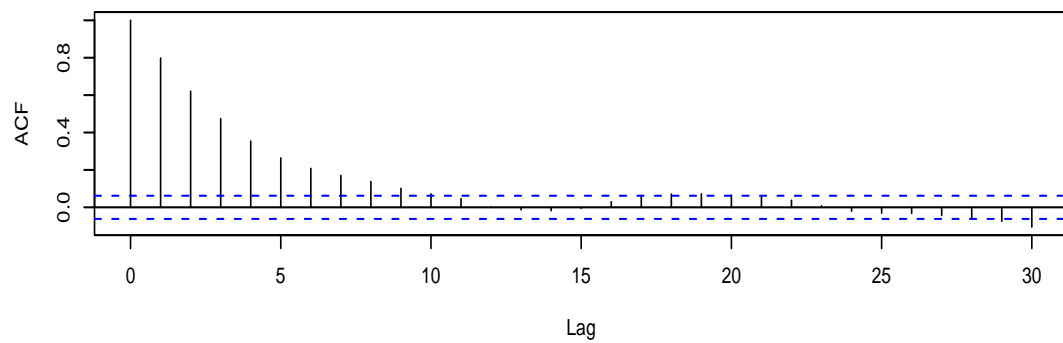
DÉFINITION 4.22. Le processus $(X_t)_{t \geq 0}$ est un processus *SARIMA*(p, d, q, T) si le processus $Y_t = \Delta_T \circ \Delta_1^d X_t$ est un processus *ARMA*(p, q).

Les processus *SARIMA*(p, d, q, T) sont donc bien adaptés à l'étude des séries temporelles qui présentent une saisonnalité de période T et qui ont une tendance polynômiale de degré $d - 1$.

REMARQUE 4.23. Attention, il existe dans la littérature des définitions de processus *SARIMA* plus complexes.



Series ar1



Series ar1

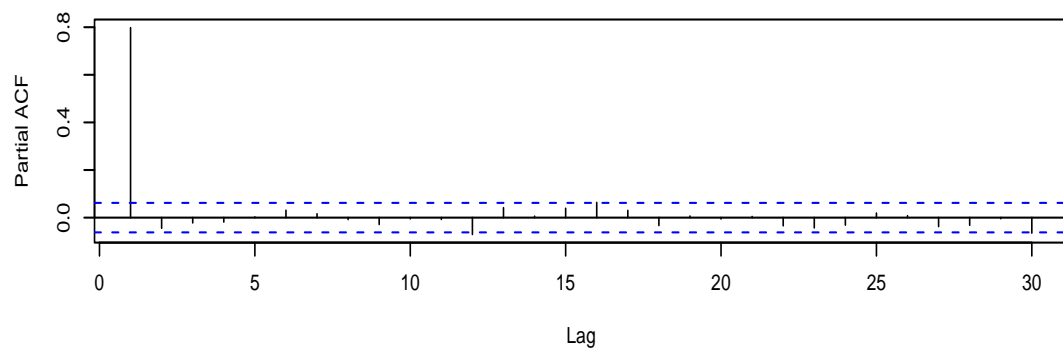
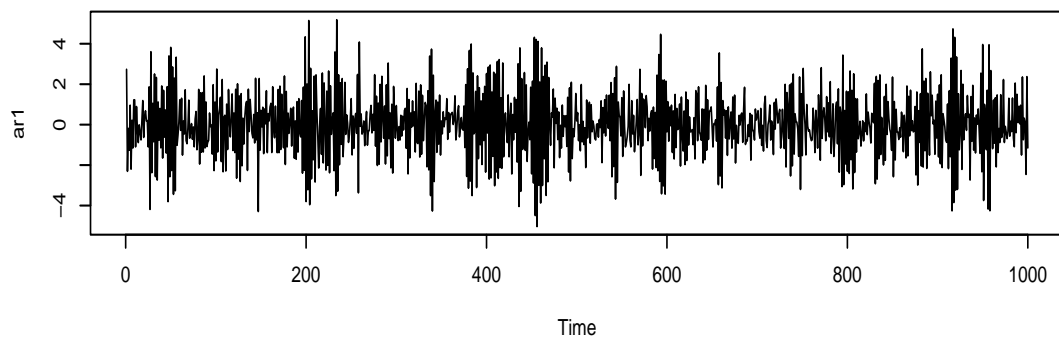
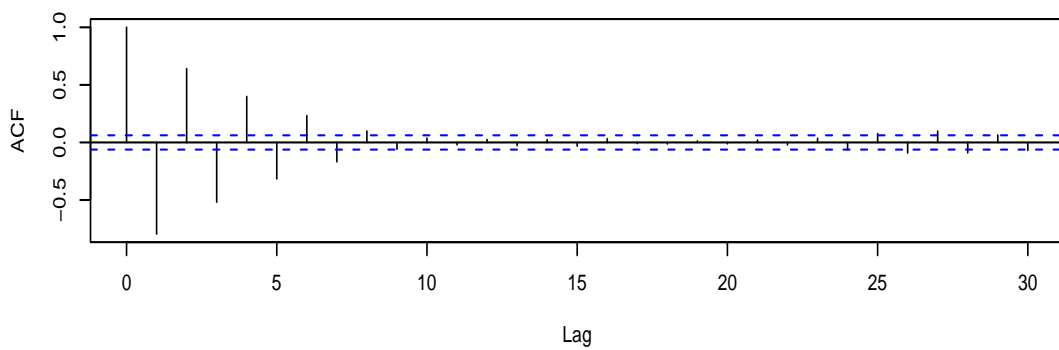


FIG. 17 – Simulation d'un $AR1$: $X_t = 0.8X_{t-1} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar1



Series ar1

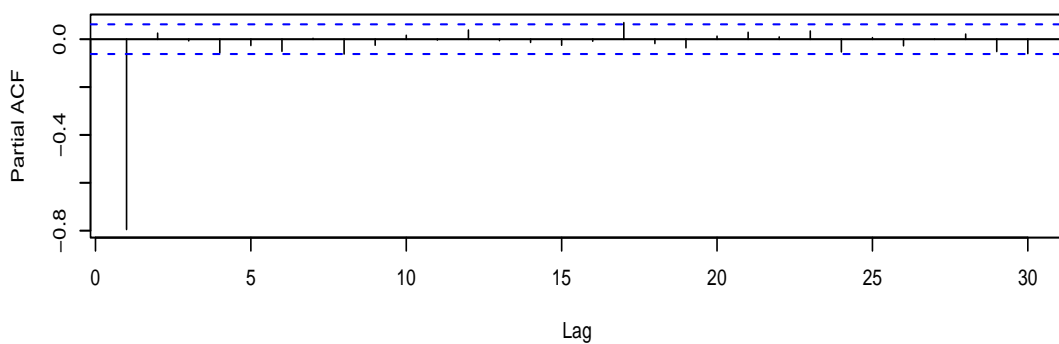
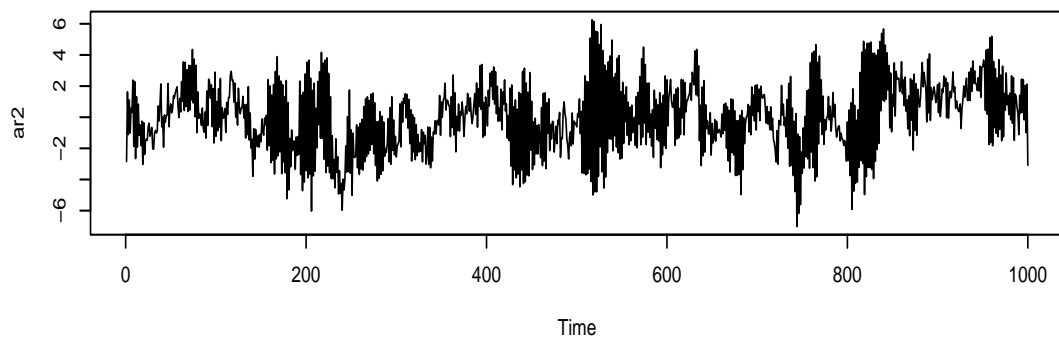
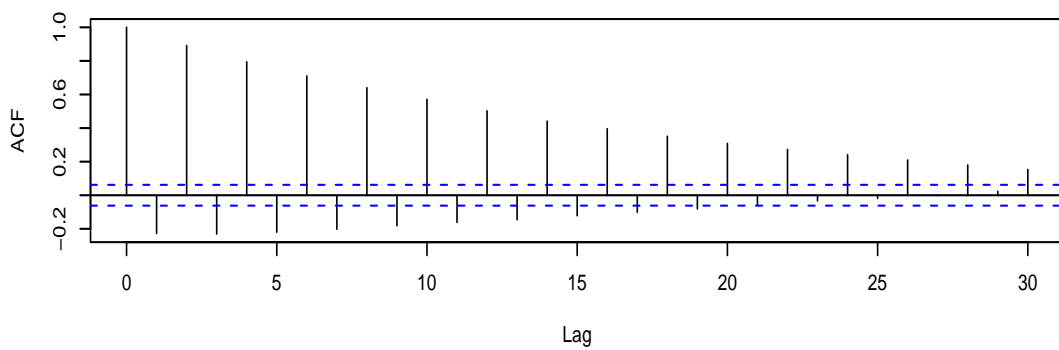


FIG. 18 – Simulation d'un $AR1$: $X_t = -0.8X_{t-1} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar2



Series ar2

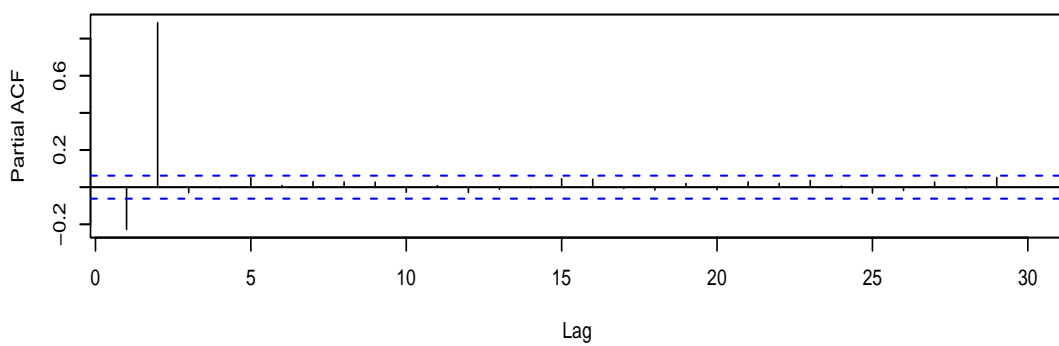
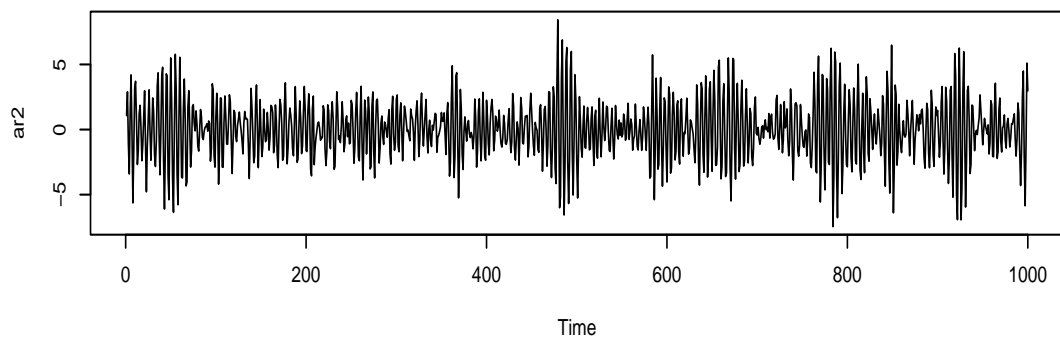
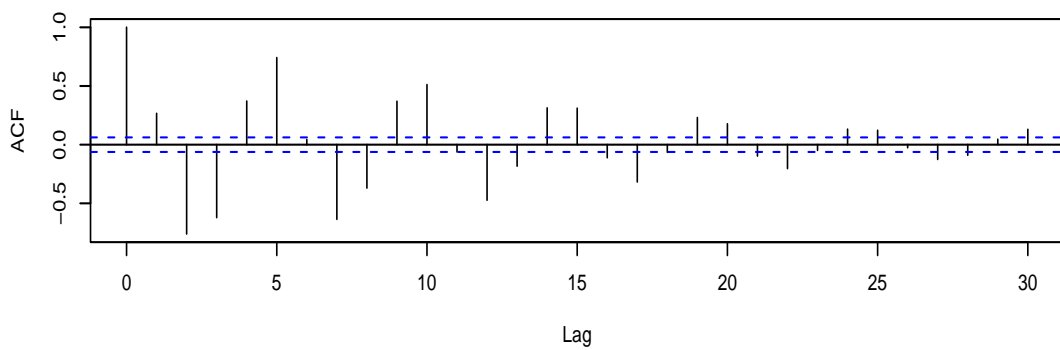


FIG. 19 – Simulation d'un AR_2 : $X_t = 0.9X_{t-2} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ar2



Series ar2

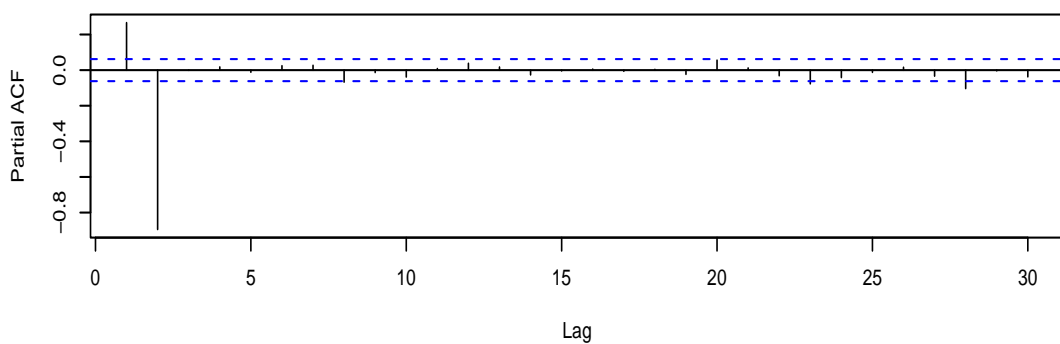
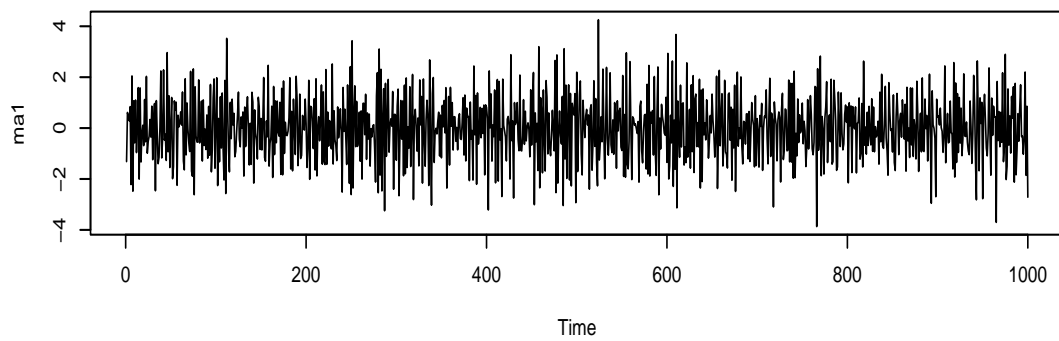
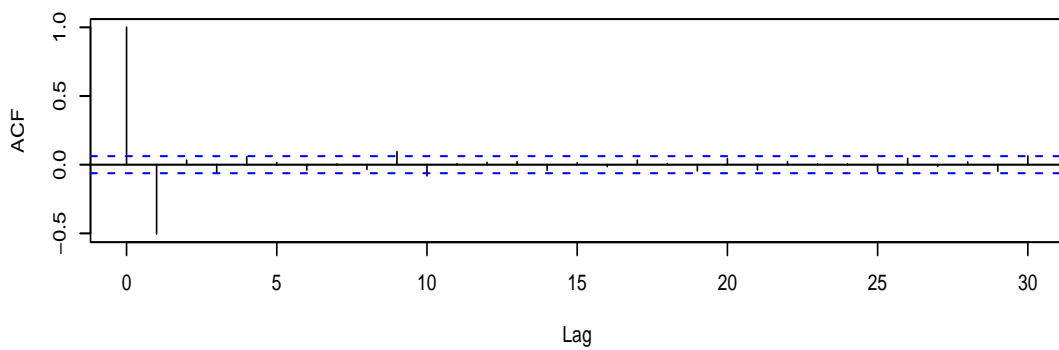


FIG. 20 – Simulation d'un AR_2 : $X_t = -0.5X_{t-1} - 0.9X_{t-2} + \epsilon_t$, auto-corrélation et auto-corrélation partielle.



Series ma1



Series ma1

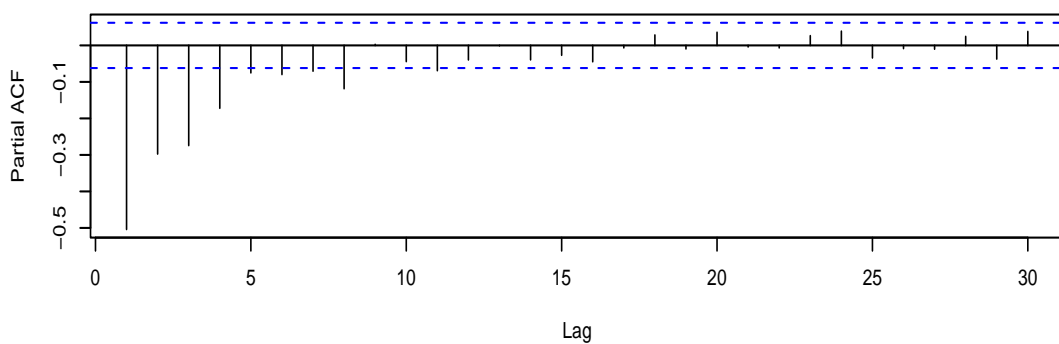
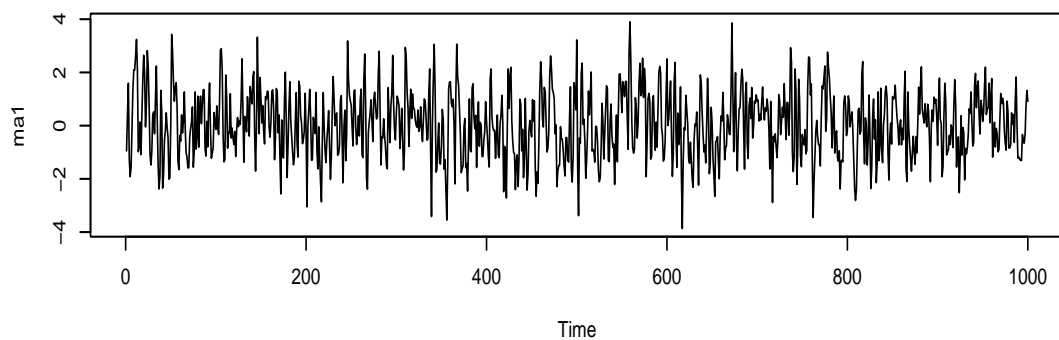
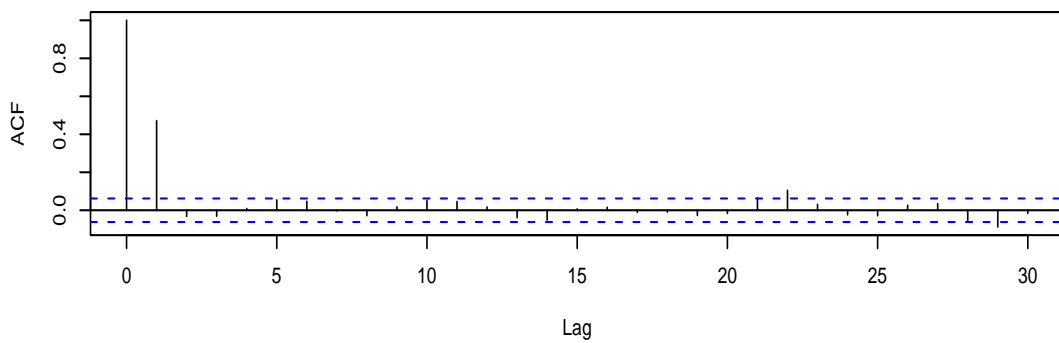


FIG. 21 – Simulation d'un MA_1 : $X_t = \epsilon_t - 0.8\epsilon_{t-1}$, auto-corrélation et auto-corrélation partielle.



Series ma1



Series ma1

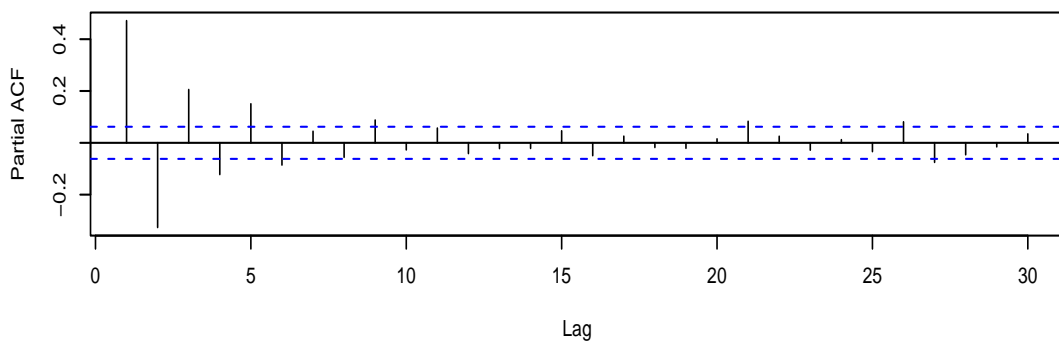
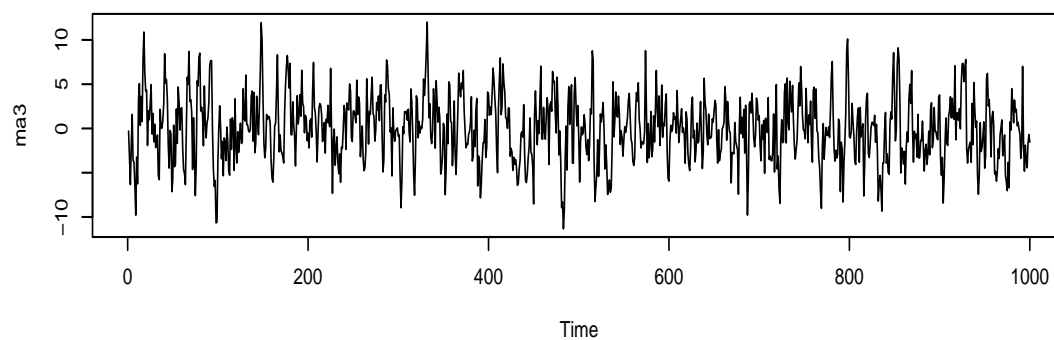
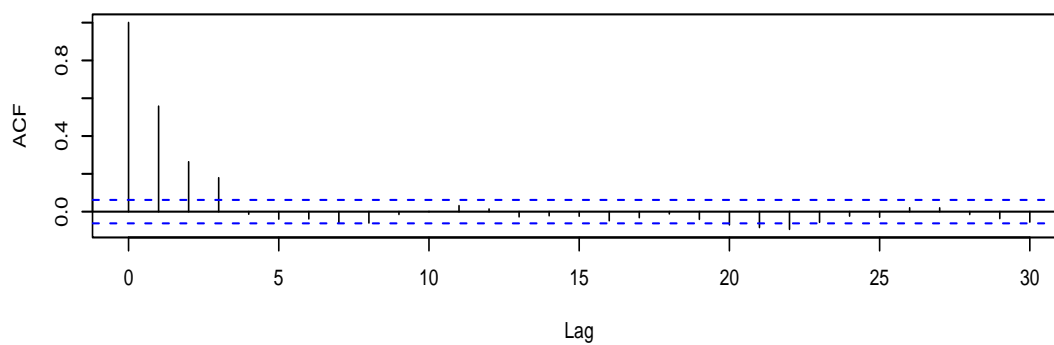


FIG. 22 – Simulation d'un MA_1 : $X_t = \epsilon_t + 0.8\epsilon_{t-1}$, auto-corrélation et auto-corrélation partielle.



Series ma3



Series ma3

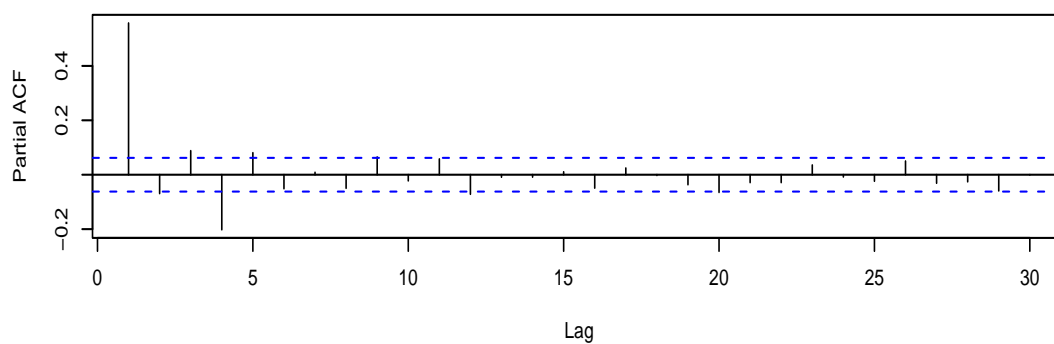


FIG. 23 – Simulation d'un MA_3 , auto-corrélation et auto-corrélation partielle.

4.8. Feuille d'exercices numéro 4 (durée : 6h)

- (1) Soit $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, $n \geq 2$. On cherche

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b) \in \mathbb{R}^2} F(a,b), \text{ avec } F(a,b) := \sum_{t=1}^n (x_t - a - bt)^2.$$

- (a) Trouver les points critiques de F .

Formules à savoir par cœur :

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

- (b) Montrer que l'unique point critique est un minimum absolu de F .
- (2) Soit $\mathbb{R}[X]$ l'espace des polynômes à coefficients réels. Soit $T \in \mathbb{R}$. Soit $\Delta : P \in \mathbb{R}[X] \mapsto \Delta P(X) = P(X) - P(X - T)$.
- (a) On suppose que $\deg(P) \geq 1$. Montrer que $\deg(\Delta P) \leq \deg(P) - 1$. Montrer que si $\deg(P) = 0$, alors $\Delta P = 0$.
- (b) On suppose que $\Delta^k P \neq 0$. Montrer que $\deg(P) \geq k$.
- (3) Les processus (X_k) suivants sont stationnaires et centrés. Le processus $(\epsilon_n)_{n \geq 0}$ est un bruit blanc de variance α^2 et de moyenne nulle.

- (a) Quelle est la fonction d'auto-covariance du processus auto-régressif de représentation

$$X_n - \frac{1}{2}X_{n-1} = \epsilon_n ?$$

(Commencer par vérifier qu'il existe un processus stationnaire vérifiant cette équation.)

- (b) On considère le processus auto-régressif de représentation

$$X_n - \frac{1}{6}X_{n-1} - \frac{1}{6}X_{n-2} = \epsilon_n.$$

Commencer par vérifier la stationnarité du processus. Donner les valeurs de $\sigma(0)$ et $\sigma(1)$. Trouver les racines du polynôme $1 - \frac{1}{6}z - \frac{1}{6}z^2$ et en déduire l'expression de $\sigma(h)$ pour $h \geq 1$.

- (c) On considère le processus auto-régressif de représentation

$$X_n - X_{n-1} + \frac{1}{2}X_{n-2} = \epsilon_n.$$

Commencer par vérifier la stationnarité du processus. Donner les expressions de $\sigma(0)$ et $\sigma(1)$. Trouver les racines du polynôme $1 - z + \frac{1}{2}z^2$ et en déduire l'expression de $\sigma(h)$ pour $h \geq 1$.

- (4) Soit le processus AR_1 (centré) suivant

$$(4.8) \quad X_t = aX_{t-1} + \epsilon_t$$

où ϵ_t est un bruit blanc centré de variance σ^2 . On suppose $a \neq 0$.

- (a) Quelle condition doit-on imposer sur a pour qu'il existe un processus stationnaire vérifiant (4.8) ? On suppose dans la suite que cette condition est vérifiée.
- (b) Calculer la variance de ce processus.
- (c) Montrer que l'auto-covariance d'un tel processus est

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

- (d) En déduire la convergence vers 0 de l'auto-covariance lorsque $h \rightarrow +\infty$.
- (e) Calculer les auto-corrélations.

- (f) On suppose que $\mathbb{E}(X_0|X_1) = X_1/a$ (il n'y a pas de raison que ce soit toujours vrai). Calculer les auto-corrélations partielles.
- (5) Considérons le processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant $X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$ (avec (ϵ_t) bruit blanc centré de variance σ^2).
- (a) Montrer que $\sigma(0) = \sigma^2 \times \frac{1+b^2+2ab}{1-a^2}$, $\sigma(1) = \sigma^2 \times \frac{a+b+ab^2+a^2b}{1-a^2}$.
- (b) Montrer que $\rho(1) = \frac{(a+b)(1+ab)}{b^2+2ab+1}$ et $\rho(h) = a^{h-1}\rho(1)$ pour $h \geq 1$.

4.9. Corrigé de la feuille d'exercices numéro 4

- (1) (a) Nous avons

$$\begin{aligned}\frac{\partial F}{\partial a}(a, b) &= \sum_{t=1}^n -2(x_t - a - bt) \\ \frac{\partial F}{\partial b}(a, b) &= \sum_{t=1}^n -2t(x_t - a - bt).\end{aligned}$$

Notons

$$\bar{x}_n = \sum_{t=1}^n x_t, \quad \bar{s}_n = \sum_{t=1}^n tx_t.$$

Nous cherchons à résoudre le système

$$\begin{cases} -\bar{x}_n + an + b\frac{n(n+1)}{2} &= 0 \\ -\bar{s}_n + a\frac{n(n+1)}{2} + b\frac{n(n+1)(2n+1)}{6} &= 0. \end{cases}$$

Nous calculons la valeur de l'unique solution :

$$\begin{aligned}(-\bar{x}_n + an) \times \frac{(2n+1)}{3} - \left(-\bar{s}_n + a\frac{n(n+1)}{2}\right) &= 0 \\ an \left(\frac{(2n+1)}{3} - \frac{(n+1)}{2}\right) &= \bar{x}_n \times \frac{(2n+1)}{3} - \bar{s}_n \\ a &= \frac{2(2n+1)\bar{x}_n - 6\bar{s}_n}{n(n-1)},\end{aligned}$$

$$\begin{aligned}\left(-\bar{x}_n + b\frac{n(n+1)}{2}\right) \times \frac{(n+1)}{2} - \left(-\bar{s}_n + b\frac{n(n+1)(2n+1)}{6}\right) &= 0 \\ b\frac{n(n+1)}{2} \left(\frac{(n+1)}{2} - \frac{2n+1}{3}\right) &= \frac{(n+1)}{2}\bar{x}_n - \bar{s}_n \\ b &= -\frac{6(n+1)\bar{x}_n - 12\bar{s}_n}{n(n+1)(n-1)}.\end{aligned}$$

- (b) Calculons les dérivées secondes de F (pour voir une autre méthode que celle vue en cours, vous choisissez celle que vous préférez) :

$$\begin{aligned}\frac{\partial^2 F}{\partial a^2}(a, b) &= 2n \\ \frac{\partial^2 F}{\partial b^2}(a, b) &= \frac{n(n+1)(2n+1)}{3} \\ \frac{\partial^2 F}{\partial a \partial b}(a, b) &= n(n+1).\end{aligned}$$

La matrice hessienne de F en (a, b) est donc

$$F''(a, b) = \begin{bmatrix} 2n & n(n+1) \\ n(n+1) & \frac{n(n+1)(2n+1)}{3} \end{bmatrix}.$$

La forme quadratique associée est :

$$\begin{aligned}(x, y) &\mapsto 2nx^2 + \frac{n(n+1)(2n+1)}{3}y^2 + 2n(n+1)xy \\ &= \frac{n(n+1)(2n+1)}{3} \left(y - \frac{3x}{(2n+1)} \right)^2 - \frac{3n(n+1)}{(2n+1)}x^2 + 2nx^2 \\ &= \frac{n(n+1)(2n+1)}{3} \left(y - \frac{3x}{(2n+1)} \right)^2 + x^2 \left(\frac{n^2-n}{2n+1} \right).\end{aligned}$$

On voit que $n^2 - n > 0$ pour $n \geq 2$ et que donc la forme quadratique est définie positive. Donc F est strictement convexe. Donc le point critique trouvé est un minimum absolu.

(2) (vu en cours)

(a) Soit n le degré de P , on suppose $n \geq 1$. Le polynôme P peut s'écrire $P(X) = a_n X^n + a_{n-1} X^{n-1} + \dots + a_0$ (avec a_0, \dots, a_n dans \mathbb{R}). Et donc

$$\begin{aligned}-\Delta P(X) &= (a_n(X-T)^n + \dots + a_0) - (a_n X^n + \dots + a_0) \\ &= a_n \sum_{i=0}^n [C_n^i X^i (-T)^{n-i}] - a_n X^n \\ &\quad + (a_{n-1}(X-T)^{n-1} + \dots + a_0) - (a_{n-1} X^{n-1} + \dots + a_0) \\ &= a_n \sum_{i=0}^{n-1} [C_n^i X^i (-T)^{n-i}] \\ &\quad + (a_{n-1}(X-T)^{n-1} + \dots + a_0) - (a_{n-1} X^{n-1} + \dots + a_0).\end{aligned}$$

Donc $\deg(\Delta P) \leq n-1 = \deg(P) - 1$. Dans le cas $n = 0$, le même calcul montre que $\Delta P = 0$.

(b) Soit n le degré de P . Par la question précédente, nous savons que si $\deg(P) \leq k-1$, alors $\Delta^k P = 0$, ce qui n'est pas le cas. Donc $\deg(P) \geq k$.

(3) (a) Le polynôme associé à l'équation est $A(X) = 1 - X/2$. Son unique racine est 2, qui est de module > 1 . Donc il existe un processus stationnaire qui vérifie l'équation voulue. Nous avons (pour tout n)

$$\begin{aligned}\mathbb{V} \left(\left(X_n - \frac{1}{2} X_{n-1} \right) \right) &= \alpha^2 \\ \sigma(0) + \frac{1}{4} \sigma(0) - \sigma(1) &= \alpha^2 \\ \frac{5}{4} \sigma(0) - \sigma(1) &= \alpha^2,\end{aligned}$$

et

$$\begin{aligned}\sigma(1) &= \text{Cov}(X_n, X_{n-1}) \\ &= \text{Cov} \left(\frac{1}{2} X_{n-1} + \epsilon_n, X_{n-1} \right) \\ &= \frac{\sigma(0)}{2} + 0.\end{aligned}$$

D'où :

$$\begin{aligned}\frac{5}{4} \sigma(0) - \frac{\sigma(0)}{2} &= \alpha^2 \\ \sigma(0) &= \frac{4}{3} \alpha^2\end{aligned}$$

et

$$\sigma(1) = \frac{2}{3} \alpha^2.$$

La formule de récurrence pour les $\sigma(\cdot)$ est

$$\sigma(h+1) = \frac{1}{2}\sigma(h).$$

Donc, pour tout h ,

$$\sigma(h) = \frac{2^{2-h}}{3}\alpha^2.$$

(b) Nous avons (pour tout n)

$$\begin{aligned} \mathbb{V}(X_n) &= \mathbb{V}\left(\frac{1}{6}X_{n-1} + \frac{1}{6}X_{n-2} + \epsilon_n\right) \\ \sigma(0) &= \frac{1}{36}\sigma(0) + \frac{1}{36}\sigma(0) + \alpha^2 + \frac{2}{36}\sigma(1) + 0 \\ 36\sigma(0) &= 2\sigma(0) + 2\sigma(1) + 36\alpha^2 \\ 34\sigma(0) &= 2\sigma(1) + 36\alpha^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(X_n, X_{n-1}) &= \text{Cov}\left(\frac{1}{6}X_{n-1} + \frac{1}{6}X_{n-2} + \epsilon_n, X_{n-1}\right) \\ \sigma(1) &= \frac{1}{6}\sigma(0) + \frac{1}{6}\sigma(1) + 0 \\ 5\sigma(1) &= \sigma(0). \end{aligned}$$

D'où

$$\begin{aligned} 170\sigma(1) - 2\sigma(1) &= 36\alpha^2 \\ 168\sigma(1) &= 36\alpha^2 \\ \sigma(1) &= \frac{36}{168}\alpha^2 = \frac{3\alpha^2}{14} \end{aligned}$$

et

$$\sigma(0) = \frac{15}{14}\alpha^2.$$

Nous remarquons que 2 est racine de P , ce qui permet de factoriser :

$$P(z) = 1 - \frac{1}{6}z - \frac{1}{6}z^2 = \left(1 - \frac{z}{2}\right)\left(1 + \frac{z}{3}\right).$$

Donc les racines du polynôme sont $\{2; -3\}$. La relation de récurrence sur les autocovariances est la suivante :

$$\sigma(h+2) = \frac{\sigma(h+1) + \sigma(h)}{6},$$

de polynôme caractéristique $Q(z) = z^2 - z/6 - 1/6$ dont les racines sont inverses de celles de P . Les racines de Q sont donc $\{1/2; -1/3\}$. Donc $\sigma(h)$ s'écrit sous la forme

$$\sigma(h) = \frac{a}{2^h} + \frac{b}{(-3)^h}.$$

Nous déterminons a et b à partir de $\sigma(0)$ et $\sigma(1)$:

$$\begin{cases} a + b &= \sigma(0) \\ \frac{a}{2} - \frac{b}{3} &= \sigma(1). \end{cases}$$

D'où

$$\begin{aligned} a\left(\frac{1}{3} + \frac{1}{2}\right) &= \frac{\sigma(0)}{3} + \sigma(1) \\ a \times \frac{5}{6} &= \left(\frac{5}{14} + \frac{3}{14}\right)\alpha^2 \\ a &= \frac{24}{35}\alpha^2, \end{aligned}$$

et

$$\begin{aligned} b &= \sigma(0) - a \\ &= \left(\frac{15}{14} - \frac{24}{35} \right) \alpha^2 \\ &= \left(\frac{75 - 48}{70} \right) \alpha^2 = \frac{27}{70} \alpha^2. \end{aligned}$$

(c) Nous avons

$$\begin{aligned} \mathbb{V}(X_n) &= \mathbb{V}\left(X_{n-1} - \frac{1}{2}X_{n-2} + \epsilon_n\right) \\ \sigma(0) &= \sigma(0) + \frac{\sigma(0)}{4} - \sigma(1) + \alpha^2 \\ -\frac{\sigma(0)}{4} + \sigma(1) &= \alpha^2, \end{aligned}$$

et

$$\begin{aligned} \text{Cov}(X_n, X_{n-1}) &= \text{Cov}\left(X_{n-1} - \frac{1}{2}X_{n-2} + \epsilon_n, X_{n-1}\right) \\ \sigma(1) &= \sigma(0) - \frac{\sigma(1)}{2} \\ \frac{3\sigma(1)}{2} &= \sigma(0). \end{aligned}$$

D'où

$$\begin{aligned} -\frac{3\sigma(1)}{8} + \sigma(1) &= \alpha^2 \\ \sigma(1) &= \frac{8}{5}\alpha^2, \end{aligned}$$

et

$$\sigma(0) = \frac{12}{5}\alpha^2.$$

Soit le polynôme $P(z) = 1 - z + z^2/2$. Son discriminant est :

$$\Delta = 1 - \frac{4}{2} = -1.$$

Les racines de P sont donc

$$\{1 + i; 1 - i\}.$$

Le polynôme caractéristique de la relation de récurrence des $\sigma(h)$ est

$$Q(z) = z^2 - z + \frac{1}{2},$$

dont les racines sont les inverses de celles de P et sont donc

$$\left\{ \frac{1}{1+i}; \frac{1}{1-i} \right\} = \left\{ \frac{1-i}{2}; \frac{1+i}{2} \right\}$$

(nous aurions pu les calculer directement, sans passer par P). Les $\sigma(h)$ sont donc de la forme

$$\sigma(h) = a2^{-h}(1-i)^h + b2^{-h}(1+i)^h.$$

Nous déterminons a et b à partir de $\sigma(0)$ et $\sigma(1)$:

$$\begin{cases} a + b &= \sigma(0) \\ \frac{a}{2}(1-i) + \frac{b}{2}(1+i) &= \sigma(1). \end{cases}$$

D'où

$$\frac{a}{2}(1+i - (1-i)) = \frac{\sigma(0)(1+i)}{2} - \sigma(1)$$

$$\begin{aligned}
ia &= \frac{\sigma(0)(1+i)}{2} - \sigma(1) \\
a &= -i \times \left(\frac{6}{5}\alpha^2 + \frac{6i}{5}\alpha^2 - \frac{8}{5}\alpha^2 \right) \\
a &= \frac{6}{5}\alpha^2 + \frac{2i}{5}\alpha^2,
\end{aligned}$$

et

$$\begin{aligned}
b &= \sigma(0) - a \\
&= \frac{12}{5}\alpha^2 - \left(\frac{6}{5}\alpha^2 + \frac{2i}{5}\alpha^2 \right) \\
&= \frac{6}{5}\alpha^2 - \frac{2i}{5}\alpha^2.
\end{aligned}$$

Nous avons donc pour tout h (avec \mathcal{R} désignant la partie réelle)

$$\begin{aligned}
\sigma(h) &= \frac{2\alpha^2}{5} ((3+i)2^{-h}(1-i)^h + (3-i)2^{-h}(1+i)^h) \\
&= \frac{4\alpha^2}{5} \mathcal{R}((3+i)2^{-h}(1-i)^h) \\
&= \frac{4\alpha^2}{5} \mathcal{R}((3+i)2^{-h}(\sqrt{2}e^{-i\pi/4})^h) \\
&= \frac{4\alpha^2}{5} \mathcal{R}((3+i)2^{-h/2}(\cos(-h\pi/4) + i\sin(-h\pi/4))) \\
&= \frac{4\alpha^2}{5} (3 \cdot 2^{-h/2} \cos(h\pi/4) + 2^{-h/2} \sin(h\pi/4)).
\end{aligned}$$

- (4) (a) Si $a \neq 0$, le polynôme associé à ce processus est $P(z) = 1 - az$, de racine $1/a$. Dans ce cas, on sait qu'il existe une suite stationnaire vérifiant la relation voulue si $|1/a| > 1$, c'est à dire $|a| < 1$.

- (b) Nous avons (pour tout t)

$$\begin{aligned}
\mathbb{V}(X_t) &= \mathbb{V}(aX_{t-1} + \epsilon_t) \\
\sigma(0) &= a^2\sigma(0) + \sigma^2 \\
\sigma(0) &= \frac{\sigma^2}{1 - a^2}.
\end{aligned}$$

- (c) La relation de récurrence vérifiée par les auto-covariances est la suivante :

$$\sigma(h+1) = a\sigma(h).$$

Nous en déduisons que pour tout h :

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

- (d) Donc $\sigma(h) \xrightarrow{h \rightarrow +\infty} 0$.

- (e) Pour $h \geq 1$:

$$\rho(h) = \frac{\sigma(h)}{\sigma(0)} = a^h.$$

- (f) Pour tout t , nous pouvons écrire

$$\begin{aligned}
X_t &= aX_{t-1} + \epsilon_t \\
&= a^2X_{t-2} + a\epsilon_{t-1} + \epsilon_t \\
&= a^3X_{t-3} + a^2\epsilon_{t-2} + a\epsilon_{t-1} + \epsilon_t \\
&\vdots
\end{aligned}$$

$$= a^t X_0 + \sum_{i=1}^t a^{t-i} \epsilon_i.$$

Fixons maintenant n . La donnée de X_1, \dots, X_{n-1} est équivalente à la donnée de $aX_0 + \epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_{n-1}$ (les variables de la deuxième liste se déduisent de celle de la deuxième liste, et vice-versa). Nous avons donc

$$\begin{aligned} \mathbb{E}(X_n | X_1, \dots, X_{n-1}) &= \mathbb{E}(a^n X_0 + a^{n-1} \epsilon_1 + \sum_{i=2}^{n-1} a^{n-i} \epsilon_i + \epsilon_n | aX_0 + \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}) \\ &= a^n X_0 + a^{n-1} \epsilon_1 + \sum_{i=2}^{n-1} a^{n-i} \epsilon_i + \mathbb{E}(\epsilon_n | aX_0 + \epsilon_1, \epsilon_2, \dots, \epsilon_{n-1}) \\ &= a^n X_0 + a^{n-1} \epsilon_1 + \sum_{i=2}^{n-1} a^{n-i} \epsilon_i + \mathbb{E}(\epsilon_n) \\ &= a^{n-1} X_1 + \sum_{i=2}^{n-1} a^{n-i} \epsilon_i, \end{aligned}$$

et

$$\begin{aligned} \mathbb{E}(X_0 | X_1, \dots, X_{n-1}) &= \mathbb{E}(X_0 | X_1) \\ &= \frac{X_1}{a}. \end{aligned}$$

Donc

$$\begin{aligned} X_n - \mathbb{E}(X_n | X_1, \dots, X_{n-1}) &= \epsilon_n, \\ X_0 - \mathbb{E}(X_0 | X_1, \dots, X_{n-1}) &= X_0 - \frac{X_1}{a} = -\frac{\epsilon_1}{a}. \end{aligned}$$

Donc

$$\begin{aligned} \text{Cov}(X_0 - \mathbb{E}(X_0 | X_1, \dots, X_{n-1}), X_n - \mathbb{E}(X_n | X_1, \dots, X_{n-1})) &= \text{Cov}\left(-\frac{\epsilon_1}{a}, \epsilon_n\right) \\ &= 0. \end{aligned}$$

(5) (a) Nous avons (pour tout t)

$$\begin{aligned} \mathbb{V}(X_t) &= \mathbb{V}(aX_{t-1} + \epsilon_t + b\epsilon_{t-1}) \\ \sigma(0) &= a^2\sigma(0) + \sigma^2 + b^2\sigma^2 + 2ab \text{Cov}(X_{t-1}, \epsilon_{t-1}) \\ \sigma(0) &= a^2\sigma(0) + \sigma^2 + b^2\sigma^2 + 2ab \text{Cov}(aX_{t-2} + \epsilon_{t-1} + \epsilon_{t-2}, \epsilon_{t-1}) \\ \sigma(0) &= a^2\sigma(0) + (1 + b^2)\sigma^2 + 2ab\sigma^2 \\ \sigma(0) &= \frac{(1 + b^2 + 2ab)\sigma^2}{1 - a^2}. \end{aligned}$$

Et aussi :

$$\begin{aligned} \text{Cov}(X_t, X_{t-1}) &= \text{Cov}(aX_{t-1} + \epsilon_t + b\epsilon_{t-1}, X_{t-1}) \\ \sigma(1) &= a\sigma(0) + b \text{Cov}(\epsilon_{t-1}, X_{t-1}) \\ \sigma(1) &= a\sigma(0) + b\sigma^2 \\ \sigma(1) &= \frac{(a(1 + b^2 + 2ab) + b(1 - a^2))}{1 - a^2} \sigma^2 \\ \sigma(1) &= \frac{(a + b + ab^2 + a^2b)}{1 - a^2} \sigma^2. \end{aligned}$$

(b) Nous avons

$$\rho(1) = \frac{\sigma(1)}{\sigma(0)} = \frac{a + b + ab^2 + a^2b}{1 + b^2 + 2ab} = \frac{(a + b)(1 + ab)}{1 + b^2 + 2ab}.$$

La relation de récurrence pour les auto-covariances est (pour $h \geq 1$) :

$$\sigma(h+1) = a\sigma(h).$$

Donc, pour tout $h \geq 1$,

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

et donc

$$\rho(h) = a^{h-1} \frac{(a+b)(1+ab)}{1+b^2+2ab}.$$

4.10. Feuille d'exercices numéro 5 (durée : 6h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf. C'est ce qui sera demandé au partiel.

4.10.1. Simulation des processus ARMA. Penser à représenter chaque processus simulé.

- (1) Donner la définition d'un processus $ARMA_{p,q}$. Rappeler les conditions sur les coefficients pour que ce processus soit stationnaire.
- (2) À l'aide de la fonction `arma.sim`, simuler plusieurs processus AR_p et MA_q (p et q pas trop grands). Avant toute simulation, écrire la définition mathématique du processus à simuler et veillez à ce que les conditions de stationnarité soient respectées.
- (3) Observer les auto-corrélations empiriques (partielles ou non). Que constatez-vous ?
- (4) Simuler quelques $ARMA_{p,q}$, observer et interpréter les auto-corrélations empiriques (partielles ou non).
- (5) Faire de même avec un modèle $ARIMA_{p,d,q}$, avec un d assez petit.

4.10.2. Identification d'un processus ARMA. Récupérer le fichier de données `serie1.dat`².

- (1) Ce processus vous semble-t-il modélisable par un processus ARMA stationnaire ? Pourquoi ?
- (2) On travaille désormais avec la série obtenue en appliquant la fonction `diff` à la série. Quelle transformation a-t-on effectuée ?
- (3) En observant les auto-corrélations empiriques et auto-corrélations partielles empiriques, proposer des modèles AR_p ou MA_q d'ordre faible pour modéliser cette série.
- (4) Estimer les paramètres du modèle sélectionné.
- (5) Tester la blancheur des résidus (on choisira la niveau $\alpha = 0,02$).
- (6) Conclure pour choisir un modèle.
- (7) S'inspirer de la démarche pour modéliser la série `serie2.dat`.

4.10.3. Prévision dans un processus ARMA. Soit le processus $X_t - X_{t-1} + \frac{1}{2}X_{t-2} - \frac{1}{3}X_{t-3} = \epsilon_t$ avec ϵ_t un bruit blanc gaussien centré réduit.

- (1) Après avoir identifié ce processus, simuler 50 réalisations de longueur 105.
- (2) Pour chaque simulation, extraire les 100 premières valeurs et estimer les paramètres d'un AR_3 .
- (3) Pour chaque simulation, prédire les cinq valeurs suivantes.
- (4) Donner une estimation de l'erreur moyenne (biais) et de la variance de l'erreur de prévision à 1, 2, 3, 4 et 5 pas.
- (5) Recommencer en rallongeant la durée d'observation, et comparer aux résultats précédents.

2. Toujours à l'adresse : <http://math.unice.fr/~rubentha/enseignement/>

4.10.4. Précipitations mensuelles à San Fransisco entre 1932 et 1966. Récupérer la série `sanfran.dat`.

- (1) La série semble-t-elle stationnaire? Si non faites en sorte qu'elle le soit.
- (2) Proposer alors un AR_p ou un AM_q adapté. Valider votre modélisation en testant les résidus. Mettez votre idée de côté, car nous vous proposons de tester maintenant directement sur la série initiale un AR_2 avec partie saisonnière (autrement dit un $SARIMA_{2,0,0,12}$). On utilise désormais les données jusqu'à la fin de l'année 1963.
- (3) Estimer le modèle $SARIMA_{2,0,0,12}$. Afficher et tester les résidus.
- (4) Prévoir les précipitations de 1964, 1965 et 1966 à partir de ce modèle. Superposer sur un graphique prévision et valeurs réelles.
- (5) Faire de même avec le modèle AR_p sélectionné à la première question (ainsi qu'avec un lissage exponentiel de Holt-Winters, avec et sans composante saisonnière).
- (6) Quel est, graphiquement, le meilleur modèle pour prédire cette série?
- (7) Comment aurait-on pu répondre à la question précédente de façon moins subjective? Comparer les résultats à l'analyse graphique.

4.10.5. Taux d'intérêt au Royaume-Uni. Le fichier `UKinterestrates.dat` contient le spread des taux d'intérêts (différence entre taux d'intérêt à long terme et à court terme) pour le Royaume-Uni entre mars 1953 et décembre 1995. En vous inspirant des exercices précédents, proposer une modélisation de type $ARMA$, $ARIMA$ ou $SARIMA$. Justifier bien votre démarche.

Appendice : mise en œuvre sous R. Auto-corrélation partielle : utiliser la fonction `pacf`. La fonction `arma.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))` (choisir `ma` ou `ar` pour un processus MA ou AR , si $p = 1$, écrire `ar=a1`). 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 `arma.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 `arma` permet d'estimer les paramètres :

- d'un $ARMA_{p,q}$: `out<-arma(serie,order=c(p,0,q))`
- d'un $ARIMA_{p,d,q}$: `out<-arma(serie,order=c(p,d,q))`
- d'un $SARIMA_{p,d,q,T}$:
`out<-arma(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.

Analyse spectrale

On cherche à décrire une série stationnaire comme un mélange de fonctions périodiques (plus du bruit).

EXEMPLE 5.1.

$$Y_t = \mu + \sum_{j=1}^p [\alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t)] + \epsilon_t$$

Ici les ω_j sont des fréquences. La période de $t \mapsto \cos(\omega_j t)$ est $2\pi/\omega_j$, pour tout j .

5.1. Densité spectrale

On suppose que $(X_t)_{t \geq 0}$ est un processus stationnaire centré de covariances $(\sigma(h))_{h \geq 0}$, avec

$$(5.1) \quad \sum_{h=-\infty}^{+\infty} |\sigma(h)| < +\infty.$$

DÉFINITION 5.2. La densité spectrale de (X_t) est la fonction

$$\lambda \in \mathbb{R} \mapsto f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\lambda} \sigma(h)$$

(rappel : $e^{i\lambda} = \cos(\lambda) + i \sin(\lambda)$, $i^2 = -1$).

La fonction f est bien définie à cause de l'équation (5.1).

PROPOSITION 5.3. La fonction f vérifie :

- (1) f est paire,
- (2) f est 2π -périodique,
- (3) pour tout $\lambda \in \mathbb{R}$, $f(\lambda) \in \mathbb{R}^+$,
- (4) pour tout k dans \mathbb{Z} , $\sigma(k) = \int_{-\pi}^{\pi} e^{i\lambda k} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(\lambda k) f(\lambda) d\lambda$.

DÉMONSTRATION. (1) Pour tout λ ,

$$\begin{aligned} f(-\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{i\lambda h} \sigma(h) \\ (\sigma \text{ est paire}) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{i\lambda h} \sigma(-h) \\ (\text{changement } h \text{ en } -h) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-i\lambda h} \sigma(h). \end{aligned}$$

- (2) Pour tout $h \in \mathbb{Z}$, $h \mapsto e^{-i\lambda h}$ est 2π -périodique donc f est 2π -périodique.
- (3) Pour h dans \mathbb{N}^* , on rassemble les termes h et $-h$, nous avons donc, pour tout $\lambda \in \mathbb{R}$,

$$2\pi f(\lambda) = \sigma(0) + \sum_{h \geq 1} \sigma(h) e^{-i\lambda h} + \sigma(-h) e^{i\lambda h}$$

$$\begin{aligned}
&= \sigma(0) + \sum_{h \geq 1} \sigma(h)(e^{-i\lambda h} + e^{i\lambda h}) \\
(5.2) \quad &= \sigma(0) + \sum_{h \geq 1} \sigma(h) \times 2 \cos(\lambda h).
\end{aligned}$$

(4) Nous avons, pour tout k ,

$$\begin{aligned}
\int_{-\pi}^{\pi} |e^{ikh} f(\lambda)| d\lambda &= \int_{-\pi}^{\pi} |f(\lambda)| d\lambda \\
&= \int_{-\pi}^{\pi} \left| \sum_{h \in \mathbb{Z}} e^{-i\lambda h} \sigma(h) \right| d\lambda \\
&\leq \int_{-\pi}^{\pi} \sum_{h \in \mathbb{Z}} |\sigma(h)| d\lambda \\
&= 2\pi \sum_{h \in \mathbb{Z}} |\sigma(h)| \\
&< \infty.
\end{aligned}$$

Donc le théorème de Fubini-Tonelli (cours de L3), nous dit que

$$(5.3) \quad \int_{-\pi}^{\pi} e^{i\lambda k} f(\lambda) d\lambda = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \int_{-\pi}^{\pi} e^{i\lambda k} e^{-i\lambda h} \sigma(h) d\lambda.$$

Or, pour tout $k \in \mathbb{Z}^*$,

$$\int_{-\pi}^{\pi} e^{i\lambda k} d\lambda = \left[\frac{e^{i\lambda k}}{ik} \right]_{-\pi}^{\pi} = 0$$

car $\lambda \mapsto e^{i\lambda k}$ est 2π -périodique. Et pour $k = 0$,

$$\int_{-\pi}^{\pi} e^0 d\lambda = 2\pi.$$

Donc

$$\int_{-\pi}^{\pi} e^{i\lambda k} f(\lambda) d\lambda = \sigma(k).$$

D'où, pour tout k , en prenant la partie réelle,

$$\begin{aligned}
\Re(\sigma(k)) &= \Re \left(\int_{-\pi}^{\pi} e^{i\lambda k} f(\lambda) d\lambda \right) \\
&= \int_{-\pi}^{\pi} \Re(e^{i\lambda k}) f(\lambda) d\lambda \\
&= \int_{-\pi}^{\pi} \cos(\lambda k) f(\lambda) d\lambda.
\end{aligned}$$

□

PROPOSITION 5.4. Une fonction $f :]-\pi; \pi] \rightarrow \mathbb{R}$ est la densité spectrale d'un processus stationnaire si et seulement si

- (1) f est paire,
- (2) f est à valeurs dans \mathbb{R}^+ ,
- (3) $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$.

EXEMPLE 5.5. Soit $X_t = \omega_t$, les ω_t étant i.i.d., centrés, de variance σ^2 ((X_t) est alors un bruit blanc). La fonction de covariance de (X_t) est alors

$$\sigma(h) = \begin{cases} \sigma^2 & \text{si } h = 0, \\ 0 & \text{sinon.} \end{cases}$$

D'où la densité spectrale : $f(\lambda) = \frac{\sigma^2}{2\pi}$ pour tout λ .

EXEMPLE 5.6. On prend un processus $AR(1)$ (stationnaire) : $X_t = \phi X_{t-1} + Z_t$ (Z_t) bruit blanc de variance σ^2 , $\phi \in]0; 1[$). Nous avons vu en exercice que la fonction d'auto-covariance est

$$\sigma(h) = \frac{\sigma^2}{1 - \phi^2} \times \phi^{|h|}, \text{ pour tout } h \geq 0.$$

Nous calculons alors la densité spectrale (en utilisant la parité de σ) :

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-i\lambda h} \phi^{|h|} \sigma(0) \\ &= \frac{\sigma(0)}{2\pi} \left[1 + \sum_{h \geq 1} \phi^h e^{-i\lambda h} + \sum_{h \geq 1} \phi^h e^{i\lambda h} \right] \\ &= \frac{\sigma(0)}{2\pi} \left[1 + \frac{\phi e^{-i\lambda}}{1 - \phi e^{-i\lambda}} + \frac{\phi e^{i\lambda}}{1 - \phi e^{i\lambda}} \right] \\ &= \frac{\sigma(0)}{2\pi} \times \frac{(1 - \phi e^{-i\lambda})(1 - \phi e^{i\lambda}) + \phi e^{-i\lambda}(1 - \phi e^{i\lambda}) + \phi e^{i\lambda}(1 - \phi e^{-i\lambda})}{(1 - \phi e^{-i\lambda})(1 - \phi e^{i\lambda})} \\ &= \frac{\sigma(0)}{2\pi} \times \frac{1 + \phi^2 - 2\phi \cos(\lambda) + 2\phi \cos(\lambda) - 2\phi^2}{1 + \phi^2 - 2\phi \cos(\lambda)} \\ &= \frac{1}{2\pi} \times \frac{1}{1 + \phi^2 - 2\phi \cos(\lambda)}. \end{aligned}$$

Voir le graphique de f dans la figure 5.1.1.

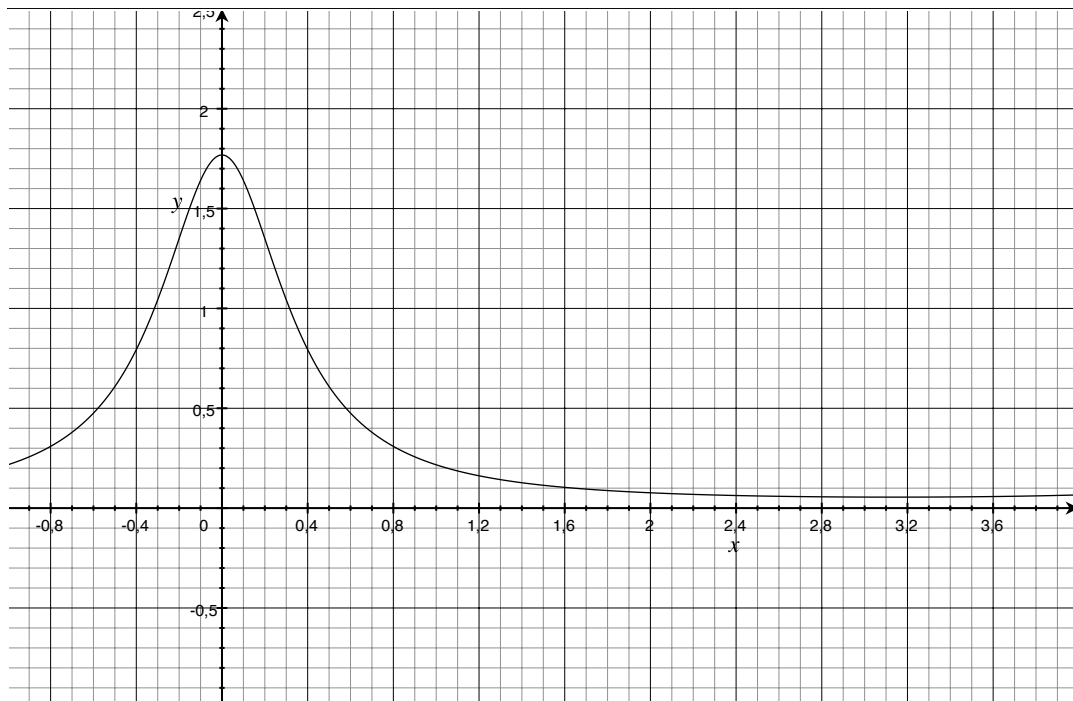


FIGURE 5.1.1. Densité spectrale

5.2. Le périodogramme

DÉFINITION 5.7. Le périodogramme de la série temporelle (x_1, x_2, \dots, x_n) est la fonction

$$I_n : \lambda \in \mathbb{R} \mapsto \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-i\lambda t} \right|^2.$$

DÉFINITION 5.8. La transformée de Fourier discrète de la série temporelle (x_1, x_2, \dots, x_n) est la fonction

$$\begin{aligned} d : \{ \omega_j = \frac{j}{n}, 0 \leq j \leq n-1 \} &\longrightarrow \mathbb{R} \\ w_j &\longmapsto d(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-2i\pi\omega_j t}. \end{aligned}$$

Les ω_j ($0 \leq j \leq n-1$) sont appelés les fréquences fondamentales de Fourier.

PROPOSITION 5.9. *Le périodogramme est 2π -périodique et pair (donc aussi symétrique par rapport à π). Il présente des maxima en les fréquences du processus (si on imagine que le processus contient un mélange de fonctions périodiques).*

REMARQUE 5.10. Attention, \mathbb{R} trace (parfois) $I_n(\lambda/2\pi)$ au lieu de I_n (on s'en rend compte en regardant les graduations en abscisse).

Posons

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|j| \leq m_n} W_n(j) I_n \left(g(n, j) + \frac{2\pi j}{n} \right)$$

avec

$$\begin{aligned} m_n &\xrightarrow{n \rightarrow +\infty} +\infty, \quad \frac{m_n}{n} \xrightarrow{n \rightarrow +\infty} 0, \\ W_n(j) &\geq 0 \text{ et } W_n(j) = W_n(-j) \text{ pour tout } j, \\ \sum_{|j| \leq m_n} W_n(j) &= 1 \text{ et } \sum_{|j| \leq m_n} W_n(j)^2 \xrightarrow{n \rightarrow +\infty} 0, \\ g(n, \lambda) &= \arg \min_j \left| \frac{2\pi j}{n} - \lambda \right|. \end{aligned}$$

PROPOSITION 5.11. *Pour tout λ ,*

$$\hat{f}(\lambda) \xrightarrow{n \rightarrow +\infty} f(\lambda).$$

On pourrait montrer, au prix de quelques calculs, que $I_n(\lambda) \xrightarrow{n \rightarrow +\infty} f(\lambda)$. La convolution par un noyau $W_n(\cdot)$ a pour effet de lisser l'estimateur (le dessin est donc plus joli).

EXEMPLE 5.12. Le \hat{f} ci-dessus est un estimateur de f dit « estimateur à noyau ». Un choix possible de m_n est le suivant : $m_n = \lfloor \sqrt{n} \rfloor$. Et on peut prendre :

$$W_n(j) = \frac{1}{|j|} \times \frac{1}{\sum_{|i| \leq m_n} (1/|i|)}.$$

Nous avons bien

$$\sum_{|j| \leq m_n} W_n(j) = 1$$

et

$$\sum_{|j| \leq m_n} W_n(j)^2 = \left(\sum_{|j| \leq m_n} \frac{1}{|j|^2} \right) \times \frac{1}{\left(\sum_{|j| \leq m_n} \frac{1}{|j|} \right)^2}.$$

Or, $\sum_{j \in \mathbb{Z}} 1/|j|^2 < \infty$ et $\sum_{j \in \mathbb{Z}} 1/|j| = +\infty$ donc

$$\sum_{|j| \leq m_n} W_n(j)^2 \xrightarrow{n \rightarrow +\infty} 0.$$

5.3. Récupération des composantes périodiques

On a déjà vu que la périodicité de (X_t) se retrouve dans $\sigma(\dots)$. On admet que si (X_t) a plusieurs composantes périodiques alors $\sigma(\dots)$ a les mêmes composantes périodiques.

EXEMPLE 5.13. On suppose $\sigma(h) = \phi^{|h|} \cos(\omega h)$, pour tout $h \in \mathbb{Z}$ ($\phi \in]0; 1[$), Calculons :

$$\begin{aligned}
 f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-i\lambda h} \sigma(h) \\
 &= \frac{1}{2\pi} + \frac{1}{2\pi} \sum_{h \geq 1} (e^{-i\lambda h} + e^{i\lambda h}) \phi^h \cos(\omega h) \\
 &\quad (\text{car } \cos \text{ est paire}) \\
 &= \frac{1}{2\pi} + \frac{1}{2\pi} \sum_{h \geq 1} (e^{-i\lambda h} + e^{i\lambda h}) \phi^h \frac{(e^{i\omega h} + e^{-i\omega h})}{2} \\
 &= \frac{1}{2\pi} + \frac{1}{2\pi} \sum_{h \geq 1} \frac{\phi^h}{2} (e^{ih(\lambda+\omega)} + e^{ih(\lambda-\omega)} + e^{ih(-\lambda+\omega)} + e^{ih(-\lambda-\omega)}) \\
 &= \frac{1}{2\pi} + \frac{1}{4\pi} \left(\frac{\phi e^{i(\lambda+\omega)}}{1 - \phi e^{i(\lambda+\omega)}} + \frac{\phi e^{i(\lambda-\omega)}}{1 - \phi e^{i(\lambda-\omega)}} + \frac{\phi e^{i(-\lambda+\omega)}}{1 - \phi e^{i(-\lambda+\omega)}} + \frac{\phi e^{i(-\lambda-\omega)}}{1 - \phi e^{i(-\lambda-\omega)}} \right) \\
 &= \frac{1}{2\pi} + \frac{1}{4\pi} \left(\frac{\phi e^{i(\lambda+\omega)}(1 - \phi e^{i(-\lambda-\omega)}) + \phi e^{i(-\lambda-\omega)}(1 - \phi e^{i(\lambda+\omega)})}{(1 - \phi e^{i(\lambda+\omega)})(1 - \phi e^{i(-\lambda-\omega)})} \right. \\
 &\quad \left. + \frac{\phi e^{i(\lambda-\omega)}(1 - \phi e^{i(-\lambda+\omega)}) + \phi e^{i(-\lambda+\omega)}(1 - \phi e^{i(\lambda-\omega)})}{(1 - \phi e^{i(\lambda-\omega)})(1 - \phi e^{i(-\lambda+\omega)})} \right) \\
 &= \frac{1}{2\pi} + \frac{1}{4\pi} \left(\frac{2 \cos(\lambda + \omega) - 2\phi^2}{1 + \phi^2 - 2 \cos(\lambda + \omega)} + \frac{2 \cos(\lambda - \omega) - 2\phi^2}{1 + \phi^2 - 2 \cos(\lambda - \omega)} \right) \\
 &= \frac{1}{4\pi} \left(\frac{1 - \phi^2}{1 + \phi^2 - 2 \cos(\lambda + \omega)} + \frac{1 - \phi^2}{1 + \phi^2 - 2 \cos(\lambda - \omega)} \right).
 \end{aligned}$$

La fonction

$$\lambda \in \mathbb{R} \mapsto \frac{1 - \phi^2}{1 + \phi^2 - 2 \cos(\lambda + \omega)}$$

est 2π -périodique et a un maximum en $-\omega$. La fonction

$$\lambda \in \mathbb{R} \mapsto \frac{1 - \phi^2}{1 + \phi^2 - 2 \cos(\lambda - \omega)}$$

est 2π -périodique et a un maximum en ω . Voir la figure 5.3.1 pour le graphe de f (avec $\phi = 0,7$ et $\omega = 0,5$). La fonction f présente des maxima en ω et $-\omega$ sur $[-\pi; \pi]$. Attention, seuls les maxima sur $[0; \pi]$ sont à prendre en compte. Nous voyons ici que le maximum en ω reflète bien le fait que $\sigma(\dots)$ a une composante de fréquence ω . Si (X_t) est un processus de fonction de covariance $\sigma(\dots)$, on déduit du graphique de f que (X_t) a une composante périodique de fréquence ω , et donc de période $2\pi/\omega$.

Pour un processus qui n'a qu'une composante périodique, nous aurions pu détecter cette composante en regardant $\sigma(\dots)$ (voir remarque 1.7). Nous verrons en exercice des cas où l'analyse spectrale permet de détecter plusieurs composantes périodiques.

5.4. Feuille d'exercices numéro 6 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf.

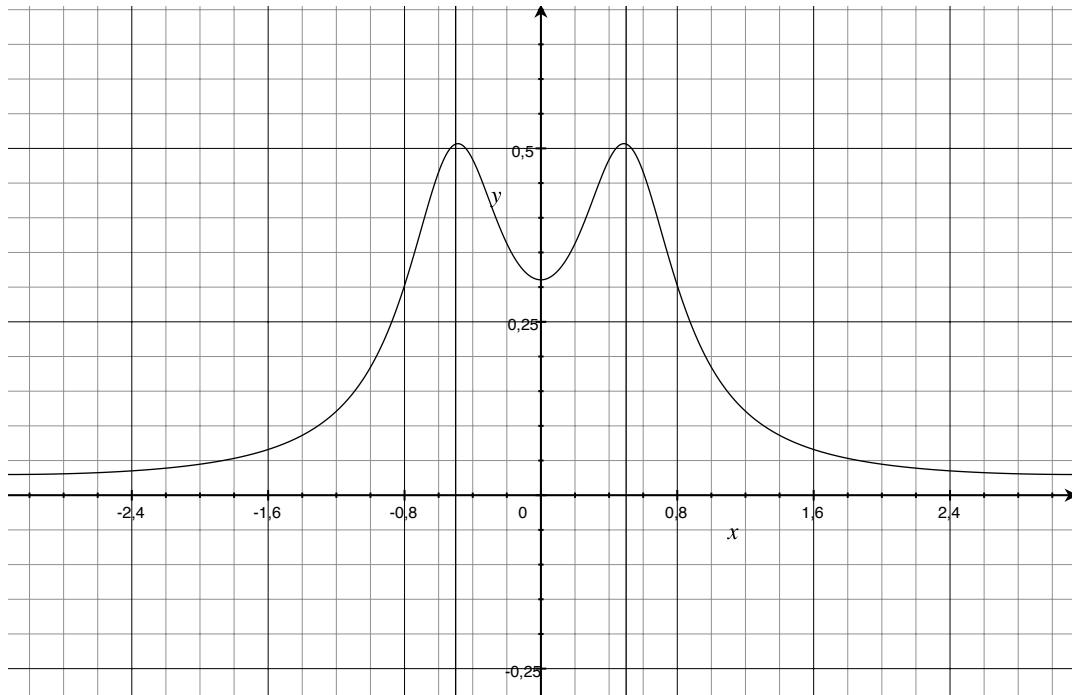


FIGURE 5.3.1. Densité spectrale

5.4.1. Données simulées.

- (1) Soient les fonctions (de \mathbb{Z} dans \mathbb{Z}) $x_1(t) = 2 \cos(2\pi \times t \times 5/100)$, $x_2(t) = 2 \cos(2\pi \times t \times 10/100)$, $x_3(t) = 2 \cos(2\pi \times t \times 20/100)$, $x(t) = x_1(t) + x_2(t) + x_3(t)$. Tracer les graphes de x_1 , x_2 , x_3 , x pour t dans $\{1, 2, \dots, 100\}$. On range les ordonnées du tracé de x dans une série temporelle \mathbf{x} .
- (2) Tracer le périodogramme de \mathbf{x} .
- (3) Tracer une estimée de la densité spectrale de \mathbf{x} . Bien noter la différence d'échelle en abscisse entre le tracé de la densité spectrale et celui du périodogramme.
- (4) Au vu des graphiques, quelles sont les composantes périodiques de \mathbf{x} (attention à ne pas se laisser tromper par l'échelle en abscisse) ?
- (5) Tracer une série y de type AR_3 avec $a_1 = 0$, $a_2 = -0,5$, $a_3 = 0.2$.
- (6) Tracer une estimée de la densité spectrale de y .
- (7) Au vu du graphique, quelles sont les composantes périodiques de y (attention à ne pas se laisser tromper par l'échelle en abscisse) ?

5.4.2. Données réelles.

- (1) La série temporelle `sunspot.year` (contenue dans `R`) contient le nombre moyen de taches solaires observées par mois (on fait à chaque fois la moyenne sur une année). Tracer la densité spectrale de cette série.
- (2) Au vu du graphique, quelles sont les composantes périodiques de `sunspot.year` (attention à ne pas se laisser tromper par l'échelle en abscisse) ?
- (3) Récupérer le fichier `soi.clean.dat`¹. Il contient les différences de pression mensuelles entre Tahiti et Darwin dans lesquelles les oscillations annuelles ont déjà été gommées. Cette série est un indicateur des événements « El Niño ». Tracer la densité spectrale de cette série (que l'on appellera \mathbf{x}). Tracer sur le même graphique la ou les ligne(s) verticale(s) correspondant aux périodes entières.

1. <http://math.unice.fr/~rubentha/enseignement/soi.clean.dat>

- (4) Au vu du graphique, quelles sont les composantes périodiques de \mathbf{x} (attention à ne pas se laisser tromper par l'échelle en abscisse) ?

Appendice : mise en œuvre sous R. Pour tracer une fonction : remplir un vecteur avec les abscisses des points que l'on veut tracer (exemple : $\mathbf{t}=(1:100)/(2*\pi)$), remplir un vecteur avec les ordonnées des points que l'on veut tracer (exemple : $\mathbf{x}=\cos(\mathbf{t})$), puis utiliser la commande `plot.ts` (exemple : `plot.ts(x,ylim=c(-2,2),main=expression(exemple))`, l'option `ylim` permet de fixer les bords du cadre en ordonnée, l'option `main(...)` permet d'écrire un titre).

Pour tracer le périodogramme d'une série \mathbf{x} de n points : $\mathbf{P}=\text{abs}(\text{fft}(\mathbf{x})/n)^2$;
 $\mathbf{Fr}=(0:(n-1))*2*\pi/n$; `plot(Fr,P,type='o',xlab='fréquence',ylab='périodogramme')` (les options `xlab`, `ylab` permettent d'écrire des légendes sur les axes).

Pour tracer une estimée de la densité spectrale d'une série \mathbf{x} : choisir des poids $W_n(\cdot)$ comme dans le cours (rentrer l'instruction `k=kernel("daniell",4)`), puis tracer la densité spectrale avec l'instruction `spec.pgram(x,k,taper=0,log='no')`.

5.4.3. Exercices sur table.

- (1) (a) Soit (X_t) un processus $AR(1)$: $X_t = \phi X_{t-1} + Z_t$ avec $\phi = 0,99$, (Z_t) un bruit blanc centré de variance 1. Calculer les covariances de (X_t) .
 (b) Calculer la densité spectrale de (X_t) .
 (c) Soit $Y_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$. Calculer les covariances de (Y_t) .
 (d) Calculer la densité spectrale de (Y_t) .
- (2) Soient (X_t) , (Y_t) des processus stationnaires centrés indépendants de densités spectrales respectivement f_X , f_Y . Soit $Z_t = X_t + Y_t$. Calculer la densité spectrale de Z en fonction de f_X , f_Y .
- (3) Soit (W_t) un bruit blanc de variance 1 (centré). Soit $X_t = W_t - \theta W_{t-1}$ ($\theta \in \mathbb{R}$)
 (a) Est-ce que (X_t) est stationnaire ? Calculer les covariances de (X_t) .
 (b) Calculer la densité spectrale de (X_t) .

5.4.4. Corrigés des exercices sur table.

- (1) (a) Nous avons vu (Chapitre 4, section 4.8, exercice 4) que pour un tel processus,

$$\sigma(h) = \frac{\phi^{|h|}}{1 - \phi^2}.$$

- (b) Nous calculons, pour tout λ dans \mathbb{R} ,

$$\begin{aligned} f_X(\lambda)(1 - \phi^2) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\lambda} \sigma(h) \times (1 - \phi^2) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\lambda} \phi^h \\ &= \frac{1}{2\pi} \left(1 + \sum_{h \geq 1} \phi^h e^{-i\lambda h} + \sum_{h \geq 1} \phi^h e^{i\lambda h} \right) \\ &= \frac{1}{2\pi} \left(1 + \frac{\phi e^{-i\lambda}}{1 - \phi e^{-i\lambda}} + \frac{\phi e^{i\lambda}}{1 - \phi e^{i\lambda}} \right) \\ &= \frac{1}{2\pi} \left(1 + \frac{\phi e^{-i\lambda}(1 - \phi e^{i\lambda}) + \phi e^{i\lambda}(1 - \phi e^{-i\lambda})}{(1 - \phi e^{i\lambda})(1 - \phi e^{-i\lambda})} \right) \\ &= \frac{1}{2\pi} \left(1 + \frac{2\phi \cos(\lambda) - 2\phi^2}{1 + \phi^2 - 2\phi \cos(\lambda)} \right) \end{aligned}$$

$$= \frac{1}{2\pi} \times \frac{1 - \phi^2}{1 + \phi^2 - 2\phi \cos(\lambda)}.$$

Donc

$$f_X(\lambda) = \frac{1}{2\pi(1 + \phi^2 - 2\phi \cos(\lambda))}.$$

(c) Nous avons, pour tout t et tout $h > 0$,

$$\begin{aligned} \mathbb{E}(Y_t Y_{t+h}) &= \frac{1}{9} \mathbb{E}((X_{t-1} + X_t + X_{t+1})(X_{t+h-1} + X_{t+h} + X_{t+h+1})) \\ &= \frac{1}{9} (\sigma(h) + \sigma(h+1) + \sigma(h+2) + \sigma(h-1) + \sigma(h) + \sigma(h+1) \\ &\quad + \sigma(h-2) + \sigma(h-1) + \sigma(h)) \\ &= \frac{1}{9} (\sigma(h-2) + 2\sigma(h-1) + 3\sigma(h) + 2\sigma(h+1) + \sigma(h+2)) \end{aligned}$$

(d) Nous calculons, pour tout λ dans \mathbb{R} ,

$$\begin{aligned} f_Y(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-i\lambda h} \frac{(\sigma(h-2) + 2\sigma(h-1) + 3\sigma(h) + 2\sigma(h+1) + \sigma(h+2))}{9} \\ &= \frac{1}{18\pi(1 - \phi^2)} \sum_{h=-\infty}^{+\infty} \phi^{|h|} \left(e^{-i\lambda(h+2)} + 2e^{-i\lambda(h+1)} + 3e^{-i\lambda h} + 2e^{-i\lambda(h-1)} + e^{-i\lambda(h-2)} \right) \\ &= \frac{(e^{-2i\lambda} + 2e^{-i\lambda} + 3 + 2e^{i\lambda} + e^{2i\lambda})}{9} f_X(\lambda) \\ &= \frac{(\cos(2\lambda) + 2\cos(\lambda) + 3)}{9} \times \frac{1}{2\pi(1 + \phi^2 - 2\phi \cos(\lambda))}. \end{aligned}$$

(2) Nous calculons les covariances (pour tout h) :

$$\begin{aligned} \sigma_Z(h) &= \mathbb{E}(Z_t Z_{t+h}) \\ &= \mathbb{E}((X_t + Y_t)(X_{t+h} + Y_{t+h})) \\ &= \mathbb{E}(X_t X_{t+h}) + \mathbb{E}(Y_t Y_{t+h}) + 0 \\ &= \sigma_X(h) + \sigma_Y(h). \end{aligned}$$

Et donc, pour tout λ ,

$$\begin{aligned} f_Z(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-i\lambda h} (\sigma_X(h) + \sigma_Y(h)) \\ &= f_X(h) + f_Y(h). \end{aligned}$$

(3) (a) Nous avons pour tout t , $\mathbb{E}(X_t) = 0$ et

$$\begin{aligned} \mathbb{E}(X_t^2) &= \mathbb{E}((W_t - \theta W_{t-1})^2) \\ &= 1 + \theta^2, \end{aligned}$$

$$\begin{aligned} \mathbb{E}(X_t X_{t+1}) &= \mathbb{E}((W_t - \theta W_{t-1})(W_{t+1} - \theta W_t)) \\ &= -\theta, \end{aligned}$$

et pour $h \geq 2$

$$\mathbb{E}(X_t X_{t+h}) = 0.$$

Donc la fonction de covariance de (X_t) est telle que : $\sigma(0) = 1 + \theta^2$, $\sigma(1) = -\theta$, $\sigma(h) = 0$ pour $h \geq 2$.

(b) Nous calculons, pour tout λ dans \mathbb{R} ,

$$f(\lambda) = \frac{1}{2\pi} (-2\theta \cos(\lambda) + (1 + \theta^2)).$$

Processus ARCH et GARCH

Dans les modèles précédent, on pouvait se ramener à un processus stationnaire (en soustrayant la tendance, la saisonnalité ...). Il arrive que l'on tombe sur un processus qui semble non stationnaire (par exemple, de moyenne ou de variance non constante).

6.1. Processus ARCH

DÉFINITION 6.1. Un processus $ARCH(p)$ est un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifiant une relation de récurrence du type

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

avec des coefficients $\alpha_0, \dots, \alpha_p \geq 0$.

On peut décrire la simulation de ce processus de manière récursive : si on a simulé X_0, \dots, X_{t-1} , on calcule σ_t , on peut alors simuler X_t . On remarque que si X_t est très grand alors σ_{t+1} est très grande et X_{t+1} a de grandes chances d'être très grand aussi.

6.1.1. Rappels de probabilité. Si (X, Y) est un couple de variables aléatoires de densité f (par exemple dans \mathbb{R}^2), la densité conditionnelles de X sachant $Y = y$ est définie par

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

où f_Y est la densité de Y . On peut alors calculer l'espérance conditionnelle de n'importe quelle fonction de X :

$$\mathbb{E}(\varphi(X) | Y = y) = \int_{\mathbb{R}} \varphi(x) f_{X|Y}(x) dx.$$

Notons $\mathcal{I}_s = \{X_{s-1}, X_{s-2}, \dots\}$ (pour tout $s \geq 1$). Nous avons alors, pour tout processus (X_t) ,

$$\text{Var}(X_t | \mathcal{I}_{t-1}) = \mathbb{E}(X_t^2 | \mathcal{I}_{t-1}) - \mathbb{E}(X_t | \mathcal{I}_{t-1})^2.$$

Dans le cas d'un processus $GARCH$, nous avons directement la loi de X_t sachant X_{t-1}, X_{t-2}, \dots (donc pas besoin de la formule ci-dessus). Si nous posons $Y = (X_{t-1}, X_{t-2}, \dots)$, $X_t = X$, nous pouvons écrire

$$f_{X|Y}(x) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{x^2}{2\sigma_t^2}\right)$$

(σ_t est une fonction de Y). Comme annoncé dans la formule ci-dessus, Y est présent en tant que paramètre dans $f_{X|Y}$. On trouve bien $\text{Var}(X_t | Y) = \sigma_t^2$.

6.1.2. Propriétés. Soit (X_t) un processus vérifiant la définition ci-dessus.

PROPOSITION 6.2. Nous avons pour tout t et $h \geq 1$:

- $\mathbb{E}(X_t | \mathcal{I}_{t-1}) = 0$,
- $\mathbb{E}(X_t) = 0$,
- $\text{Var}(X_t | \mathcal{I}_{t-1}) = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$,
- $\text{Var}(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i}$, si $\sum_{i=1}^p \alpha_i < 1$,
- $\text{Cov}(X_t, X_{t+h} | \mathcal{I}_{t-1}) = 0$, pour $h \geq 1$,
- $\text{Cov}(X_t, X_{t+h}) = 0$, pour $h \geq 1$.

DÉMONSTRATION. Par définition, $\mathbb{E}(X_t|\mathcal{I}_{t-1}) = 0$. Donc $\mathbb{E}(X_t) = 0$. Par définition :

$$\text{Var}(X_t|\mathcal{I}_{t-1}) = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2.$$

Si $\sum_{i=1}^p \alpha_i < 1$, puisque X_t est centré, nous avons

$$\begin{aligned} \text{Var}(X_t) &= \mathbb{E}(X_t^2) \\ &= \mathbb{E}(\mathbb{E}(X_t^2|\mathcal{I}_{t-1})) \\ &= \mathbb{E}(\sigma_t^2) \\ &= \mathbb{E}(\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2) \\ X \text{ étant stationnaire} &= \alpha_0 + \sum_{i=1}^p \alpha_i \text{Var}(X_t). \end{aligned}$$

D'où

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i}.$$

Nous avons

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}(X_t X_{t+h}) - \mathbb{E}(X_t)\mathbb{E}(X_{t+h}) \\ &= \mathbb{E}(\mathbb{E}(X_t X_{t+h}|\mathcal{I}_{t+h-1})) - 0 \\ &= \mathbb{E}(X_t \mathbb{E}(X_{t+h}|\mathcal{I}_{t+h-1})) \\ &= 0, \end{aligned}$$

et

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}|\mathcal{I}_{t-1}) &= \mathbb{E}(X_t X_{t+h}|\mathcal{I}_{t-1}) - \mathbb{E}(X_t|\mathcal{I}_{t-1})\mathbb{E}(X_{t+h}|\mathcal{I}_{t-1}) \\ &= \mathbb{E}(\mathbb{E}(X_t X_{t+h}|\mathcal{I}_{t+h-1})|\mathcal{I}_{t-1}) - 0 \\ &= 0. \end{aligned}$$

□

REMARQUE 6.3. Un processus ARCH est conditionnellement hétéroscédastique (la variance de X_t conditionné aux valeurs X_{t-1}, X_{t-2}, \dots n'est pas constante) mais il est inconditionnellement homoscedastique (la variance de X_t ne dépend pas de t).

PROPOSITION 6.4. Si $\sum_{i=1}^p \alpha_i < 1$ alors il existe un processus ARCH stationnaire de paramètres $(\alpha_0, \alpha_1, \dots, \alpha_p)$.

DÉMONSTRATION. On se donne une collection de variable $(e_t)_{t \in \mathbb{Z}}$, i.i.d. de loi $\mathcal{N}(0; 1)$. Posons pour tout t ,

$$\begin{cases} \sigma_t^2 = \alpha_0 + \sum_{k=1}^{+\infty} \sum_{1 \leq i_1, \dots, i_k \leq p} \alpha_0 \times \alpha_{i_1} \cdots \alpha_{i_k} \times e_{t-i_1}^2 e_{t-i_1-i_2}^2 \cdots e_{t-i_1-\dots-i_k}^2, \\ X_t = \sigma_t e_t. \end{cases}$$

Montrons que les σ_t ainsi définis sont finis p.s. Nous avons

$$\begin{aligned} \mathbb{E}(\sigma_t^2) &= \alpha_0 + \alpha_0 \sum_{k=1}^{+\infty} \sum_{1 \leq i_1, \dots, i_k \leq p} \alpha_0 \times \alpha_{i_1} \cdots \alpha_{i_k} \\ &= \alpha_0 \left(1 + \sum_{k=1} (\alpha_1 + \cdots + \alpha_p)^k \right) \\ &< +\infty \end{aligned}$$

car $\alpha_1 + \cdots + \alpha_p \in]0; 1[$. Vérifions maintenant que nous avons la relation de récurrence voulue. Nous avons, pour tout t ,

$$\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \sigma_{t-i}^2 e_{t-i}^2$$

$$= \alpha_0 + \sum_{i=1}^p \left(\alpha_i e_{t-i}^2 \sum_{k=0}^{+\infty} \alpha_0 \times \sum_{1 \leq i_1, \dots, i_k \leq p} \alpha_{i_1} \dots \alpha_{i_p} e_{t-i-i_1}^2 \dots e_{t-i-i_1-\dots-i_k}^2 \right),$$

avec la convention $\sum_{1 \leq i_1, \dots, i_k \leq p} (\dots) = 1$ si $k = 0$. Et donc

$$\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 = \alpha_0 + \alpha_0 \sum_{k=1}^{+\infty} \sum_{1 \leq i_1, \dots, i_k \leq p} \alpha_{i_1} \dots \alpha_{i_p} e_{t-i_1}^2 \dots e_{t-i_1-\dots-i_k}^2.$$

Par construction, la suite (X_t) est stationnaire. \square

REMARQUE 6.5. Soit (X_t) un processus ARCH de variance σ^2 .

— La « skewness » de (X_t) (définie par $\mathbb{E}((X_t - \mathbb{E}(X_t))^3)/\sigma^3$) est nulle.

— La « kurtosis » de (X_t) (définie par $\mathbb{E}((X_t - \mathbb{E}(X_t))^4)/\sigma^4 - 3\mathbb{E}(X_t^2)^2/\sigma^4$) est > 0

Si $Z_t \sim \mathcal{N}(0, \sigma^2)$ alors $\mathbb{E}(Z_t^4) = 3\mathbb{E}(Z_t^2)^2$ donc Z_t est de kurtosis nulle. La kurtosis > 0 du ARCH veut dire que sa densité est plus étalée que celle de la gaussienne.

6.2. Processus GARCH

DÉFINITION 6.6. Un processus GARCH(p, q) est un processus $(X_t)_{t \in \mathbb{Z}}$ vérifiant une relation de récurrence du type :

$$\begin{cases} X_t = \sigma_t e_t \text{ (avec des } e_t \text{ i.i.d. } \sim \mathcal{N}(0; 1)), \\ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2, \end{cases}$$

avec $\alpha_0 > 0$, $\alpha_i \geq 0$ pour i dans $\{1, 2, \dots, p\}$, $\beta_i \geq 0$ pour i dans $\{1, 2, \dots, q\}$.

REMARQUE 6.7. — Un processus GARCH peut être vu comme un processus ARCH d'ordre infini.

— Un GARCH($p, 0$) est un ARCH(p).

PROPOSITION 6.8. Soit (X_t) un processus GARCH(p, q), supposé stationnaire. Notons, pour tout t , $\mathcal{I}_t = (X_t, \sigma_t, X_{t-1}, \sigma_{t-1}, \dots)$. Nous avons les propriétés suivantes pour tout t .

- $\mathbb{E}(X_t | \mathcal{I}_{t-1}) = 0$,
- $\mathbb{E}(X_t) = 0$,
- $\text{Cov}(X_t, X_{t+h} | \mathcal{I}_{t-1}) = 0$ pour tout $h > 0$,
- $\text{Cov}(X_t, X_{t+h}) = 0$ pour tout $h > 0$.

DÉMONSTRATION. Soit $t \geq 1$. Par définition, $\mathbb{E}(X_t | \mathcal{I}_{t-1}) = 0$. Et donc $\mathbb{E}(X_t) = 0$. Si $h > 0$, nous avons

$$\begin{aligned} \mathbb{E}(X_t X_{t+h} | \mathcal{I}_{t-1}) &= \mathbb{E}(X_t \mathbb{E}(X_{t+h} | \mathcal{I}_{t+h-1}) | \mathcal{I}_{t-1}) \\ &= 0. \end{aligned}$$

Et donc $\mathbb{E}(X_t X_{t+h}) = 0$. \square

PROPOSITION 6.9. Si (X_t) est un processus GARCH(p, q) alors (X_t^2) est un processus ARMA(m, q) avec $m = \sup(p, q)$.

DÉMONSTRATION. Nous avons, pour tout t ,

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2,$$

$$\begin{aligned} \sigma_t^2 + X_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + X_t^2 + \beta_1 X_{t-1}^2 + \dots + \beta_q X_{t-q}^2 \\ &\quad - \beta_1 (X_{t-1}^2 - \sigma_{t-1}^2) - \dots - \beta_q (X_{t-q}^2 - \sigma_{t-q}^2). \end{aligned}$$

Nous posons $W_t = X_t^2 - \sigma_t^2$. Nous obtenons alors

$$X_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) X_{t-i}^2 + W_t - \sum_{i=1}^q \beta_i W_t,$$

avec la convention $\alpha_i = 0$, si $i > p$ et $\beta_i = 0$, si $i > q$. Les (W_t) forment un bruit blanc (nous ne démontrons pas ce point), donc (X_t^2) est un ARMA(m, q). \square

En conséquence de la proposition ci-dessus, pour identifier un $GARCH(p, q)$, on identifiera d'abord le processus $ARMA(p, q)$ qui modélise (X_t^2) . Pour identifier p dans le cas $m = q$ (c'est à dire $p \leq q$), il faut estimer les coefficients $\alpha_1, \dots, \alpha_q$ (possible car on peut estimer les coefficients du processus $ARMA(m, q)$) (sont-ils significativement non nuls?).

6.3. Feuille d'exercices numéro 7 (durée : 3h)

Préliminaires. Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf.

6.3.1. EuStockMarket (valeurs à la fermeture de divers indices des marchés européens). Charger les logarithmes des accroissements de l'indice DAX en exécutant :
`dax <- diff(log(EuStockMarkets))[, "DAX"]`. Soit `dax2=dax^2` (les carrés des composantes de `dax`).

- (1) On cherche à ajuster un modèle $ARMA_{m,q}$ avec $m \geq q$ sur `dax2`. Choisir (m, q) dans $\{(1, 0), (1, 1), (2, 0), (2, 1), (2, 2)\}$ de manière à minimiser la quantité AIC (utiliser les instructions du TP 5).
- (2) Proposer un modèle $GARCH_{p,q}$ modélisant `dax`. Tester les résidus.

6.3.2. NYSE (New-York stock exchange daily return). Charger le fichier `nyse.dat` à l'adresse

`http://math.unice.fr/~rubentha/enseignement/nyse.dat`

- (1) On suppose que `nyse` suit un modèle $GARCH(1, 1)$. On notera (X_t) le processus. Estimer les paramètres du modèle.
- (2) Tracer dans la même fenêtre : `nyse` entre les temps 900 et 1000 et, pour chaque temps t dans $\{900, \dots, 1000\}$, les deux extrémités d'un intervalle $[-m; m]$ tel que

$$\mathbb{P}(X_t \in [-m; m] | X_{t-1}, X_{t-2}, \dots) = 0,32.$$

Mise en œuvre sous R. Pour utiliser les fonctions spécifiques à l'étude des modèles $ARCH$ et $GARCH$, il faut avant tout charger le package `tseries` à l'aide de la commande `library(tseries)` (et/ou en cherchant dans le menu « packages » de R). La fonction `garch` permet d'estimer un $GARCH_{p,q}$: `serie<-garch(data,order=c(p,q))` Parmi les sorties de cette fonction : `coef`, `residuals`, `fitted.values`. La sortie `serie$fitted.values` contient pour chaque temps t les valeurs $-\sigma_t$ et σ_t (avec la notation standard : $X_t = \sigma_t e_t$ ($e_t \sim \mathcal{N}(0, 1)$) et σ_t^2 est la variance calculée à partir de $X_{t-1}, X_{t-2}, \dots, \sigma_{t-1}, \sigma_{t-2}, \dots$. On rappelle que si $Z \sim \mathcal{N}(0, 1)$, $\mathbb{P}(Z \leq 1) = 0,8413$ (approximativement).

Instructions utile : si `A` est un tableau à deux colonnes alors `A[,j]` est la j -ème colonne de `A`. Pour fixer la taille de la fenêtre graphique en ordonnée : `plot(x,ylim=c(-0.7,0.9))` (c'est un exemple).

6.4. Feuille d'exercices numéro 8 (révisions)

- (1) Soit X un $AR(1)$ (stationnaire) défini par :

$$\forall t \in \mathbb{Z}, X_t = \phi X_{t-1} + Z_t,$$

avec Z un bruit blanc centré de variance σ^2 et $|\phi| < 1$. On pose :

$$\forall t \in \mathbb{Z}, Y_t = X_t - X_{t-1}.$$

- (a) Montrer que Y est un processus stationnaire centré, puis que c'est un $ARMA(p, q)$. Préciser les valeurs de p et q et donner l'équation de récurrence vérifiée par Y .
- (b) Que vaut la variance de Y_t (pour tout t) ?

- (2) Soit X un processus $ARMA$ (stationnaire) vérifiant

$$\forall t \in \mathbb{Z}, X_t - \frac{7}{6}X_{t-1} + \frac{1}{3}X_{t-2} = Z_t - \frac{1}{4}Z_{t-1} - \frac{1}{8}Z_{t-2},$$

avec Z un bruit blanc (centré). Montrer que :

$$\forall t \in \mathbb{Z}, X_t - \frac{2}{3}X_{t-1} = Z_t + \frac{1}{4}Z_{t-1}.$$

- (3) La série `snkf.rda` contient le trafic mensuel sur les lignes SNCF de janvier 1963 à décembre 1980 en millions de passagers par kilomètres. Rappel : Δ_T est l'opérateur des différences avec un décalage de T .

- (a) Notons X la série à étudier. Tracer la densité spectrale de X et les autocorrélations de X pour trouver la période T de la composante saisonnière de X .
- (b) Notons $Y_1 = \Delta_1 X$, $Y_2 = \Delta_T X$, $Y_3 = \Delta_T \circ \Delta_1 X$. Pour chacune de ces séries $Y...$,
 - (i) tracer les autocorrélations et les autocorrélations partielles,
 - (ii) en déduire un modèle $ARMA$ pour $Y...$ (on doit pouvoir voir le nombre de coefficients non nuls sur les graphiques précédents) et estimer les coefficients,
 - (iii) tracer la série des résidus et tester la blancheur des résidus.
(On doit trouver que seul Y_3 est un $ARMA$.)
 - (iv) Trouver, parmi les $(p, q) \in \{0, 1, 2\}^2$, le couple minimisant le critère AIC pour Y_3 .
- (c) Prédire le trafic de l'année 1981.

- (4) On étudie la série `serie1.rda` (disponible sur <http://math.unice.fr/~rubentha/enseignement/serie1.rda>, sauvegardez le fichier sur l'ordinateur puis tapez `load('chemin_complet/serie1.rda')`, la série `serie1` sera chargée dans `R`). Le modèle choisi pour cette série est

$$(6.1) \quad \forall t \in \{1, 2, \dots, n\}, x_t = m(t) + s(t) + \epsilon_t,$$

où m est une tendance polynômiale, s une saisonnalité et ϵ un processus $ARMA$ (stationnaire) centré (conduit par un bruit blanc gaussien).

- (a) Tracer la série `serie1.rda` et son périodogramme. Quelle est la période p de la saisonnalité s ?
 - (b) Estimer une tendance polynômiale et une saisonnalité de période p par moindres carrés ordinaires sur la série `serie1.rda`.
 - (c) Sur le résidu de cette estimation, proposer un modèle $ARMA$.
 - (d) Une fois identifié l'ordre du bruit $ARMA$, ajuster directement avec `arima` un modèle de type (6.1) sur `serie1.rda`. Pensez à valider votre modèle et donnez les valeurs estimées des paramètres de l' $ARMA$.
 - (e) Représenter l'estimation de $m + s$.
 - (f) En utilisant la fonction `predict`, prédire les 12 prochaines valeurs de la série `serie1.rda`.
- (5) Dans cet exercice, on étudie la série `serie2.rda` (disponible sur <http://math.unice.fr/~rubentha/enseignement/serie2.rda>, à charger dans `R` à l'aide de l'instruction `load`).
- (a) Tracer la série `serie2.rda`. Est-elle stationnaire ?
 - (b) Ajuster un modèle $ARMA$ sur cette série. Pensez à valider votre modèle.
 - (c) Peut-on considérer cet $ARMA$ comme étant de moyenne nulle ?

6.5. Corrigé de la feuille d'exercices numéro 8

- (1) (a) Pour tout t , $\mathbb{E}(X_t) = \phi \mathbb{E}(X_{t-1}) + 0$. Comme X est stationnaire, $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1})$, et donc, puisque $|\phi| < 1$, $\mathbb{E}(X_t) = 0$. Donc, pour tout t , $\mathbb{E}(Y_t) = \mathbb{E}(X_t) - \mathbb{E}(X_{t-1}) = 0$. Pour tout t , $\mathbb{E}(Y_t^2) = \text{Var}(X_t) + \text{Var}(X_{t-1}) - 2\text{Cov}(X_t, X_{t-1})$, et puisque X est stationnaire, cette quantité ne dépend pas de t . Pour tout t et $h > 0$,

$$\begin{aligned}\mathbb{E}(Y_t Y_{t+h}) &= \mathbb{E}((X_t - X_{t-1})(X_{t+h} - X_{t+h-1})) \\ &= \text{Cov}(X_t, X_{t+h}) - \text{Cov}(X_t, X_{t+h-1}) - \text{Cov}(X_{t-1}, X_{t+h}) + \text{Cov}(X_{t-1}, X_{t+h-1}),\end{aligned}$$

qui ne dépend pas de t puisque X est stationnaire. Donc Y est stationnaire.

On calcule :

$$\begin{aligned}Y_t &= \phi X_{t-1} + Z_t - (\phi X_{t-2} + Z_{t-1}) \\ &= \phi Y_{t-1} + Z_t - Z_{t-1}.\end{aligned}$$

Donc Y est un $ARMA(1, 1)$.

- (b) On calcule (pour t quelconque) :

$$\text{Var}(X_t) = \phi^2 \text{Var}(X_{t-1}) + \sigma^2.$$

Donc

$$\text{Var}(X_t) = \frac{\sigma^2}{1 - \phi^2}.$$

Nous avons pour tout t :

$$\begin{aligned}\mathbb{E}(X_t X_{t-1}) &= \phi \text{Var}(X_{t-1}) + 0 \\ &= \frac{\phi \sigma^2}{1 - \phi^2}.\end{aligned}$$

D'où :

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(X_t) + \text{Var}(X_{t-1}) - 2\text{Cov}(X_t, X_{t-1}) \\ &= \frac{2\sigma^2}{1 - \phi^2} - 2\left(\frac{\phi \sigma^2}{1 - \phi^2}\right) \\ &= \frac{2\sigma^2(1 - \phi)}{1 - \phi^2} = \frac{2\sigma^2}{1 + \phi}.\end{aligned}$$

- (2) Nous voulons éliminer les X dans l'expression $X_t - (2/3)X_{t-1}$ (t quelconque fixé). Nous commençons donc le calcul par

$$\begin{aligned}X_t - \frac{2}{3}X_{t-1} &= X_t - \frac{7}{6}X_{t-1} + \frac{1}{3}X_{t-2} + \frac{X_{t-1}}{2} - \frac{1}{3}X_{t-2} \\ &= \left(Z_t - \frac{Z_{t-1}}{4} - \frac{Z_{t-2}}{8}\right) + \frac{1}{2}\left(X_{t-1} - \frac{7}{6}X_{t-2} + \frac{X_{t-3}}{3}\right) \\ &\quad + \frac{1}{4}X_{t-2} - \frac{1}{6}X_{t-3} \\ &= \left(Z_t - \frac{Z_{t-1}}{4} - \frac{Z_{t-2}}{8}\right) + \frac{1}{2}\left(X_{t-1} - \frac{7}{6}X_{t-2} + \frac{X_{t-3}}{3}\right) \\ &\quad + \frac{1}{4}\left(X_{t-2} - \frac{7}{6}X_{t-3} + \frac{X_{t-4}}{3}\right) + \frac{1}{8}X_{t-3} - \frac{1}{12}X_{t-4}.\end{aligned}$$

Ce qui nous donne l'idée de calculer $\sum_{k \geq 0} \left[\frac{1}{2^k} \left(X_{t-k} - \frac{7}{6}X_{t-k-1} + \frac{1}{3}X_{t-k-2} \right) \right]$. Nous commençons par montrer que cette suite est bien définie (en utilisant le fait que les Z_t sont centrés et tous de même variance)

$$\mathbb{E} \left(\left(\sum_{k \geq 0} \left[\frac{1}{2^k} \left(X_{t-k} - \frac{7}{6}X_{t-k-1} + \frac{1}{3}X_{t-k-2} \right) \right] \right)^2 \right)$$

$$\begin{aligned}
&= \mathbb{E} \left(\left(\sum_{k \geq 0} \left[\frac{1}{2^k} \left(Z_{t-k} - \frac{1}{4} Z_{t-k-1} - \frac{1}{8} Z_{t-k-2} \right) \right] \right)^2 \right) \\
&= \sum_{k \geq 0} \mathbb{E}(Z_{t-k}^2) \times \left(\frac{1}{2^k} - \frac{1}{4} \frac{1}{2^{k-1}} \mathbb{1}_{k \geq 1} - \frac{1}{8} \frac{1}{2^{k-2}} \mathbb{1}_{k \geq 2} \right)^2 \\
&< +\infty \text{ car série géométrique.}
\end{aligned}$$

Donc la somme qui nous intéresse est finie presque sûrement. Nous calculons alors

$$\begin{aligned}
&\sum_{k \geq 0} \left[\frac{1}{2^k} \left(X_{t-k} - \frac{7}{6} X_{t-k-1} + \frac{1}{3} X_{t-k-2} \right) \right] \\
&= X_t + X_{t-1} \left(-\frac{7}{6} + \frac{1}{2} \right) + \sum_{k \geq 2} \frac{X_{t-k}}{2^k} \left(1 - \frac{7}{6} \times 2 + \frac{1}{3} \times 2^2 \right) \\
&= X_t + X_{t-1} \times \frac{2}{3}.
\end{aligned}$$

Nous avons par ailleurs

$$\begin{aligned}
&\sum_{k \geq 0} \left[\frac{1}{2^k} \left(X_{t-k} - \frac{7}{6} X_{t-k-1} + \frac{1}{3} X_{t-k-2} \right) \right] \\
&= \sum_{k \geq 0} \frac{1}{2^k} \left(Z_t - \frac{Z_{t-1}}{4} - \frac{1}{8} Z_{t-2} \right) \\
&= Z_t + Z_{t-1} \left(-\frac{1}{4} + \frac{1}{2} \right) + \sum_{k \geq 2} \frac{Z_{t-k}}{2^k} \left(1 - \frac{1}{4} \times 2 - \frac{1}{8} \times 2^2 \right) \\
&= Z_t + Z_{t-1} \left(-\frac{1}{4} + \frac{1}{2} \right).
\end{aligned}$$

Ce qui finit la démonstration.

- (3) (a) Pour tracer le périodogramme : `k=kernel("daniell",4); spec.pgram(sncf,k,taper=0,log='no')`. On obtient la figure 6.5.1. Rappel : les abscisses vont de 0 à π (l'échelle affichée n'est pas correcte). Le pic le plus important se trouve donc en $\pi/6$, ce qui correspond à une période $2\pi/(\pi/6) = 12$ (ce qui est bien cohérent avec le graphique de la série et les autocorrélations
`(png(filename="blabla/tp8/5.1.png") ; par(mfrow=c(2,1)) ; plot(sncf) ; acf(sncf,lag.max=50) ; dev.off(), figure 6.5.2)`
- (b) Calculons les séries demandées : `y1=diff.ts(sncf,lag=1)` ; `y2=diff.ts(sncf,lag=12)` ; `y3=diff.ts(y1,lag=12)`.
- On commence par `y1` (figure 6.5.3). On se dit que Y_1 pourrait être un $AR(3)$, et donc que X pourrait être un $ARIMA(3,1,0)$. On estime les coefficients par : `out<-arima(sncf,order=c(3,1,0))`. On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid,lag=10)` renvoie une p -valeur de 0,002345 (le test calcule une statistique et la p -valeur est la probabilité d'observer une statistique aussi grande sous l'hypothèse « bruit blanc »). Comme la p -valeur est plus petite que 0,05, on rejette l'hypothèse « bruit blanc ». Notre processus ne peut pas être un $ARIMA(3,1,0)$.
 - On continue avec `y2` (figure 6.5.4). On se dit que Y_2 pourrait être un $AM(1)$ et donc que X pourrait être un $SARIMA(0,0,1,12)$. On estime les coefficients par `out<-arima(sncf,order=c(0,0,1),seasonal=list(order=c(0,0,1),period=12))`. On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid,lag=10)` renvoie une p -valeur de $2,2 \times 10^{-16}$,

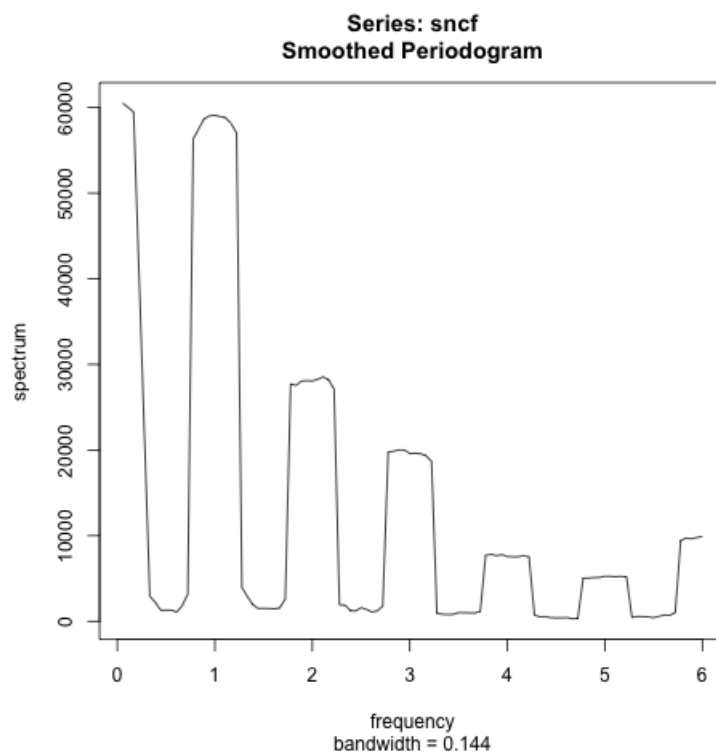


FIGURE 6.5.1. périodogramme

donc on rejette l'hypothèse « bruit blanc ». Notre processus ne peut pas être un $SARIMA(0, 0, 1, 12)$.

- On continue avec Y_3 (figure 6.5.5) On se dit que Y_3 pourrait être un $MA(2)$ et que donc X pourrait être un $SARIMA(0, 1, 2, 12)$. On estime les coefficients par `out<-arima(sncf,order=c(0,1,2),seasonal=list(order=c(0,1,2),period=12))`. On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid,lag=10)` renvoie une p -valeur de 0,4596 ($> 0,05$) donc on reste sur l'hypothèse « bruit blanc ». Donc notre processus X a bien l'air d'être un $SARIMA(0, 1, 2, 12)$.

- (c) On exécute une boucle

```
p=0
q=0
out<-arima(sncf,order=c(0,1,2),seasonal=list(order=c(0,1,2),period=12))
m=out$aic
for (i in 0:2)
{
  for (j in 0:2)
  {
    if (i+j>0)
    {
      out<-arima(sncf,order=c(i,1,j),seasonal=list(order=c(i,1,j),period=12))
      if (out$aic>m) { p=i q=j }
    }
  }
}
cat(p,q)
```

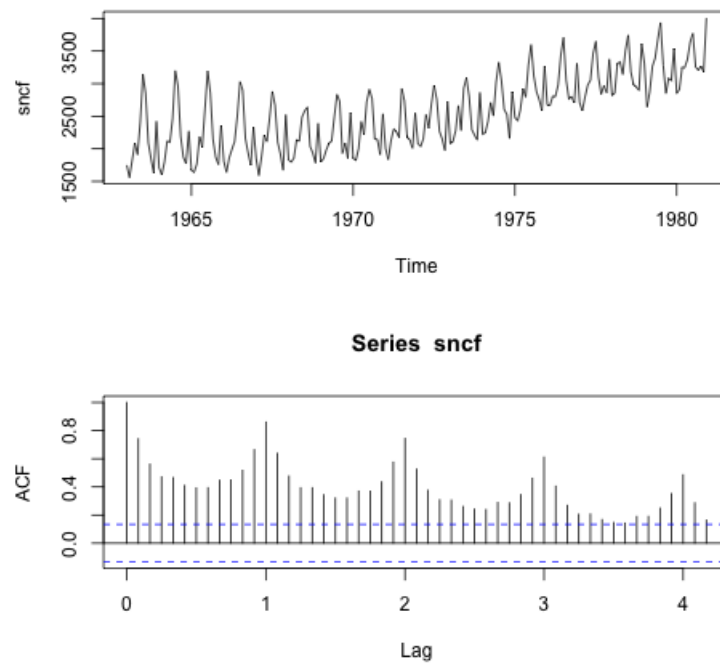


FIGURE 6.5.2. graphique et ACF

Le meilleur couple maximisant le critère AIC est $p = 2$, $q = 0$ (qui ne correspond pas aux graphiques ACF et PACF ci-dessus).

- (d) On exécute le code (graphique dans la figure 6.5.6)
- ```
out02<-arima(sncf,order=c(0,1,2),seasonal=list(order=c(0,1,2),period=12))
outpred<-predict(out02,n.ahead=12)
pred=outpred$pred
ts.plot(sncf,pred,col=c('black','red'))
```
- Pour avoir un graphique convaincant, il faudrait calculer des intervalles de confiance.

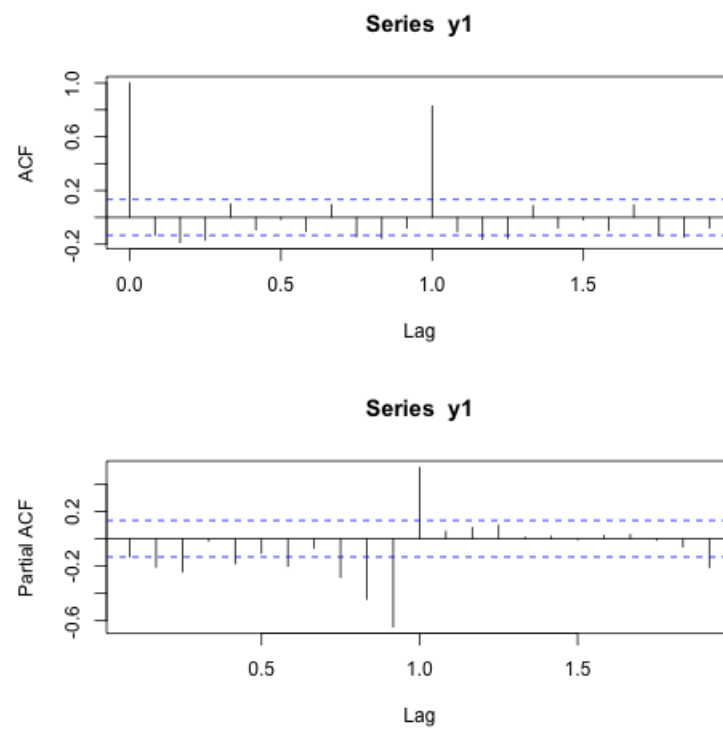


FIGURE 6.5.3.

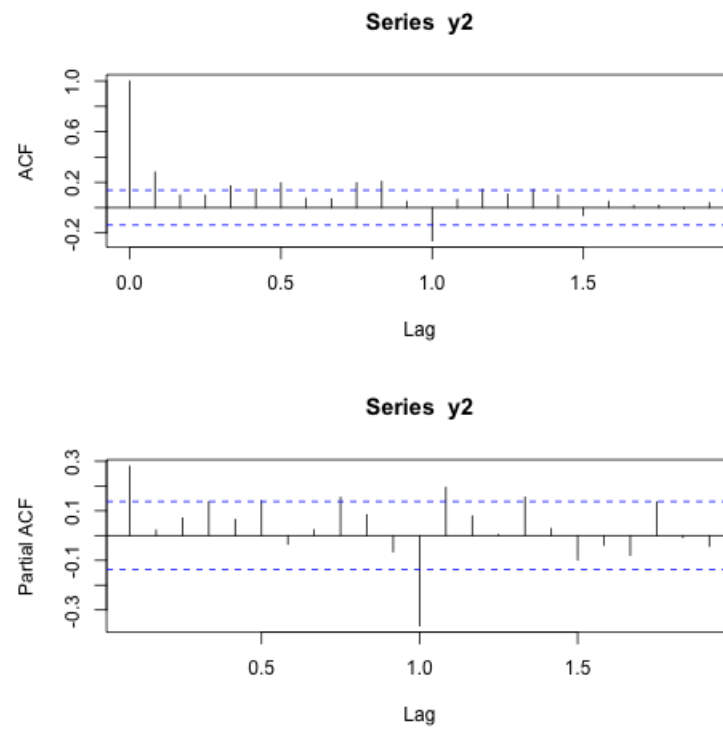


FIGURE 6.5.4.

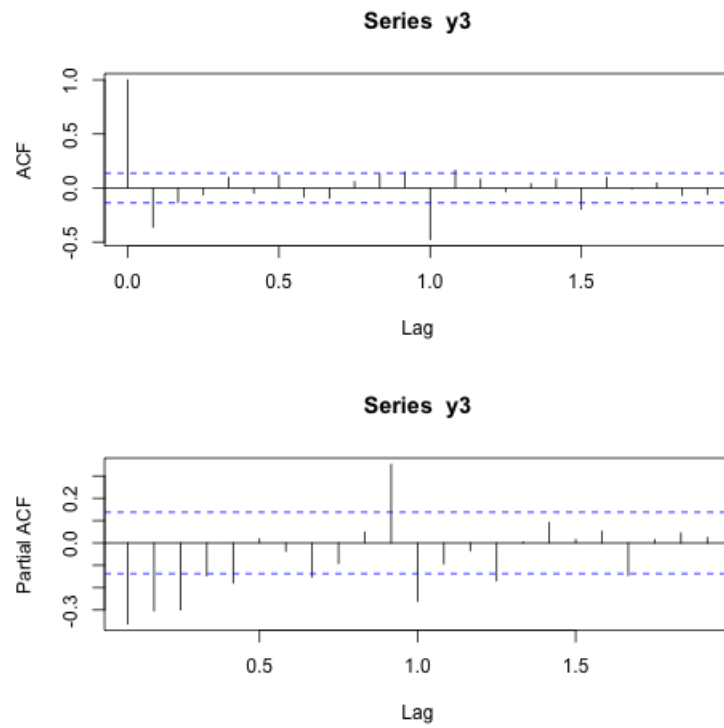


FIGURE 6.5.5.

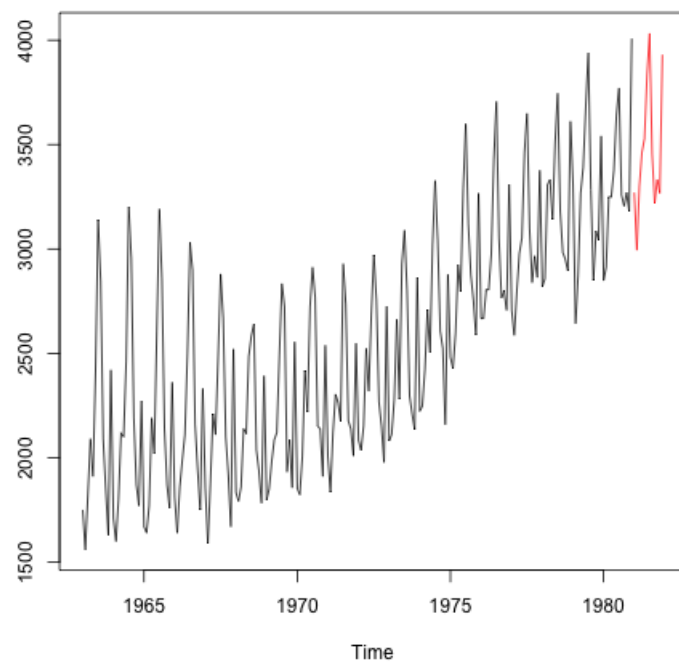


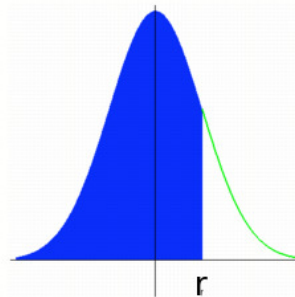
FIGURE 6.5.6.





## Table de la loi normale

$$P(X \leq r) \text{ avec } X \sim N(0,1)$$



|     | 0      | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0   | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1   | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2   | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |



## Bibliographie

- [Ara11] Yves ARAGON : *Séries temporelles avec R—méthodes et cas*. Pratique R. Springer, Paris, 2011. With a preface by Dominique Haughton.
- [BD02] Peter J. BROCKWELL et Richard A. DAVIS : *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer-Verlag, New York, second édition, 2002. With 1 CD-ROM (Windows).
- [Ber17] Florent BERTHELIN : *Équations différentielles*. Cassini, 2017.
- [BJR08] George E. P. BOX, Gwilym M. JENKINS et Gregory C. REINSEL : *Time series analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, fourth édition, 2008. Forecasting and control.
- [DDRT] J. J. DAUBIN, C. DUBY, S. ROBIN et P. TRÉCOURT : Analyse de séries chronologiques. Disponible à l'adresse <http://www.agroparistech.fr/IMG/pdf/Polychro.pdf>.
- [GM97] Christian GOURIEROUX et Alain MONTFORT : *Séries temporelles et modèles dynamiques*. second édition, 1997.
- [Gou08] Xavier GOURDON : *Les maths en tête : Analyse*. second édition, 2008.
- [Ham94] James D. HAMILTON : *Time series analysis*. Princeton University Press, Princeton, NJ, 1994.
- [HSD13] Morris W. HIRSCH, Stephen SMALE et Robert L. DEVANEY : *Differential equations, dynamical systems, and an introduction to chaos*. Elsevier/Academic Press, Amsterdam, third édition, 2013.
- [Jac] Julien JACQUES : Introduction aux séries temporelles. Disponible à l'adresse <http://eric.univ-lyon2.fr/~jjjacques/Download/Cours/ST-Cours.pdf>.
- [OPV] G. OPPENHEIM, A. PHILIPPE et M.-C. VIANO : Cours de séries temporelles. Disponible à l'adresse <http://www.math.sciences.univ-nantes.fr/~philippe/download/coursTime-seriesM2.pdf>.
- [RC10] Christian P. ROBERT et George CASELLA : *Introducing Monte Carlo methods with R. Use R!* Springer, New York, 2010.
- [SS11] Robert H. SHUMWAY et David S. STOFFER : *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, third édition, 2011. With R examples.
- [WH99] Beverly WEST et John HUBBARD : *Équations différentielles et systèmes dynamiques*. Cassini, 1999.



## Liste des symboles

|                         |                             |
|-------------------------|-----------------------------|
| $\mathbb{C}$            | Nombres complexes           |
| Cov                     | Covariance                  |
| $\mathbb{E}$            | Espérance                   |
| $\mathbb{N}$            | Entiers naturels            |
| $\mathbb{P}$            | Probabilité                 |
| $\Re$                   | Partie réelle               |
| $\mathbb{R}$            | Nombres réels               |
| Var                     | Variance                    |
| $\mathbb{Z}$            | Entiers relatifs            |
| $\lfloor \dots \rfloor$ | Partie entière (inférieure) |



# Index

## A

AIC, 43  
ARIMA, 44  
Auto-corrélation empirique, 2  
Auto-corrélation partielle, 35  
Auto-covariance empirique, 2

## B

BIC, 43  
Bruit blanc, 23  
Bruit d'innovation, 35

## C

Chi-deux, 29

## D

Densité spectrale, 63, 64

## F

Fonction d'auto-corrélation empirique, 2  
Fonction d'auto-covariance, 23  
Fonction d'auto-covariance empirique, 2  
Fréquences fondamentales de Fourier, 66

## G

GARCH, 73

## H

Hétéroscédastique, 72  
Homoscédastique, 72

## I

i.i.d., 2  
Innovations, 35  
Intervalle de confiance, 44

## K

Kurtosis, 73

## L

Lissage exponentiel double, 16  
Lissage exponentiel simple, 15

## M

Maximum de vraisemblance, 43  
Moindres carrés, 23  
Moyenne empirique, 2  
Moyenne mobile, 26

## P

Périodogramme, 66  
Prévisions météo, 16  
Processus auto-régressif, 35

Processus auto-régressif en moyenne mobile, 39  
Processus en moyenne mobile, 39  
Processus stationnaire, 23

## S

SARIMA, 44  
Skewness, 73  
Statistique, 29

## T

Tendance, 2  
Test d'hypothèses, 29

## V

Variance empirique, 2