

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Dr Moulay Tahar
Faculté des Sciences
Département de Mathématiques
Saïda

Séries chronologiques

Cours, exercices et travaux pratiques

Fatiha Mokhtari ¹

1. e-mail : fatiha.mokhtari@univ-saida.dz

Table des matières

1	Introduction	6
1.1	Définition	8
1.2	Exemples	8
1.3	Domaines d'application	10
1.4	Objectifs principaux	11
2	Description d'une série chronologique	12
2.1	Composantes fondamentales	12
2.1.1	La tendance ou trend	12
2.1.2	La composante saisonnière ou saisonnalité	12
2.1.3	La composante résiduelle	13
2.1.4	Les phénomènes accidentels	13
2.2	Modélisation	13
2.2.1	les modèles déterministes.	13
2.2.2	les modèles stochastiques	14
2.3	Modélisation déterministe	14
2.3.1	Le modèle additif	14
2.3.2	Le modèle multiplicatif	15
2.3.3	Le modèle mixte	16
2.4	Choix du modèle	16
2.4.1	Méthode de profile	16
2.4.2	Méthode de la bande	17
2.4.3	Méthode analytique	18
2.5	Mise en œuvre sous R	19
2.6	Exercices et feuille de travaux pratiques 1	20
3	Analyse de la tendance	29
3.1	Rappels sur la régression linéaire	29
3.1.1	La méthode des moindres carrés	29
3.2	Ajustement tendanciel linéaire	30

3.3	Ajustement tendanciel linéaire par points médians	31
3.4	Ajustements tendanciels non linéaires	31
3.4.1	Exemple	32
3.5	Lissage de la tendance par moyenne mobile	32
3.5.1	Définition	32
3.5.2	Notation	33
3.5.3	Cas particulier : Moyenne mobile arithmétique.	33
3.5.4	Moyenne mobile d'ordre pair	34
3.6	Propriétés d'un lissage par moyenne mobile	35
3.6.1	Effet d'une moyenne mobile sur une tendance	35
3.6.2	Effet d'une moyenne mobile sur une composante saisonnière	35
3.6.3	Effet d'une moyenne mobile sur les fluctuations irrégulières	35
3.6.4	Choix pratique de l'ordre d'une moyenne mobile	35
3.7	Mise en œuvre sous R	36
3.8	Exercices et feuille de travaux pratiques 2	37
4	Calcul des variations saisonnières et accidentelles.	41
4.1	Calcul des données sans tendance	41
4.2	Calcul des coefficients saisonniers S_j	41
4.2.1	Exemple	42
4.3	Série désaisonnalisée ou série CVS	43
4.3.1	Définition	43
4.3.2	Intérêts	43
4.4	La série ajustée	44
4.4.1	Ce qu'elle représente	44
4.5	Variations accidentelles ou résiduelles	44
4.6	Description schématique de l'étude complète d'une série chronologique	44
4.7	Exercices et feuille de travaux pratiques 3	45
5	Lissages exponentiels	54
5.1	lissage exponentiel simple (LS)	54
5.1.1	Définition	55
5.1.2	Formule de mise à jour	56
5.1.3	Choix du α	56
5.2	Lissage exponentiel double	57
5.3	Méthode de Holt-Winters	61
5.3.1	Méthode non saisonnière	61
5.3.2	Méthode saisonnière additive	62
5.3.3	Méthode saisonnière multiplicative	63

5.4	Mise en œuvre sous R	63
5.5	Exercices et feuille de travaux pratiques 4	64
6	Appendice	72
6.1	Indices descriptifs d'une série temporelle	72
6.1.1	Indice de tendance centrale	72
6.1.2	Indices de dispersion	72
6.1.3	Indices de dépendance	72
6.2	Commandes usuelles de R	73
6.2.1	Création de données	73
6.2.2	Extraction de données et fonctions de base	74
6.2.3	Importer et exporter des données	76
6.2.4	Variables et attributs	77
6.2.5	Graphiques	77
6.2.6	Modèles et analyses statistiques	78
6.2.7	Programmation	78
	Bibliographie	79

Chapitre 1

Introduction

Le monde tel qu'on le connaît est rempli de phénomènes dépendant du temps. Une série chronologique est une suite formée d'observations au cours du temps. L'analyse des séries chronologiques est un outil couramment utilisé de nos jours. Ce domaine possède beaucoup d'applications en finance, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. Par exemple, en finance, on s'intéresse à modéliser le taux de change d'une devise. En météorologie, les scientifiques modélisent par exemple la température maximale d'Alger dans la dernière année pour prédire la température maximale qu'il fera le mois prochain.

L'idée est de prendre un échantillon de données et de construire le meilleur modèle qui ajuste ces données. Ce modèle nous permet de tirer certaines conclusions sur la série. Par exemple, on peut établir une formule pour la prédiction de données, détecter certains pics ou modéliser la tendance (orientation) de la série. Un autre aspect important de la série est la composante saisonnière, c'est-à-dire la présence de cycles. Un autre concept intéressant serait le phénomène de causalité, c'est-à-dire l'influence d'une série sur une autre.

L'un des objectifs principaux de l'étude d'une série chronologique est la prévision, cela très souvent pour des raisons économiques (prévoir l'évolution des ventes d'un certain produit, prévoir la consommation d'électricité pour ajuster au mieux la production). Les prévisions sont effectuées en s'appuyant sur ce qui s'est passé avant l'instant à partir duquel démarre la prévision. Et parfois en utilisant des renseignements annexes. La particularité des séries chronologiques tient la présence d'une relation d'antériorité entre les temps, qui ordonne l'ensemble des informations. Dans la prévision des ventes par exemple pour l'année $T + 1$ on s'appuiera sur l'évolution des ventes durant les années $T, T - 1, \dots$, mais on tiendra compte aussi, éventuellement, d'indices exogènes (la conjoncture économique, certains événements récents inatten-

dus, relation température, vitesse de vent et niveau d'humidité etc ...).

Cependant, il n'est pas toujours évident de choisir le bon modèle, car, normalement, plusieurs peuvent être de bons candidats. La plupart de ceux-ci auront tendance à minimiser la variance des résidus. Il faut aussi considérer le nombre de paramètres à estimer. Normalement, on utilise le principe de parcimonie qui nous propose de choisir le modèle demandant le moins de paramètres avec une variance faible des résidus. Il y a donc plusieurs critères à considérer dans la création d'un modèle.

Il y a toujours une erreur de prévision, et les bonnes méthodes fournissent non pas une prévision, mais un intervalle de prévision. Il arrive souvent qu'une amélioration minime de la qualité de la prévision ait une grosse incidence sur les coûts.

Parmi les méthodes de prévision les plus utilisées on a :

- Vous connaissez la plus rudimentaire, qui consiste à faire une régression et à se baser sur elle pour prévoir. On estime les coefficients par une méthode de régression. On valide le résultat, puis on effectue des prévisions.
- Les lissages exponentiels sont une batterie de procédures extrêmement simples à mettre en oeuvre. Ils sont exposés au chapitre 5.
- Les méthodes de prévision les plus populaires depuis une trentaine d'années sont celles liées à la modélisation ARMA qui n'est pas malheureusement l'objet de ce travail. Elles consistent en gros à dégager les tendances évidentes dans le phénomène qu'on observe (croissance, périodicités etc. . .) et à se concentrer sur ce qui reste lorsqu'on les a supprimées. On procède à une modélisation du résidu obtenu et on s'appuie sur cette modélisation pour obtenir la prévision. Ces méthodes, plus sophistiquées mais plus performantes que la régression et que les lissages exponentiels, sont coûteuses en temps de calcul. Leur utilisation systématique dans tous les domaines, que ce soit économique, industriel, physique etc . . ., est due au développement exponentiellement rapide de l'outil informatique. Tout l'art du statisticien, face à des données difficiles, est de confronter plusieurs méthodes et de choisir celle qui convient le mieux.

Ce cours est structuré de la façon suivante : Dans le premier chapitre, on donne les définitions nécessaires à la compréhension du corps. Ensuite, le deuxième, le troisième et le quatrième chapitre présentent une description détaillée des différentes composantes, de la relation qui les relie et de leurs approximations. Le lissage exponentielle

est abordé dans le chapitre cinq. L'appendice est essentiellement consacré à donner quelques fonctions de base du langage R.

1.1 Définition

Une suite d'observations d'une famille de variables aléatoires réelles $(X_t)_{t \in \Theta}$ indexée par le temps est appelée série chronologique (ou temporelle). où l'ensemble Θ est appelé espace de temps qui peut être :

- **discret** (nombre d'étudiants inscrits à l'université, température maximale ...). Dans ce cas, $\Theta \subset \mathbb{Z}$. Les dates d'observations sont le plus souvent équidistantes : par exemple relevés mensuels, trimestriels ... Ces dates équidistantes sont alors indexées par des entiers : $t = 1, 2, \dots, T$ et T est le nombre d'observations. On dispose donc des observations des variables X_1, X_2, \dots, X_T . Ainsi si h est l'intervalle de temps séparant deux observations et t_1 l'instant de la première observation, on a le schéma suivant :

$$\begin{array}{ccccccc} t_1 & t_1 + h & \dots & t_1 + (T-1)h & & & \\ X_{t_1} & X_{t_1+h} & \dots & X_{t_1+(T-1)h} & & & \\ & X_1 & X_2 & \dots & X_T & & \end{array}$$

- **continu** (signal radio, résultat d'un électrocardiogramme ...). L'indice de temps est à valeurs dans un intervalle de \mathbb{R} et on dispose (au moins potentiellement) d'une infinité d'observations issues d'un processus $(X_t)_{t \in \Theta}$ où Θ est un intervalle de \mathbb{R} . Un tel processus est dit à temps continu.

1.2 Exemples

Le tableau suivant et la Figure 1.1 représente l'évolution du nombre mensuel des passagers aériens par mois entre 1949 et 1960.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

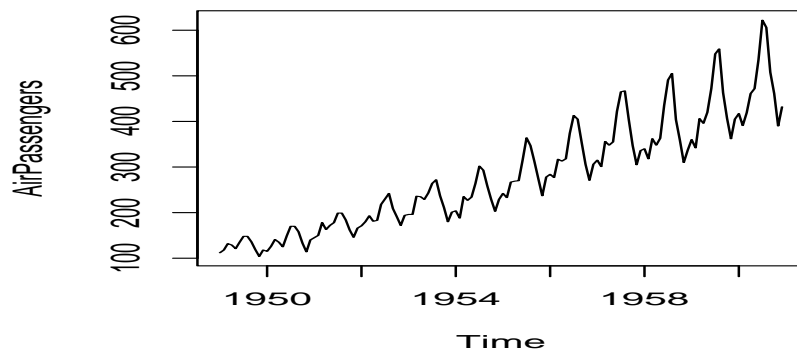


Figure 1.1 La série AirPassengers

Figure 1.2 et 1.3 représentent respectivement la température mensuelle moyenne à Nottingham entre 1920-1939 et la concentration atmosphérique de CO_2 à Maunaloa.

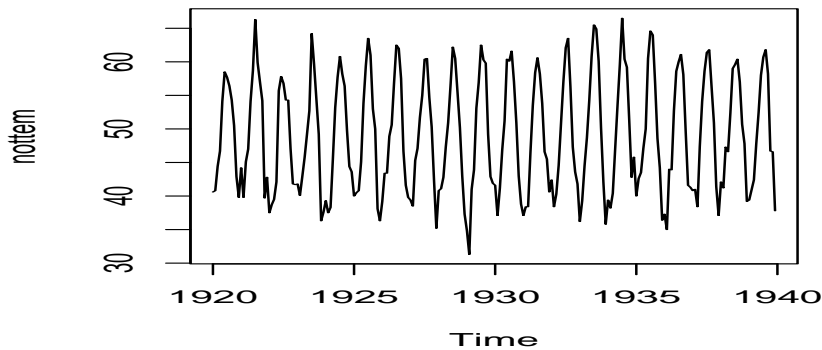


Figure 1.2 la température à Nottingham

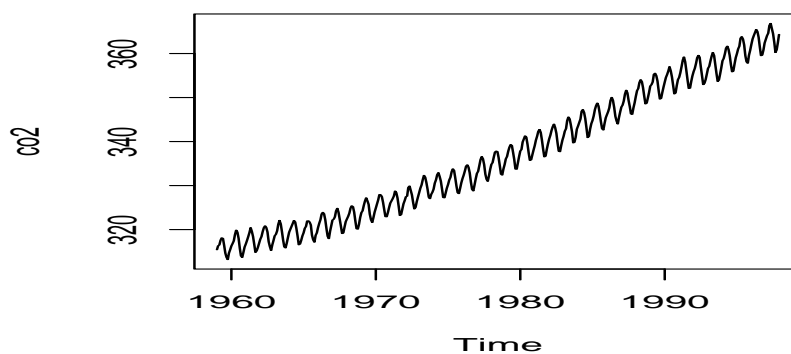


Figure 1.3 Concentration du CO_2

1.3 Domaines d'application

On trouve des exemples de séries chronologiques univariées dans de très nombreux domaines. La liste suivante n'est qu'un échantillon :

- finance et économétrie : évolution des indices boursiers, des prix, des données économiques des entreprises, des ventes et achats de biens, des productions agricoles ou industrielles,
- assurance : analyse des sinistres,
- médecine / biologie : suivi des évolutions des pathologies, analyse d'électroencéphalogrammes et d'électrocardiogrammes,
- sciences de la terre et de l'espace : indices de marées, variations des phénomènes physiques (météorologie), évolution des taches solaires, phénomènes

- d'avalanches,
- traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole,
- traitement des données : mesures successives de position ou de direction d'un objet mobile (trajectographie).

1.4 Objectifs principaux

L'étude d'une série chronologique permet **d'analyser, de décrire et d'expliquer** un phénomène au cours du temps et d'en tirer des conséquences pour des prises de décision (marketing, ...).

Cette étude permet aussi de faire un **contrôle**, par exemple pour la gestion des stocks, le contrôle d'un processus chimique.

Mais l'un des objectifs principaux de l'étude d'une série chronologique est **la prévision** qui consiste à prévoir les valeurs futures X_{T+h} ($h = 1, 2, 3, \dots$) de la série à partir de ses valeurs observées jusqu'au temps T : X_1, X_2, \dots, X_T .

Il existe encore bien d'autres objectifs immédiats relatifs à l'étude des séries chronologiques par exemple :

si deux séries sont observées, on peut se demander quelle **influence elles exercent l'une sur l'autre**. En notant X_t et Y_t les deux séries en question, on examine s'il existe par exemple des relations du type

$$Y_t = a_1 X_{t-1} + a_3 X_{t-3}.$$

Ici, deux questions se posent : tout d'abord, la question de **la causalité** i.e. quelle variable (ici (X_t)) va expliquer l'autre (ici (Y_t)), ce qui amène à la deuxième question, celle **du décalage temporel** : si une influence de (X_t) sur (Y_t) existe, avec quel délai et pendant combien de temps la variable explicative (X_t) influence-t-elle la variable expliquée (Y_t) ?

Un dernier problème important de la macroéconométrie est de déterminer les relations persistances (de long terme) des autres relations (de court terme).

Chapitre 2

Description d'une série chronologique

On considère qu'une série chronologique X_t est la résultante de différentes composantes fondamentales.

2.1 Composantes fondamentales

2.1.1 La tendance ou trend

(C_t) représente l'évolution à long terme de la série étudiée. Elle traduit le comportement "moyen" de la série. Elle est le plus souvent modélisée par une fonction linéaire ou polynomiale du temps. Par exemple, la série de la Figure 1.1 a tendance à augmenter de façon linéaire et indique que le nombre de voyageurs a augmenté de manière régulière au cours du temps.

2.1.2 La composante saisonnière ou saisonnalité

S_t correspond à un phénomène qui se répète à intervalles de temps réguliers (périodiques). En général, c'est un phénomène saisonnier d'où le terme de variations saisonnières. Les variations saisonnières sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz, vente des crèmes solaires, ...) ou encore par des activités économiques et sociales (fêtes, vacances, soldes, ...).

C'est un phénomène qui se reproduit de manière analogue sur chaque intervalle de temps successif. Lorsqu'on veut mettre en évidence ce phénomène à l'aide d'un graphique, on peut découper la série en sous-séries de longueur de période p du saisonnier et représenter ces sous-séries sur un même graphique. Dans la suite, la plupart du temps, on suppose que la composante saisonnière est constante sur chaque période P , c'est-à-dire $S_{t+P} = S_t$. Lorsque $P = 4$, la série est trimestrielle; lorsque $P = 12$, la série

est mensuelle. Sur les graphiques Figure 1.1 et Figure 1.3, on voit bien une similarité des différentes courbes annuelles liée aux saisons météorologiques.

2.1.3 La composante résiduelle

. (ou bruit ou résidu) (ε_t)

Correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. Ces variables aléatoires sont centrées. On considère le plus souvent un bruit blanc, c'est-à-dire une suite de variables aléatoires telles que (ε_t), d'espérance nulle et de variance constante c'est-à-dire $\mathbb{E}(\varepsilon_t)$ et $Var(\varepsilon_t)$ ne dépendent pas de t et tel que $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0$ pour $h \neq 0$. Si les variables sont gaussiennes, le bruit blanc est dit gaussien et on a dans ce cas de plus l'indépendance des ε_t .

Par exemple, la série de la Figure 1.2 a un comportement assez irrégulier, il y a comme une sorte de bruit de faible amplitude qui perturbe les données.

Les modèles présentés dans la suite tiennent compte de ces trois composantes (tendance, saisonnalité et fluctuations irrégulières). Il faut cependant remarquer que l'on pourrait envisager d'autres composantes.

2.1.4 Les phénomènes accidentels

sont des valeurs isolées anormalement élevées ou faibles de courte durée. Ces variations brusques de la série sont généralement explicables (Février 2022, crise entre la Russie et l'Ukraine, Mai 68, réunification de l'Allemagne, conditions météorologiques exceptionnelles, grèves, crash financier. . .). La plupart du temps, ces accidents sont intégrés dans la série des bruits.

2.2 Modélisation

Un modèle est une image simplifiée de la réalité qui vise à traduire les mécanismes de fonctionnement du phénomène étudié. On distingue principalement deux types de modèles :

2.2.1 les modèles déterministes.

Ils sont de la forme :

$$X_t = f(t, \varepsilon_t).$$

Les deux modèles de ce type les plus usités sont les suivants :

le modèle additif

X_t s'écrit comme la somme de trois termes :

$$X_t = C_t + S_t + \varepsilon_t.$$

le modèle multiplicatif

La variable X_t s'écrit

$$X_t = C_t S_t \varepsilon_t.$$

2.2.2 les modèles stochastiques

Ils sont du même type que les modèles déterministes à ceci près que les variables de bruit ε_t ne sont pas iid mais possèdent une structure de corrélation non nulle : ε_t est une fonction des valeurs passées (lointaines suivant le modèle) et d'un terme d'erreur η_t ,

$$\varepsilon_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \eta_t).$$

La classe des modèles de ce type la plus fréquemment utilisée est la classes des modèles SARIMA (et de ses sous-modèles ARIMA, ARMA, ...).

Les deux types de modèles ci-dessus induisent des techniques de prévision bien particulières.

2.3 Modélisation déterministe

2.3.1 Le modèle additif

$$X_t = C_t + S_t + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Dans ce modèle, l'amplitude de la série reste constante au cours du temps. Cela se traduit graphiquement par des fluctuations autour de la tendance C_t constantes au bruit près (cf Figure 2.1).

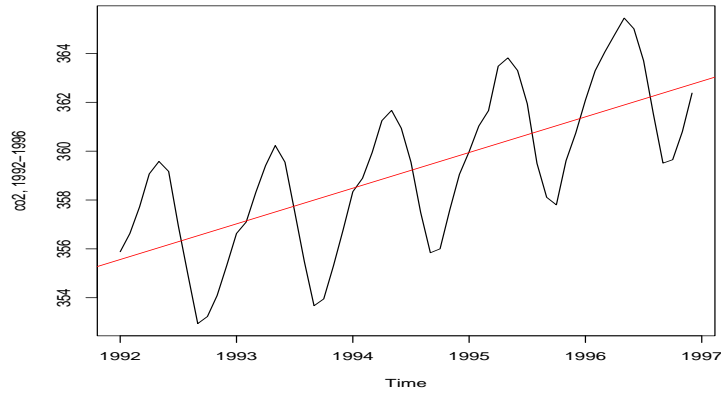


Figure 2.1

Remarque 2.3.1. À première vue, la notion de composante périodique pourrait être suffisante. Cependant, ce n'est pas le cas pour la raison décrite ci-après. Considérons le modèle additif

$$X_t = C_t + S_t + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Cette décomposition n'est pas unique en l'absence d'hypothèses supplémentaires. En effet, si S_t est une composante périodique de période p , alors il existe une constante c telle que

$$S_{t+1} + \dots + S_{t+p} = c, \quad t \in \mathbb{Z}.$$

Dans ce cas, on a aussi la décomposition suivante

$$X_t = \hat{C}_t + \hat{S}_t + \varepsilon_t, \quad t \in \mathbb{Z},$$

où pour tout $t \in \mathbb{Z}$, $\hat{C}_t = C_t + c/p$ et $\hat{S}_t = S_t - c/p$. On a donc trouvé une autre décomposition du signal X_t . On remarquera que \hat{S}_t est une composante de somme nulle sur la période p . C'est pourquoi on impose à toute composante saisonnière d'être périodique et de somme nulle sur une période ($\frac{1}{p} \sum_{i=1}^p S_i = 0$).

2.3.2 Le modèle multiplicatif

Nous considérons dans cette section une série $X = (X_t)_t$ admettant une décomposition multiplicative

$$X_t = C_t S_t \varepsilon_t, \quad t = 1 \dots T,$$

Dans ce cas l'amplitude de la série n'est plus constante au cours du temps : elle varie au cours du temps proportionnellement à la tendance C_t au bruit près (cf Figure 2.2).

Dans ce modèle, on suppose que $\frac{1}{p} \sum_{i=1}^p S_i = 1$.

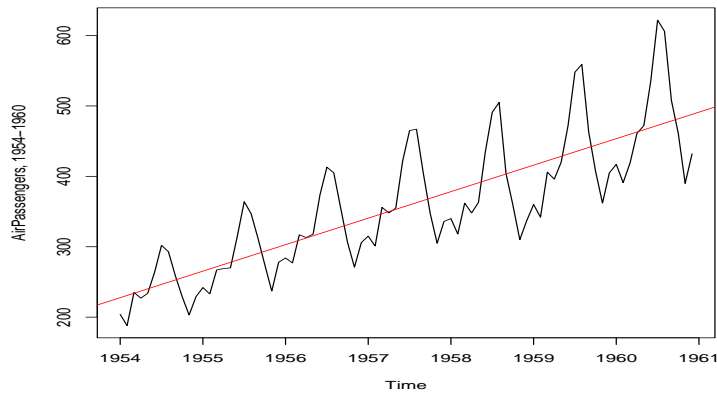


Figure 2.2

Remarque 2.3.2.1. *Le modèle multiplicatif est généralement utilisé pour des données de type économique.*

2.3.3 Le modèle mixte

Il s'agit de modèles où addition et multiplication sont utilisées. On peut supposer par exemple que la composante saisonnière agit de façon multiplicative alors que les fluctuations irrégulières sont additives :

$$X_t = C_t S_t + \varepsilon_t, \quad t = 1 \dots T,$$

2.4 Choix du modèle

Avant toute modélisation et étude approfondie du modèle, on tente d'abord de déterminer si on est en présence d'un modèle multiplicatif ou additif.

2.4.1 Méthode de profil

Pour faire la détermination entre modèle additif et modèle multiplicatif graphiquement, on peut par exemple superposer les saisons représentées par des courbes de profil sur un même graphique. Si ces courbes sont parallèles, le modèle est additif, autrement le modèle est multiplicatif (cf Figure 2.3).

Sur le graphique de notre exemple Figure 2.4, les courbes de profil semblent parallèles pour toutes les saisons. On peut donc supposer que le modèle est additif.

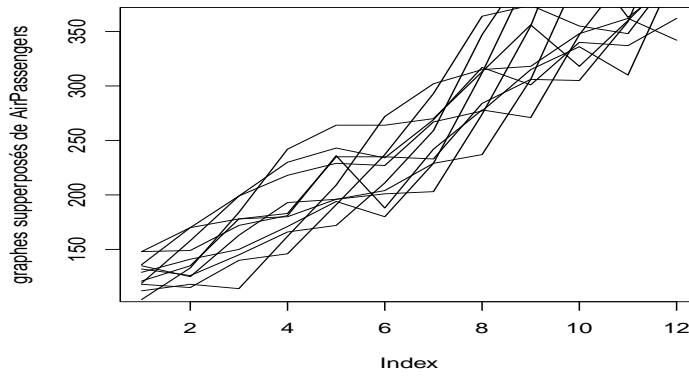


Figure 2.3

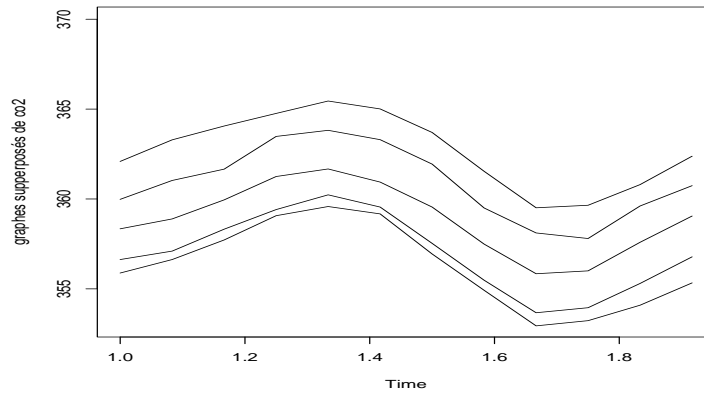


Figure 2.4

2.4.2 Méthode de la bande

On fait un graphique représentant la série chronologique, puis on trace une droite passant respectivement par les minima et par les maxima de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif (cf Figure 2.5). Dans le cas contraire, c'est un modèle multiplicatif (cf Figure 2.6)

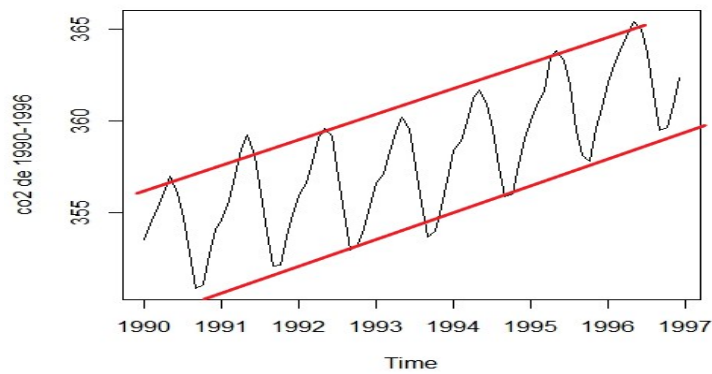


Figure 2.5

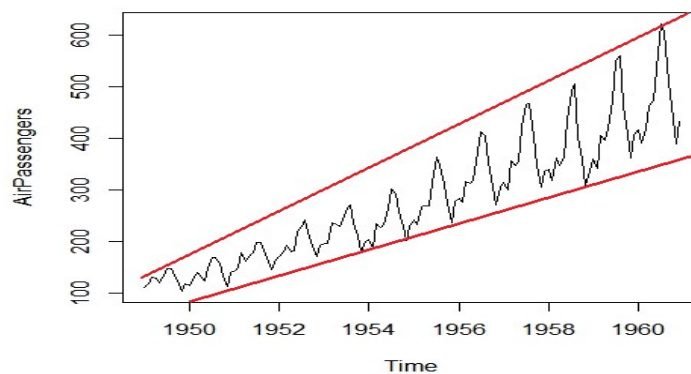


Figure 2.6

2.4.3 Méthode analytique

On calcule les moyennes et les écarts-types pour chacune des périodes considérées puis la droite des moindres carrés $\sigma = a\bar{X} + b$. Si a est nul, c'est un modèle additif, sinon c'est un modèle multiplicatif.

Exemple 2.4.1. *Le tableau 1 et la Figure 2.7 représente série des nouvelles immatriculations de voitures particulières, commerciales et utilitaires neuves selon le mois.*

Pour déterminer la nature du modèle, on calcule la droite des moindres carrés (cf tableau 2.4.2), on obtient $a = 0.195$ et $b = 289.037$, ce qui confirme que pour cet exemple, nous sommes bien en présence d'un modèle multiplicatif.

tableau 2.4.1. Série des nouvelles immatriculations de voitures.

Année	Jan	Fév	Mars	Avril	Mai	Juin	Juil	Aout	Sept	Oct	Nov	Déc
1996	2006	3224	3789	4153	3100	2527	3015	1504	1847	2314	1673	1602
1997	2247	3862	3586	4047	2838	2727	2730	1648	2007	2450	1966	1695
1998	2433	3723	4325	4493	3399	3083	3247	1928	2377	2831	2388	2126
1999	3127	4434	5478	4384	3552	3678	3611	2260	2699	3071	2510	2182
2000	3016	4871	5218	4746	4814	3545	3341	2439	2637	3085	2737	2055

tableau 2.4.2. Les moyennes et les écarts-types.

Année	Moyenne	Ecart-type
1996	2562.8	850.7
1997	2650.3	782.3
1998	3029.4	803.6
1999	3415.8	946.6
2000	3525.3	1023.4

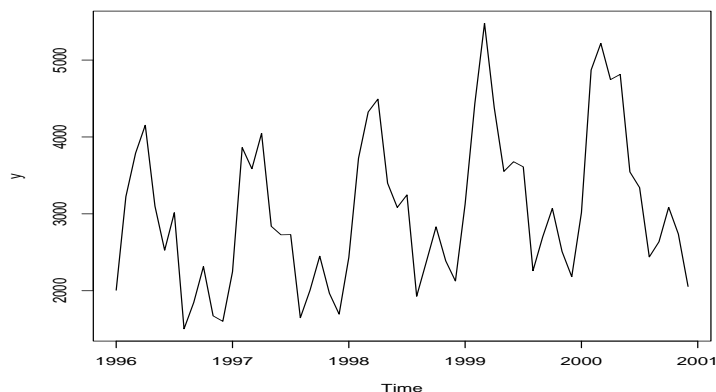


Figure 2.7

2.5 Mise en œuvre sous R

R comporte beaucoup d'objets, parmi ces objets on trouve par exemple les séries chronologiques. Pour définir une série chronologique (univariée) en R, on utilise la fonction `ts` :

`ts (data = NA, start = 1, end = numeric(0), frequency = 1).`

Quatre objets sont nécessaires :

- L'objet `data` contenant les données proprement dites. Ces données peuvent être par exemple mises sous la forme d'un vecteur. Cet objet est obligatoire pour définir une série chronologique.
- L'objet `start`, qui indique la date à laquelle commence la série. Si on a une série mensuelle, `start` peut par exemple être égal Février 1990 (`c(1990,2)`).
- L'objet `frequency` qui spécifie le nombre de réalisations observées sur une unité de temps. Par exemple, dans le cas de données mensuelles prélevées sur plusieurs années, l'unité de temps est l'année et la fréquence des observations est de 12 par an.
- L'objet `end` qui détermine la date de la dernière réalisation du processus.

Nous aurons aussi besoin de :

- La commande `par(mfrow=c(n,p))` permet d'afficher $n \times p$ graphique sur la même page.
- Découper une série temporelle de `c(i,j)` à `c(k,l)`, `xd<-window(x,c(i,j),c(k,l))`
- Obtenir un vecteur aléatoire de longueur n de composantes des variables i.i.d. $\mathcal{N}(0,1)$: `e=rnorm(n)`.
- La fonction d'autocorrélation, autocorrélation temporelle de la série pour un décalage temporel h mesure le degré de linéarité de la relation qui lie les observations aux temps t et $t-h$. On utilise pour cela :
`acf(maserie,lag.max = 20, plot = FALSE ,type="correlograme")`.

2.6 Exercices et feuille de travaux pratiques 1

Exercice 1. Soit les séries suivantes :

- $y_1(t) = t \times S_t \times \varepsilon_t$;
- $y_2(t) = 0.4t + S_t + \varepsilon_t$;
- $y_3(t) = \varepsilon_t$;
- $y_4(t) = S_t + \varepsilon_t$;
- $y_5(t) = \cos(t/30) + \varepsilon_t$;
- $y_6(t) = \cos(t/30) \times \varepsilon_t$;

où ε_t est un bruit aléatoire gaussien centré de variance 1, S_t est une série d'effets saisonniers de période 12 dont les 12 premiers termes sont tirés de façon indépendante dans la loi uniforme sur $[0,6]$.

- Faire correspondre chaque série à son graphique.

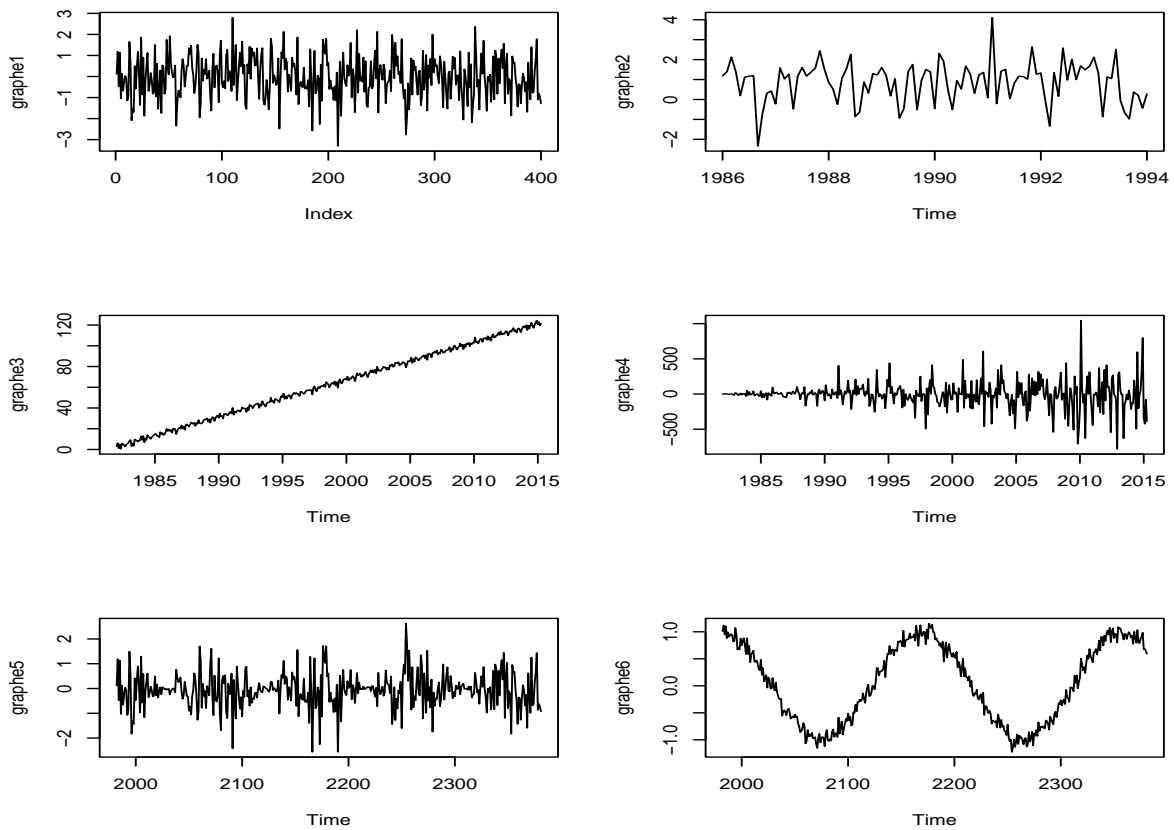


Figure 2.8

correction de l'exercice 1 On a : graphe1 \rightarrow y_3 , graphe2 \rightarrow y_4 , graphe3 \rightarrow y_2 , graphe4 \rightarrow y_1 , graphe5 \rightarrow y_6 , graphe6 \rightarrow y_5 .

Feuille de travaux pratiques 1

Exercice 2. 1. Simuler un Bruit blanc et tracer son graphe.

2. Supposons maintenant que l'on désire créer un objet série chronologique qui contient les données :

$$\text{simplevector} = c(1,2,3,4,5,6,7,2,3,4,5,6,7,8,3,4,5,6,7,8, \\ 9,4,5,6,7,8,9,10,5,6,7,8,9,10,11,6,7,8,9,10,11,12)$$

et qui, par construction, comporte six périodes de longueur sept. La série chronologique que nous allons constituer est donc telle que : $\text{frequency} = 7$. Imaginons que cette série représente l'évolution d'une variable jour après jour, ça commence, disons, le mercredi de la cinquième semaine.

3. Créer une série chronologique mensuelle contenant les données précédente commençant le mois de mars de la 5 ème années et tracer son corrélogramme et son corrélogramme partielle.

Correction de l'exercice 2.

```
1 #1)
2 a = ts(rnorm(100,mean = 0,sd =1),1980,f = 4)
3 plot(a)
```

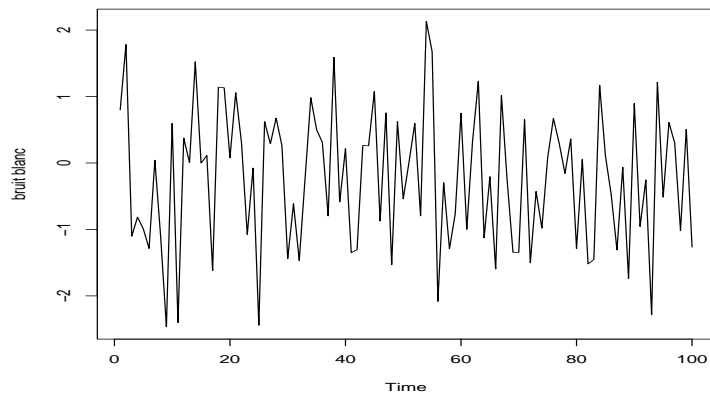


Figure 2.9

```
1 #2)
2 maserie = ts(simplevector , start = 5+(3/7), frequency = 7, end= 11+(1/7))
3 #start = 5 + (3/7), pour spécifier la date de la première réalisation de
4 #la série (avec 0/7 #=Dimanche, #1/7 = Lundi, etc.). Nous n'avons donc
5 plus le choix pour spécifier le dernier #objet, qui doit être#:end = 11
6 + #(1/7)
7
8 #3)
9 maserie = ts(simplevector , start = 5+(2/12), frequency = 12)
10 maserie = ts(simplevector , start =c(5,2), frequency = 12)
11 tsp(maserie)
12
13 acf ( maserie , lag.max = 50, plot = TRUE , main = "Fonction d'autocorré
14 lation",xlab = "Decalage temporel")
15 acf ( maserie , lag.max = 50, plot = FALSE ,type="corrélograme")
16 acf(maserie) #pour tracer le corrélogramme
17 pacf(maserie) #pour tracer le corrélogramme partiel
```

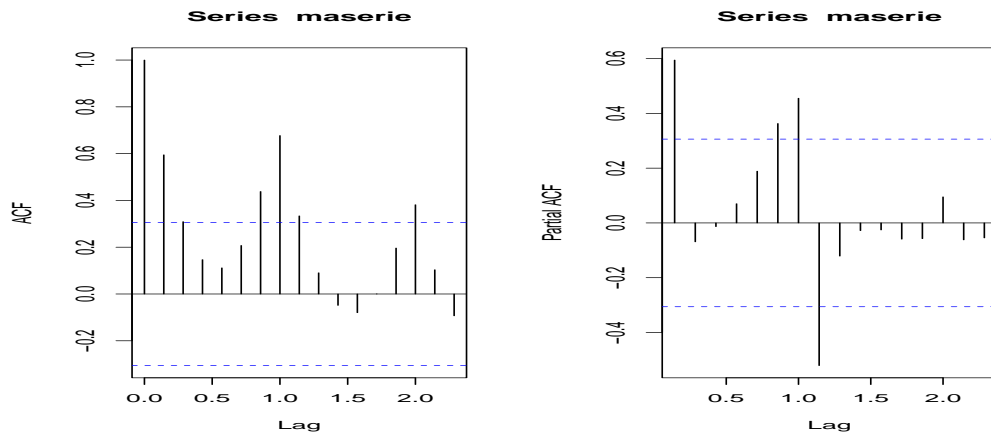


Figure 2.10

Exercice 3. Simuler les séries temporelles suivantes de taille 1000 :

- $X_1(t) = \varepsilon_t$;
- $X_2(t) = 0.4t + 2\varepsilon_t$;
- $X_3(t) = S_t + \varepsilon_t$;
- $X_4(t) = 0.4t + S_t + \varepsilon_t$;
- $X_5(t) = tS_t\varepsilon_t$;
- $X_6(t) = \cos(t/30)\varepsilon_t$;
- $X_7(t) = \cos(t/30) + \varepsilon_t$;
- $X_8(t) = 0.4X_8(t-1) + \varepsilon_t$; ($y_0 = 0$);
- $X_9(t) = \varepsilon(t) + 0.7\varepsilon_{t-1}$;
- $X_{10}(t) = 0.6X_{10}(t-1) + 0.4X_{10}(t-2) + \varepsilon_t$

où ε_t est un bruit aléatoire gaussien centré de variance 1, S_t est une série d'effets saisonniers de période 12 dont les 12 premiers termes sont tirés de façon indépendante dans la loi uniforme sur $[0, 6]$.

Correction de l'exercice 3

```

1
2 T=1000
3 epsilon=rnorm(T,0,1)
4 St=runif(12,0,6)
5 t=1:T
6 St=rep(St,length=T)
7 #1
8 X1=epsilon
9 #2
10 X2=0.4*t+epsilon
11 #3
12 X3=St+epsilon
13 #4

```

```

14 X4=0.4*t+St+epsilon
15 #5
16 X5=0.4*t* St*epsilon
17 #6
18 X6=cos(t/30)*epsilon
19 #7
20 X7= cos(t/30) + epsilon
21 X8=rep(0,T)
22 X9=rep(0,T)
23 X10=rep(0,T)
24
25 # 8,9,10
26 for(t in 3:T)
27 {
28   X8[t]=0.4*X8[t-1]+epsilon[t]
29   X9[t]=epsilon[t]+0.7*epsilon[t-1]
30   X10[t]=0.6*X[t-1]+0.4*X[t-2]+epsilon[t]
31 }
32 }
33 X1=ts(X1, f=12,1982)
34 X2=ts(X2, f=12,1982)
35 X3=ts(X3, f=12,1982)
36 X4=ts(X4, f=12,1982)
37 X5=ts(X5,1982)
38 X6=ts(X6,1982)
39 X7=ts(X7,1982)
40 X9=ts(X2,1982)
41 X10=ts(X10,1982)
42 X11()
43 par(mfrow=c(3,2))
44 plot(X1, type="l")
45 plot(window(X2, type='l',1986,1988))
46 plot(X3, type='l')
47 plot(X4, type='l')
48 plot(X5, type='l')
49 plot(X6, type='l')

```

Exercice 4. Récupérer le fichier contenant la température maximale mensuelle d'Alger entre 1957 et 2004 et le nombre de cas de varicelle relevés à New-York de janvier 1931 à juin 1972.

- Créer un objet de type série temporelle contenant cette série. Représenter graphiquement la série. (Voir mise à jours pour les instructions utiles en R.)
- Analyser qualitativement cette série, c'est-à-dire repérer d'éventuelles tendances et/ou

saisonnalités.

- *Quel est le nombre de cas de varicelle mensuel moyen ?*
- *Tracer les 25 premières auto-corrélations. Interpréter ces résultats.*
- *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).*
- *Tracer sur un graphique l'évolution annuelle du nombre de cas de varicelle.*
- *Ces deux dernières questions vous permettent-elles d'améliorer vos conclusions de la question 2 ?*

Correction de l'exercice 4

```
1 #1)
2 f=file.choose()
3 gg=read.table(f)
4 serie=matrix(c(gg[[1]],gg[[2]],gg[[3]],gg[[4]],gg[[5]],gg[[6]],gg[[7]],gg
  [[8]],gg[[9]],gg[[10]],gg[[11]],gg[[12]]),ncol=12)
5 serie1=ts(as.numeric(t(serie)),1957,f=12)
6 X11()
7 plot(serie1,ylab="Température à Alger")
```

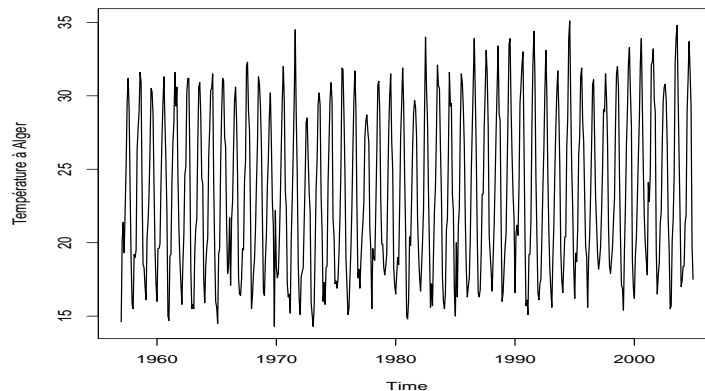


Figure 2.11

```
1 #2) On ne voit pas de tendance mais on voit une saisonnalité
2 #Pour mieux voir la saisonnalité, on peut zoomer sur une durée plus petite
  (les 5 premières années)
3
4 X11()
```

```

5 wserie1=window(serie1 ,1957 ,c(1962,12))
6 plot(wserie1 ,ylab="Température à Alger 1957-1962").

```

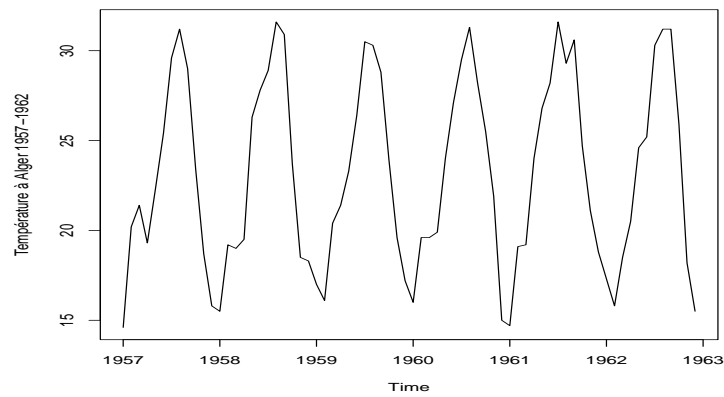


Figure 2.12

```

1 #3)
2 mean(serie)
3 # et on trouve 732, 4076 comme moyenne des données de la varicelle et
  23.35017 pour les données de la température.
4 #(4) On rentre l'instruction
5 acf(serie1 ,ylab="ACF de la température à Alger")
6 #et obtient le graphique de la figure 2.11. La fonction d'autocorrction est
  périodique , ce qui indique une périodicité la série temporelle.

```

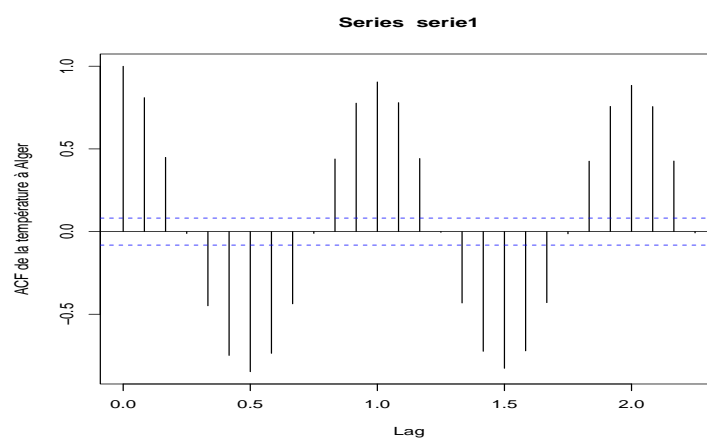


Figure 2.13

```

1 #(5) On utilise le code :

```

```

2 plot(serie[1,],ylim=c(14,35),ylab="graphes superposés")
3 for(i in 1:48)
4 lines(serie[i,],type="l",col=i)

```

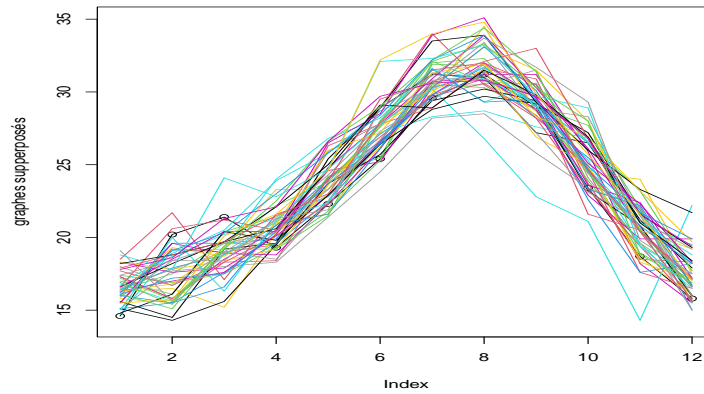


Figure 2.14

```

1 #(6) Code pour l'évolution annuelle de la température maximale:
2 v=apply(serie,1,mean)
3 plot(1:41,v)
4 #Et on obtient le graphique : voir Figure 2.12.
5 #(7) Le dernier graphique semble indiquer une tendance un peu prés
    croissante

```

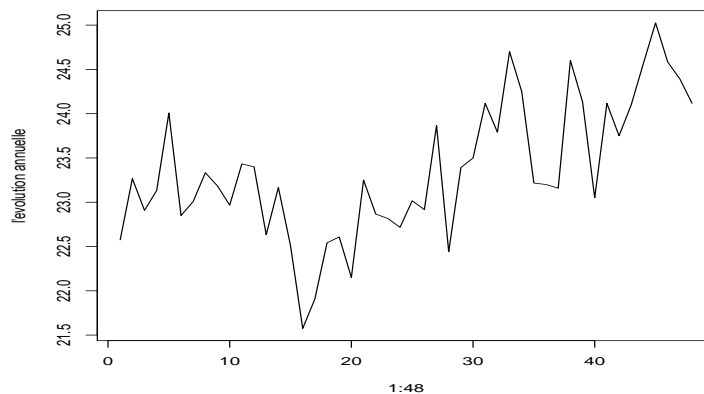


Figure 2.15

Remarque 2.6.1. • *R* arrange donc l'objet de type *ts* sous la forme d'un tableau dont les lignes représentent l'unité de temps et dont le nombre de colonnes est déterminé par la fréquence de la série.

- *la série est univariée et la lecture de ses données doit se faire de gauche à droite, puis de haut en bas. Ainsi, si on veut connaître la 20-eme réalisation de maserie, il faut écrire `maserie[20]`.*
- *Pour définir une série chronologique, seul l'objet `data` est obligatoire. On pourrait en effet définir une série chronologique `maserie2` par `maserie2 = ts(monvecteur)`, les autres objets de la série prennent une valeur par défaut qui est `start=1; frequency=1` et `end=longueur de la série`.*
- *Une corrélation positive entre deux variables traduit la tendance de ces variables à varier dans le même sens (et inversement pour une corrélation négative).*

Chapitre 3

Analyse de la tendance

Nous nous plaçons dans le cadre d'un modèle composé uniquement d'une tendance et de fluctuations irrégulières et donnons différentes méthodes permettant d'estimer la tendance.

3.1 Rappels sur la régression linéaire

Lorsqu'une liaison linéaire forte entre deux variables X et Y semble raisonnable au vu du nuage de points, on a alors une relation du type :

$$Y \cong aX + b,$$

où les coefficients a et b sont inconnus. Le problème est donc d'estimer ces coefficients grâce aux valeurs observées sur l'échantillon. Nous cherchons maintenant la droite qui passe au plus près des points du nuage. Pour cela, il faut donc mesurer l'éloignement des points du nuage par rapport à une droite (D) d'équation $y = ax + b$ puis minimiser un critère d'erreur donné. On peut envisager de minimiser

- la somme des erreurs en valeur absolue : $\min_{a,b} \sum_i |y_i - (ax_i + b)|$.
- $\min_{a,b} \sum_i (y_i - (ax_i + b))^2$.

Mais la méthode des moindres carrés minimisant le second critère est la plus usité.

3.1.1 La méthode des moindres carrés

Pour estimer la tendance par la méthode des moindres carrés on doit minimiser :

$$S(a, b) = \sum_i (y_i - ax_i + b)^2.$$

Pour cela on cherche a, b tels que

$$\frac{\partial S(a, b)}{\partial a} = 0; \frac{\partial S(a, b)}{\partial b} = 0,$$

et on trouve dans ce cas :

$$\hat{a} = \frac{\text{cov}(X, Y)}{\text{var}(X)}, \hat{b} = \bar{Y} - \hat{a}\bar{X}.$$

où $\bar{X} = \frac{1}{T} \sum_{i=1}^T x_i$, $\text{cov}(X, Y) = \sum_{i=1}^T x_i y_i - \bar{X}\bar{Y}$ et $\text{var}(X) = \text{cov}(X, X)$.

La droite d'équation $y = \hat{a}x + \hat{b}$ est appelée droite de régression de Y en X

Remarque 3.1.1. Afin de confirmer qu'il est raisonnable d'approximer le nuage de points par une droite, on calcule le coefficient de corrélation linéaire :

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

on considère que la corrélation linéaire est **faible** lorsque $-0,3 < r(X, Y) < 0$, **moyenne** lorsque $-0,7 \leq r(X, Y) \leq -0,3$ et **forte** lorsque $r < -0,7$.

3.2 Ajustement tendanciel linéaire

Supposons que l'on observe T valeurs d'une série dont la tendance semble être linéaire comme dans l'exemple de la Figure 3.1. La méthode des moindres carrés décrite au paragraphe précédent consiste à estimer la tendance par une fonction linéaire.

$$\hat{C}_t = \hat{a}t + \hat{b}.$$

Le couple solution (\hat{a}, \hat{b}) est donné par :

$$\hat{a} = \frac{\text{cov}(X, t)}{\text{var}(t)}, \hat{b} = \bar{X} - \hat{a}\bar{t}.$$

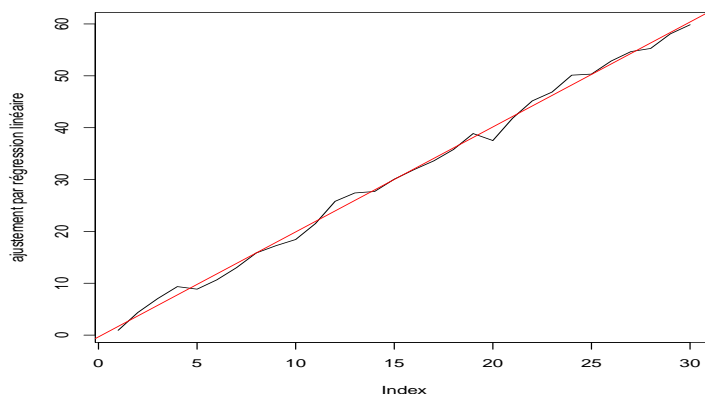


Figure 3.1

3.3 Ajustement tendanciel linéaire par points médians

On suppose ici aussi que la tendance est linéaire. Cette méthode est empirique et ne repose sur aucun critère d'erreur à minimiser. Elle peut cependant s'avérer efficace en présence de valeurs aberrantes. On choisit deux points de coordonnées (t_a, X_a) et (t_b, X_b) et on fait passer la droite par ces deux points. Les coefficients a, b de la droite vérifient :

$$a = \frac{X_b - X_a}{t_b - t_a}, b = X_a - at_a.$$

Pour choisir les deux points, on constitue deux sous-séries d'observations en général d'effectifs égaux (à 1 près). Puis on prend les points médians de chaque sous-série. On peut également prendre les points moyens ou choisir à la main des points judicieux cf(figure 3.2).

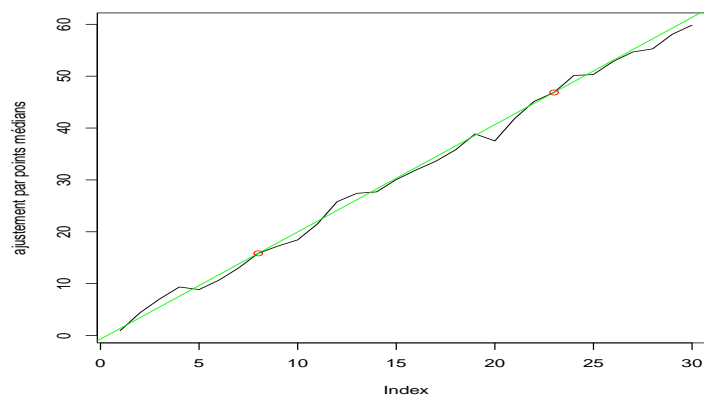


Figure 3.2

3.4 Ajustements tendanciels non linéaires

Dans certaines situations, la modélisation linéaire de la tendance peut être trop simplificatrice. Dans le cas de la Figure 3.3 par exemple, on s'attend plutôt à une tendance quadratique. Lorsque la tendance n'est pas linéaire, une technique simple consiste à se ramener à un ajustement linéaire après un changement de variable approprié. Évidemment ce procédé n'est pas toujours possible et on verra plus loin qu'il existe des méthodes d'ajustement non linéaires directes. C'est la représentation graphique qui motive le choix du changement de variable.

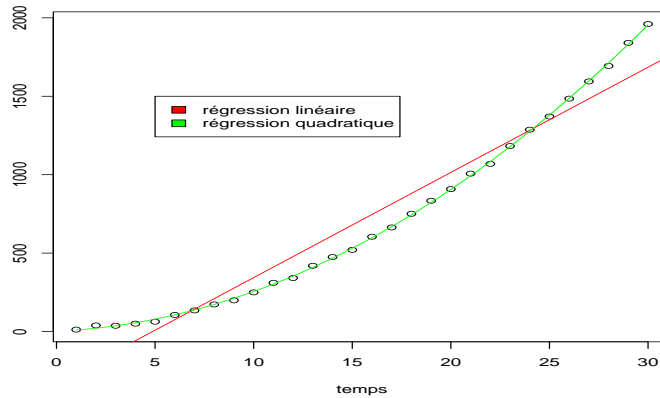


Figure 3.3

3.4.1 Exemple

- Si $Z_t = at^2 + b$, en posant $Y_t = t^2$, on se ramène à $Z_t = aY_t + b$ et on peut faire un ajustement linéaire entre Y_t et Z_t .
- Si $Z_t = be^{at}$, en posant $Y_t = \ln(Z_t)$, on se ramène à $Y_t = at + \ln(b)$ et on peut faire un ajustement linéaire entre Y_t et t . Dans le cas de l'exemple de la Figure 3.3, on peut aussi mener des calculs analogues à ceux effectués précédemment et approcher la tendance par $Z_t = at^2 + bt + c$.

3.5 Lissage de la tendance par moyenne mobile

Dans certaines situations, il n'est pas facile de trouver le degré du polynôme d'ajustement pour C_t ou de changement de variable adéquat. L'utilisation d'un polynôme avec un degré élevé mais le nombre de paramètres à estimer serait important et rendrait les calculs fastidieux. Dans ce cas, on a recours à la théorie non paramétrique de l'estimation de la tendance et on approxime la tendance par la moyenne mobile arithmétique.

3.5.1 Définition

On appelle moyenne mobile, une transformation de X_t s'écrivant comme combinaison linéaire finie des valeurs de la série correspondant à des dates entourant t . La série transformée s'écrit :

$$MM_{m1+m2+1} = \sum_{i=-m1}^{i=m2} \alpha_i X_{t-i},$$

- On dit que la moyenne mobile est **centrée** lorsque $m1 = m2 = m$.

- Une moyenne mobile centrée est **symétrique** si et seulement si $\theta_{-i} = \theta_i, i = 1, \dots, m$.
- Une moyenne mobile arithmétique est une moyenne mobile centrée, d'ordre (impair) $2m + 1$ et telle que $\theta_i = \frac{1}{2m+1}, i = -m, \dots, m$.

3.5.2 Notation

On peut réécrire la moyenne mobile en termes d'opérateurs. On définit pour cela l'opérateur B, appelé opérateur retard, qui à tout processus $(X_t)_{t \in \mathbb{Z}}$ associe le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

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

Si on compose B avec lui-même on obtient $B^2 = B \circ B$ tel que

$$\forall t \in \mathbb{Z}, B^2 X_t = X_{t-2}.$$

On peut itérer cette application et définir par récurrence $B^k X_t = X_{t-k}, k \in \mathbb{N}$. Par convention, B^0 est l'opérateur identité I.

L'opérateur B est linéaire et inversible. Son inverse $B^{-1} = F$ est défini par $\forall t \in \mathbb{Z}, FX_t = X_{t+1}$. L'opérateur F est appelé opérateur avance.

Exemple 3.5.1.

$$(B + I + B^2 + B^{-2})X_t = X_{t-1} + X_t + X_{t-2} + X_{t+2}.$$

$$MM_{m1+m2+1} = \sum_{i=-m1}^{i=m2} \alpha_i B^i; m1 = m2 = 2, \alpha_{-2} = \alpha_1 = \alpha_0 = \alpha_2 = 1 \text{ et } \alpha_{-1} = 0.$$

3.5.3 Cas particulier : Moyenne mobile arithmétique.

La série des moyennes mobiles arithmétiques d'ordre p, notée $(M_p(t))_t$, est la série des moyennes de p observations consécutives et elle prend ses valeurs aux dates moyennes correspondantes. Plus précisément, on calcule les moyennes de p termes consécutifs pour les dates

$$\frac{t_1 + t_2 \dots + t_p}{p}, \frac{t_2 + t_3 \dots + t_{p+1}}{p}, \frac{t_3 + t_4 \dots + t_{p+2}}{p}, \dots, \frac{t_{T-p} + t_{T-p+1} \dots + t_T}{p},$$

et pour les variables

$$\frac{X_1 + X_2 \dots + X_p}{p}, \frac{X_2 + X_3 \dots + X_{p+1}}{p}, \dots, \frac{X_{T-p} + X_{T-p+1} \dots + X_T}{p}.$$

Exemple 3.5.2. *Calcul d'une moyenne mobile d'ordre 3*

<i>t</i>	1	2	3	4	5	6	7
X_t	2	3	6	5	8	2	4
$MM_3(t)$	NA	$\frac{(2+3+6)}{3}$	$\frac{(3+6+5)}{3}$	$\frac{(6+5+8)}{3}$	$\frac{(5+8+2)}{3}$	$\frac{(8+2+4)}{3}$	$\frac{(2+4+9)}{3}$
	NA	3.67	4.67	6.33	5.00	4.33	2.67

8	9	10
9	7	3
$\frac{(4+9+7)}{3}$	$\frac{(9+7+3)}{3}$	NA
4.33	6.33	NA

3.5.4 Moyenne mobile d'ordre pair

Dans la définition de la moyenne arithmétique, p est impair, $p = 2m+1$. Généralisons cette procédure à tout p . Lorsque p est impair, $p = 2m+1$, la série moyenne mobile est calculée aux mêmes instants que les observations initiales. Par contre, lorsque p est pair, $p = 2m$, la moyenne mobile est calculée entre les dates d'observations. Si l'on veut comparer la série transformée à la série initiale, on a besoin d'avoir les valeurs pour les mêmes dates d'observations. Pour pallier cet inconvénient, on prendra plutôt comme transformation :

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

Exemple 3.5.3. *Calcul des moyennes mobiles arithmétiques d'ordre 2 et 4*

Date <i>t</i>	Série X_t	Date $MM_2(t)$	$MM_2(t)$
1	5	NA	NA
2	3	$(1/2+2+3/2)=2$	$(5/2+3+4/2)/2=2.5$
3	4	$(2/2+3+4/2)=3$	$(3/2+4+9/2)/2=3.33$
4	9	$(3/2+4+5/2)=4$	$(4/2+9+3/2)/2=4.16$
5	3	$(4/2+5+6/2)=5$	$(9/2+3+5/2)/2=3.33$
6	5	NA	NA

Date t	Série X_t	Date $MM_4(t)$	$MM_4(t)$
1	5	NA	NA
2	3	NA	NA
3	4	$(1/2+2+3+4+5/2)/4=3$	$(5/2+3+4+9+3/2)/4=5$
4	9	$(2/2+3+4+5+6/2)/4=4$	$(3/2+4+9+3+5/2)/4=5$
5	3	NA	NA
6	5	NA	NA

3.6 Propriétés d'un lissage par moyenne mobile

3.6.1 Effet d'une moyenne mobile sur une tendance

- L'application d'une moyenne mobile arithmétique (paire ou impaire) ne modifie pas une tendance constante.
- L'application d'une moyenne mobile arithmétique (paire ou impaire) conserve une tendance linéaire.

3.6.2 Effet d'une moyenne mobile sur une composante saisonnière

Si la série X_t possède une composante saisonnière de période p alors l'application d'une moyenne mobile d'ordre p supprime cette saisonnalité.

3.6.3 Effet d'une moyenne mobile sur les fluctuations irrégulières

Jusqu'à présent nous ne nous sommes intéressés qu'à l'effet d'une moyenne mobile sur la partie déterministe de la série (tendance et saisonnalité). L'application d'une moyenne mobile arithmétique sur un bruit blanc réduit sa variance.

3.6.4 Choix pratique de l'ordre d'une moyenne mobile

Nous rappelons que le but d'un lissage par moyenne mobile est de faire apparaître l'allure de la tendance.

En pratique, on doit donc trouver le meilleur compromis pour le choix de l'ordre de lissage optimal. Dans la pratique, il est fréquent que la saisonnalité s'exprime par une fonction périodique, de période paire $p = 2m$. C'est le cas par exemple, des données mensuelles ($p = 12$), trimestrielles ($p = 4$). Dans ce cas, pour éliminer la saisonnalité, nous utiliserons comme définit

précédemment la moyenne mobile d'ordre p .

Nous rappelons que le but d'un lissage par moyenne mobile est de faire apparaître l'allure de la tendance. Il s'agit donc de faire disparaître la saisonnalité et de réduire au maximum le bruit blanc.

3.7 Mise en œuvre sous R

1) Pour effectuer un ajustement linéaire et non linéaire, la fonction de modélisation principale utilisée est la fonction `lm`. Le premier argument de la fonction `lm()` est un objet de type formule qui fournit à R la liste des effets à inclure dans le modèle linéaire.

- $Y \sim 1$, modéliser Y en fonction de l'intercept uniquement.
- $Y \sim X$, modéliser Y en fonction de X avec intercept.
- $Y \sim -1 + X$, modéliser Y en fonction de X sans intercept.
- $Y \sim X1 + X2$, modéliser Y en fonction de $X1$ et $X2$ avec intercept.
- $Y \sim I(f(X))$, modéliser Y en fonction de $f(X)$, une fonction quelconque de X , avec intercept. Tout cela peut bien entendu être mixé pour créer un modèle bien spécifique. Par exemple pour une régression quadratique sans intercept on utilisera $Y \sim -1 + X + I(X^2)$. Les `I()` sont importants car ils servent à éviter l'interprétation de X^2 . Si cela n'est pas clair, essayez la même commande sans les `I()`. Il est parfois préférable pour des raisons de travailler avec des polynômes orthogonaux (cela a tendance à réduire la corrélation entre les coefficients obtenus par la régression) dans ce cas on peut utiliser comme formule `poly(x,3)`. Toutefois l'interprétation des coefficients ainsi obtenus n'est pas évidente.

2) Pour calculer des moyennes mobiles, on utilise la fonction **`filter`** :

`filter(data, poids, sides=2)`

- L'objet `data` est de type série chronologique.
- L'objet `poids` contient le vecteur des coefficients θ_j .
- L'objet `sides = 2` pour une MM centrée et `sides = 1` pour une MM utilisant uniquement le passé.

On peut aussi utiliser la fonction `decompose` :

`decomp=decompose(serie,type=c("additive","multiplicative"))`

- `decomp$seasonal` donne la composante saisonnière que l'on suppose additive ou multiplicative dans l'option `type`,
- `decomp$trend` donne la tendance (via les moyennes mobiles).
- `decomp$random` donne la partie aléatoire stationnaire de la série.

3) Pour tester la blancheur d'un bruit blanc, on utilise la fonction Box.test :

Box.test(serie,lag=H)

- On prendra (H0)=bruit blanc, (H1)=pas bruit blanc, probabilité de rejeter (H0) à tort= α avec $\alpha = 0,05$.

3.8 Exercices et feuille de travaux pratiques 2

Exercice 5. La série $(X_t)_t$ ci-dessous représente la quantité d'un produit P vendu par une entreprise sur les 9 dernières années (en milliers d'articles) :

t	1	2	3	4	5	6	7	8	9
X_t	32	48	38	52	61	73	80	84	95

- Effectuer un ajustement linéaire par la méthode des moindres carrés, points médians.
- Représenter graphiquement les résultats .

Correction de l'exercice 5

1. Pour l'ajustement par moindres carrés on a :

$$\bar{t} = 5; \bar{X} = 62.56; var(t) = 8.44; cov(t, X) = 65.39; a = 7.75; b = 23.81.$$

Pour l'ajustement par points médians, on répartie la série en deux sous séries x_1, x_2 de longueur 4 et 5 respectivement et on trouve :

$$med(t_1) = 2.5; med(t_2) = 7; med(x_1) = 43; med(x_2) = 80.$$

2.

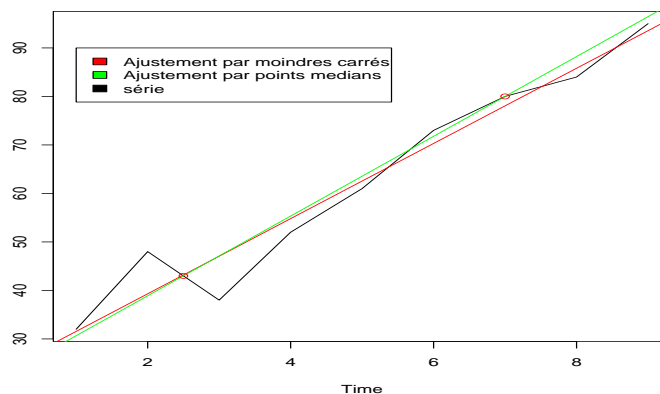


Figure 3.4

Exercice 6. Démontrer que :

- L'application d'une moyenne mobile arithmétique (paire ou impaire) ne modifie pas une tendance constante.
- L'application d'une moyenne mobile arithmétique (paire ou impaire) conserve une tendance linéaire.

Exercice 7. Calculer les séries des moyennes mobiles d'ordre 2, 3 et 4 de la série initiale X_t :

t	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
X_t	232	239	784	555	346	855	232	100	147	555	122	556

t	2013	2014	2015	2016
X_t	124	451	214	457

Feuille de travaux pratiques 2

Exercice 8. Ecrire un code R à fin de résoudre l'exercice 5.

Correction de l'exercice 8

```
1 X=c(32, 48, 38, 52, 61, 73, 80, 84, 95)
2 X=ts(X)
3 t=1:9
4 X11()
5 reg=lm(X~t)
6 reg
7 m=mean(t)
8 mx=mean(X)
9 v=9/8*var(t)
10 c=9/8*cov(t,X)
11 a=c/v
12 b=mx-a*m
13 m;mx;v;c;a;b
14 plot(X,type="l",ylab="")
15 abline(reg,col="red")
16
17 x1=median(X[1:4])
18 x2=median(X[5:9])
19 t1=median(t[1:4])
20 t2=median(t[5:9])
21 t1;t2;x1;x2
22 a=(x2-x1)/(t2-t1)
23 b=x2-a*t2
```

```

24 abline(b, a, col="green")
25 points(t1, x1, col="red")
26 points(t2, x2, col="red")
27 legend(1, 90, c("Ajustement par moindres carrés", "Ajustement par points mé
    dians", "série"), c("red", "green", "black"))

```

Exercice 9. Pour les données *AirPassengers* effectuer :

- Estimation paramétrique de la tendance
 - (a) Représenter graphiquement la série. Ce processus vous semble-t-il stationnaire? Présente-t-il des tendances et des saisonnalités?
 - (b) Estimer les paramètres d'une tendance linéaire $at + b$.
 - (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.
- Méthode des moyennes mobiles
 - (a) Appliquer la méthode des moyennes mobiles pour enlever la tendance et la saisonnalité.
 - (b) La série obtenue semble-t-elle stationnaire? Pourrait-on la modéliser par un bruit blanc?

Correction de l'exercice 9

```

1   t=1:144
2   x=AirPassengers
3   n=length(x)
4   y=t*x
5   #On utilise les formules des moindres carrés.
6   a=(6/(n*(n-1)))*(-sum(y)+(2*n+1)*n*mean(x)/3)
7   b=(12/(n*(n^2-1)))*(sum(y)-(n+1)*n*mean(x)/2)
8   print(a)
9   print(b)
10  #1c
11  xvect=as.numeric(x) #On met x sous forme vectorielle pour simplifier
    les manipulations.
12  residu=xvect-a-b*t
13  plot(t, residu, 'l')
14  print(mean(residu))
15  #1d
16  autoco=acf(residu, lag.max=40, type=c("correlation"))
17  plot(autoco)
18  #Donner une autre méthode pour calculer a et b
19  reg=lm(x~t)

```

```

20 #regression linéaire
21 reg
22 fitt=c(1/24,rep(1/12,11),1/24)
23 r=filter(x,fitt ,side=1)
24 r2=filter(x,fitt ,side=2)
25 r2=decompose(x)
26 # une autre méthode pour calculer les moyennes mobiles
27 r2$trend
28

```

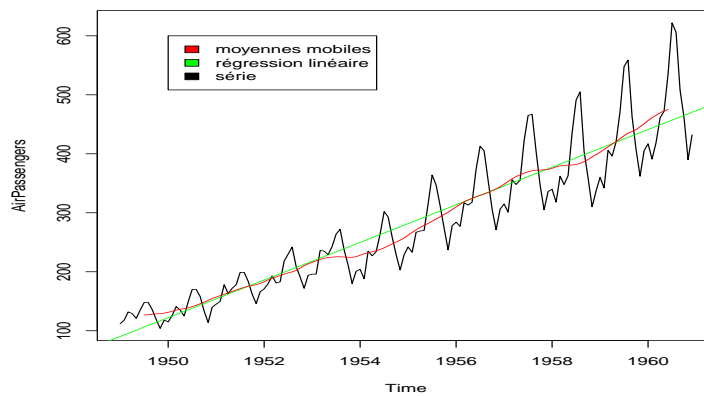


Figure 3.5

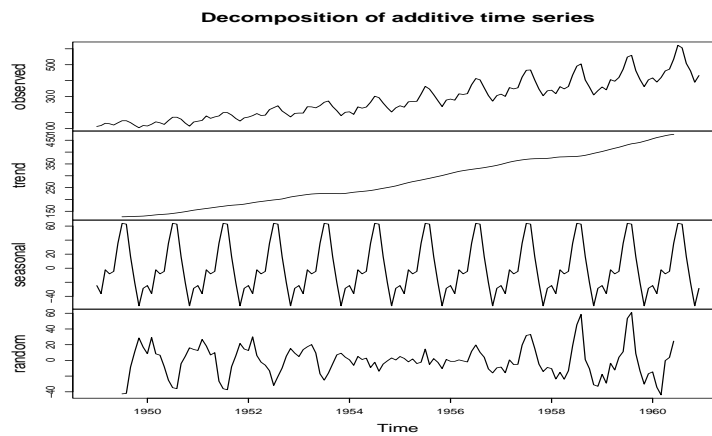


Figure 3.6

Chapitre 4

Calcul des variations saisonnières et accidentelles.

La tendance C_t est estimée par ajustement ou lissage (moyenne ou médiane mobile). On va maintenant estimer les variations saisonnières S_t .

Rappel : dans le cas d'une estimation de C_t par moyennes mobiles, il y a $p/2$ données manquantes au début et $p/2$ données manquantes à la fin.

4.1 Calcul des données sans tendance

- Cas du modèle additif.
Les données sans tendance sont $Y_t - C_t$.
- Cas du modèle multiplicatif.
Les données sans tendance sont Y_t / C_t

4.2 Calcul des coefficients saisonniers S_j

Etant donné que l'on a fait l'hypothèse que les variations saisonnières se répètent à l'identique chaque année, on estime un coefficient saisonnier pour chacun des p mois, la variation saisonnière de tous les mois j sera le coefficient saisonnier du mois j . On considère les données sans tendance, on les range par année (en ligne) et par mois (en colonne). On calcule la moyenne des données sans tendance concernant le mois j des n années, ce qui donne une première estimation du coefficient saisonnier S_j . On fait ceci pour chacun des mois j ($j = 1, 2, \dots, p$).

- Cas du modèle additif : $S_j = \frac{1}{n} \sum_{i=1}^n (X_{ij} - Z_{ij})$
- Cas du modèle multiplicatif : $S_j = \frac{1}{n} \sum_{i=1}^n (X_{ij}/Z_{ij})$

Dans le cas d'un lissage par moyennes ou médianes mobiles, les calculs ne portent pas sur toutes les années.

4.2.1 Exemple

Calculons les coefficients saisonniers de la série trimestrielle représentée par le tableau suivant :

	Qtr1	Qtr2	Qtr3	Qtr4
2000	2.70	3.51	8.39	8.77
2001	11.36	12.19	14.96	14.91
2002	18.04	20.21	23.16	24.67
2003	27.07	28.41	29.09	32.72
2004	33.95	35.14	39.31	41.09

- Calcul de la série des moyennes mobiles d'ordre 4

	Qtr1	Qtr2	Qtr3	Qtr4
2000	NA	NA	6.93	9.09
2001	11.00	12.59	14.19	16.03
2002	18.05	20.30	22.65	24.80
2003	26.57	28.32	30.18	31.88
2004	34.00	36.33	NA	NA

- Calcul de la série sans tendance

	Qtr1	Qtr2	Qtr3	Qtr4
2000	NA	NA	1.46	-0.32
2001	0.36	-0.40	0.77	-1.12
2002	-0.01	-0.09	0.51	-0.13
2003	0.50	0.09	-1.09	0.84
2004	-0.05	-1.19	NA	NA

- Calcul de la saisonnalité

1	2	3	4
0.2000	-0.40	0.41	-0.18

- Calcul des coefficients saisonniers corrigés

1	2	3	4
0.19	-0.40	0.40	-0.19

4.3 Série désaisonnalisée ou série CVS

4.3.1 Définition

On appelle série désaisonnalisée ou série corrigée des variations saisonnières **notée série CVS**, la série chronologique à laquelle on a enlevé les variations saisonnières.

Dans le cas du modèle additif : La série désaisonnalisée est $D_t = X_t - S_t$

Dans le cas du modèle multiplicatif : $D_t = X_t / S_t$

La série désaisonnalisée est D_t

Exemple 4.3.1. En s'appuyant sur les résultats de l'exemple précédent, la séries CVS est donnée par :

	Qtr1	Qtr2	Qtr3	Qtr4
2000	2.50	3.91	7.98	8.96
2001	11.16	12.59	14.55	15.10
2002	17.84	20.61	22.75	24.86
2003	26.87	28.81	28.68	32.91
2004	33.75	35.54	38.90	41.28

4.3.2 Intérêts

- La particularité de la série CVS est que les données de D_t sont directement comparables : on a enlevé l'effet des saisons et donc le caractère propre de chaque mois, on peut donc par exemple comparer les données d'un mois de janvier et celle d'un mois de juillet.
- A partir de la série CVS, on peut réévaluer la tendance par ajustement ou lissage (moindres carrés ou Mayer sur D_t , ou moyennes mobiles), afin d'avoir une meilleure estimation de la tendance.

4.4 La série ajustée

La série ajustée notée, \hat{X}_t est obtenue en recomposant les deux composantes estimées : la tendance et les variations saisonnières selon le modèle qui a été choisi.

Dans le cas du modèle additif : $\hat{X}_t = C_t + S_t$ ou encore $\hat{X}_{ij} = C_{ij} + S'_j$.

Dans le cas du modèle multiplicatif : $\hat{X}_t = C_t \times S_t$ ou encore $\hat{X}_{ij} = C_{ij} \times S'_j$.

4.4.1 Ce qu'elle représente

La série ajustée représente l'évolution qu'aurait subi la grandeur observée, si les variations saisonnières avaient été parfaitement périodiques (s'étaient répétées à l'identique d'une année sur l'autre) et s'il n'y avait pas eu de variations accidentelles.

4.5 Variations accidentelles ou résiduelles

Dans le cas du modèle additif, c'est la différence entre la série X_t et sa série ajustée représente les variations accidentelles ou résiduelles : $\varepsilon_t = X_t - \hat{X}_t$.

Dans le cas du modèle multiplicatif (1ere forme), la différence entre la série X_t et sa série ajustée Y représente les variations accidentelles ou résiduelles : $\varepsilon_t = X_t - \hat{X}_t$.

Dans le cas du modèle multiplicatif (2eme forme), le rapport entre la série X_t et sa série ajustée représente les variations accidentelles ou résiduelles : $\varepsilon_t = \frac{X_t}{\hat{X}_t}$

Remarque 4.5.1. La seule différence entre les 2 modèles multiplicatifs est le calcul de ε_t .

4.6 Description schématique de l'étude complète d'une série chronologique

Pour la décomposition et la prévision d'une série chronologique :

- (a) On trace le graphe de (X_t) et le graphique des courbes superposées.
- (b) On estime la tendance C_t , et on la trace.
- (c) On choisit le modèle de composition : additif ou multiplicatif.
- (d) On calcul la série corrigée de la tendance puis on calcul les variations saisonnières.
- (e) On calcul la série corrigée des variations saisonnières CVS.
- (f) On réestime la tendance.
- (g) On calcule la série ajustée et la série résiduelle.

- (h) On fait un diagnostic du modèle en testant la blancheur des résidus si c'est un bruit blanc on passe à l'étape (i).
- (i) Enfin, une fois ces différentes étapes réalisées, nous sommes en mesure de faire de la prédiction.

4.7 Exercices et feuille de travaux pratiques 3

Exercice 10. Soit les séries X, Y données par le tableau suivant :

t	série X_t	série Y_t
1	7	10
2	9	14
3	11	18
4	13	22
5	15	26
6	17	30
7	19	34
8	21	38
9	23	42
10	25	46
11	27	50
12	29	54
13	31	58
14	33	62
15	35	66
16	37	70
17	39	74
18	41	78
19	43	82

- Donner les moyennes mobiles d'ordre 5 de X, Y et de $Z=X+Y$.
- Conclure !

Correction de l'exercice 10

1. Rappel :

- Moyenne mobile centrée d'ordre impaire ($p = 2k + 1$) est définie par :

$$MM_p(t) = \frac{1}{p} \sum_{i=-k}^k X_{t+i}.$$

— Moyenne mobile centrée d'ordre paire ($p = 2k$) est définie par :

$$MM_p(t) = \frac{1}{p} \sum_{i=-k+1}^{k-1} X_{t+i} + \frac{1}{2}X_{t-k} + \frac{1}{2}X_{t+k}.$$

Moyenne mobile d'ordre 5 de la série X_t

t	1	2	3	4	5	6	7	8	9	10	11	12
$MM_5(X_t)$	NA	NA	11	13	15	17	19	21	23	25	27	29
t	13	14	15	16	17	18	19	20	21	22	23	24
$MM_5(X_t)$	31	33	35	37	39	41	43	45	47	49	NA	NA

Moyenne mobile d'ordre 5 de la série Y_t

t	1	2	3	4	5	6	7	8	9	10	11	12
$MM_5(Y_t)$	NA	NA	18	22	26	30	34	38	42	46	50	54
t	13	14	15	16	17	18	19	20	21	22	23	24
$MM_5(Y_t)$	58	62	66	70	74	78	82	86	90	94	NA	NA

Moyenne mobile d'ordre 5 de la série Z_t

t	1	2	3	4	5	6	7	8	9	10	11	12
$MM_5(Z_t)$	NA	NA	29	35	41	47	53	59	65	71	77	83
t	13	14	15	16	17	18	19	20	21	22	23	24
$MM_5(Z_t)$	89	95	101	107	113	119	125	131	137	143	NA	NA

2. Sachant que $X=2t+5$, $Y=4t+6$. La série Z peut se réécrire comme $Z_t = X_t + Y_t$

— Nous vérifions de la question précédente que $MM_5(X_t) = X_t$, donc la moyenne mobile laisse passer une tendance linéaire invariante.

— Nous vérifions aussi que la moyenne mobile de la somme est égale à la somme des moyennes mobiles.

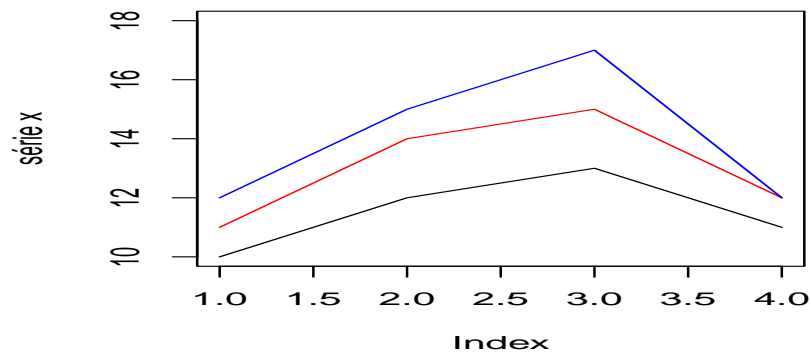
Exercice 11. On a relevé le nombre de mariages dans une petite ville du sud-ouest de l'Algérie chaque trimestre pendant 3 ans : On notera Y la variable dont on étudie l'évolution.

Trimestre	1998	1999	2000
1	10	11	12
2	12	14	15
3	13	15	17
4	11	12	12

1. Représenter graphiquement cette série chronologique (avec périodes superposées). Commenter.

2. Calculer la série des moyennes mobiles, lisser la courbe. Qu'elle est l'effet de la moyenne mobile sur la série ?.
3. Calculer l'équation de la droite de tendance et tracer cette droite sur le graphique précédent.
4. Calculer les quatre coefficients saisonniers.
5. Utiliser le modèle construit pour prévoir le nombre de mariages dans cette ville en 2002.

Correction de l'exercice 11

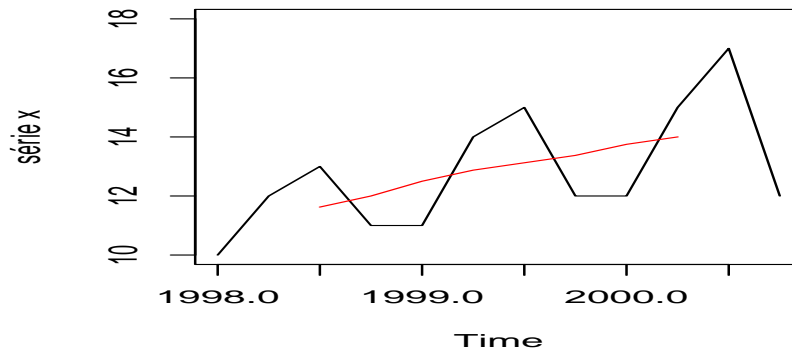


1.

Figure 4.2

2. Moyenne mobile d'ordre 4

	Qtr1	Qtr2	Qtr3	Qtr4
1998	NA	NA	11.625	12
1999	12.5	12.875	13.125	13.375
2000	13.750	14	NA	NA



3.

Figure 4.2

4. Calcule par méthode des moindres carrées

$$\bar{t} = 1999.375$$

$$\bar{X}_t = 12.833$$

$$var(t) = 0.8125$$

$$var(X_t) = 4.151$$

$$cov(t, X_t) = 1.0223$$

$$a = \frac{cov(t, X_t)}{var(t)} = 1.259; b = \bar{X} - a\bar{t}, b = -2503.862$$

Si on utilise $t=1 : 12$ on trouve les résultats suivants :

$$\bar{t} = 6.5$$

$$var(t) = 13$$

$$var(X_t) = 4.151$$

$$cov(t, X_t) = 4.09$$

$$b = 10.7879; a = 0.3147$$

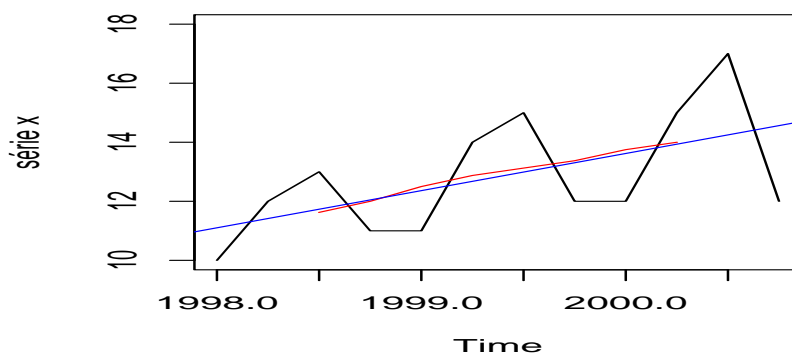


Figure 4.3

5. Série sans tendance

	Qtr1	Qtr2	Qtr3	Qtr4
1998	0.901	1.051	1.108	0.913
1999	0.890	1.104	1.155	0.902
2000	0.881	1.076	1.193	0.824

Pour la Saisonnalité on a :

S_1	S_2	S_3	S_4
0.891	1.077	1.152	0.880

On a pas besoin de corriger la saisonnalité car la moyenne des coefficients saisonniers =1.

6. Prévision pour 2002

14.37, 17.724, 19.314, 15.026.

Exercice 12. Dans le tableau suivant, On a le relevé d'une série chaque trimestre pendant 4 ans : On notera X_t la variable dont on étudie l'évolution.

X_t	trimestre 1	trimestre 2	trimestre 3	trimestre 4
2013	6.44	9.87	5.81	9.90
2014	7.75	13.54	5.65	14.90
2015	12.95	19.61	7.64	21.09
2016	16.69	26.67	7.23	23.05
2017	19.74	30.88	13.46	27.50

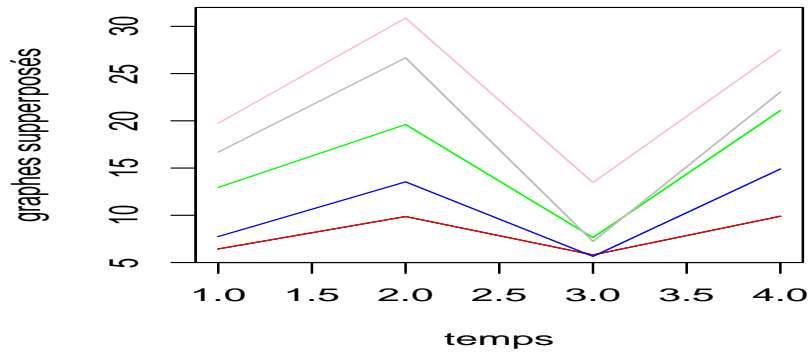
1. Représenter graphiquement cette série chronologique (avec périodes superposées). Commenter.
2. Calculer la série des moyennes mobiles, lisser la courbe. Qu'elle est l'effet de la moyenne mobile sur la série.?(compléter le tableau suivant)

X_t	trimestre 1	trimestre 2	trimestre 3	trimestre 4
2013				
2014	9.23	9.84	11.11	12.52
2015	13.53	14.55	15.79	17.14
2016	17.97	18.17	18.79	19.70
2017	21.00	22.34	NA	NA

3. Donner la série sans tendance.
4. Calculer les coefficients saisonniers.
5. Donner la série CVS. Quelle est l'intérêt de calculer cette série ?
6. Réestimer la tendance par régression linéaire en utilisant la série CVS ($\bar{t} = 10.5$, $var(t)=33.25$, $var(CVS)=30.90$, $CVS=14.98$).

7. Prédire la valeur de Janvier 2020.

Correction de l'exercice 12



1.

Figure 4.4

2. Série des moyennes mobiles :

X_t	trimestre 1	trimestre 2	trimestre 3	trimestre 4
2013	NA	NA	8.169	8.791
2014	9.23	9.84	11.11	12.52
2015	13.53	14.55	15.79	17.14
2016	17.97	18.17	18.79	19.70
2017	21.00	22.34	NA	NA

3. Série sans tendance

	Qtr1	Qtr2	Qtr3	Qtr4
2013	NA	NA	0.711	1.126
2014	0.840	1.377	0.509	1.190
2015	0.957	1.348	0.484	1.230
2016	0.929	1.468	0.385	1.170
2017	0.940	1.382	NA	NA

saisonnalité provisoire

S_t : 0.916 1.394 0.522 1.1792

saisonnalité corrigées

$\bar{S}_t = 1.002$

S_t^c : 0.913 1.390 0.521 1.176

4. Série CVS

	Qtr1	Qtr2	Qtr3	Qtr4
2013	7.050	7.100	11.161	8.420
2014	8.484	9.740	10.854	12.673
2015	14.176	14.106	14.676	17.937
2016	18.271	19.185	13.889	19.604
2017	21.609	22.213	25.856	23.389

5. ($\bar{t}=10.5, CVS =15.0196, var(t)=35 var(CVS)=32.69, cov(t, CVS)=32.129$).

$$a = 0.9362, b = 5.1881$$

6. (prévision pour Janvier 2020) = $0.9362 \cdot 29 + 5.1881 + 0.913 = 33.25$.

Exercice 13. Le tableau suivant représente un extrait de la série co2

	Jan	Feb	Mar	Apr	May	Jun	Jul
1959	315.42	316.31	316.50	317.56	318.13	318.00	316.39
1960	316.27	316.81	317.42	318.87	319.87	319.43	318.01
1961	316.73	317.54	318.38	319.31	320.42	319.61	318.42
...
1996	362.09	363.29	364.06	364.76	365.45	365.01	363.70
1997	363.23	364.06	364.61	366.40	366.84	365.68	364.52

Aug	Sep	Oct	Nov	Dec
314.65	313.68	313.18	314.66	315.43
315.74	314.00	313.68	314.84	316.03
316.63				
....
361.54	359.51	359.65	360.80	362.38
362.57	360.24	360.83	362.49	364.34

1. Compléter le tableau des moyennes mobiles suivant :

	Jan	Feb	Mar	Apr	May	Jun
1958	NA	NA	NA	NA	NA	NA
1959				316.662	316.690	316.722
1960	317.041	317.095	317.167	317.263	317.371	317.451
1961	317.981	318.050	318.129	318.187	318.216	318.269
...
1995	361.937	362.095	362.238	362.374	362.500	362.618
1996	363.234	363.311	363.385	363.464	363.584	363.736

<i>Jul</i>	<i>Aout</i>	<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
315.861	315.917	315.977	316.070	316.197	316.329
316.767	316.816	316.887	316.945	316.986	317.017
317.529	317.608	317.692	317.786	317.850	317.903
318.331	318.386	318.415	318.455	318.540	318.627
...
362.734	362.814	362.869	362.960	363.086	363.172
NA	NA	NA	NA	NA	NA

2. Compléter le tableau de la série sans tendance suivant :

	<i>Jan</i>	<i>Feb</i>	<i>Mar</i>	<i>Apr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Aug</i>	<i>sep</i>	<i>oct</i>
1958	NA	NA	NA	NA	NA	NA	0.529	-1.267	-2.297	-2.890
1959	-0.186	0.241	0.793	2.208	3.180	2.708	1.243	-1.076	-2.887	
1960	-0.311	0.445	1.213	2.047	3.049	2.159	0.891	-0.978	-2.862	-2.626
1995	0.153	1.195	1.822	2.386	2.950	2.392	0.966	-1.274	-3.359	-3.310
1996	-0.004	0.749	1.225	2.936	3.256	1.944	NA	NA	NA	NA

<i>nov</i>	<i>dec</i>
-1.537	-0.899
-2.146	-0.987
-1.910	-1.053
-2.286	-0.792
NA	NA

3. Comment calculer les coefficients saisonniers ?

4. Après les calculs on trouve les coefficients saisonniers suivants :

<i>t</i>	1	2	3	4	5	6	7	8	9	10	11	12
S_t	-0.052	0.612	1.377	2.519	3.002	2.331	0.815	-1.249	-3.053	-3.250	-2.068	-0.90

Corriger les variations saisonnières si nécessaires.

5. Compléter le tableau (de la série CVS) suivant :

	<i>Jan</i>	<i>Feb</i>	<i>Mar</i>	<i>Apr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Aug</i>
1958			315.124	315.043	315.130	315.671	315.577	315.901
1959	316.324	316.199	316.044	316.353	316.870	317.101	317.197	316.991
1960	316.784	316.929	317.004	316.793	317.420	317.281	317.607	317.881
..
1995	362.144	362.679	362.684	362.243	362.450	362.681	362.887	362.791
1996	363.284	363.449	363.234	363.883	363.840	363.351	363.707	363.821

<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
316.735	316.432	316.730	316.395
317.055	316.932	316.910	316.995
317.885	318.412	318.010	317.815
..
362.565	362.902	362.870	363.345
363.295		364.560	365.305

6. Faire une régression linéaire sur la série CVS

($\bar{t} = 234.5$, $CVS = 337.0536$; $var(t) = 18291$, $var(CVS) = 220.7469$, $cov(t, CVS) = 1997.49$)

7. Faire des prévisions pour l'année 1997 et calculer l'erreur absolue moyenne sachant que les vraies valeurs sont :

1997	<i>Jan</i>	<i>Feb</i>	<i>Mars</i>	<i>Avril</i>	<i>Mai</i>
Vraies valeurs	363.23	364.06	364.61	366.40	366.84

<i>Juin</i>	<i>Juillet</i>	<i>Aout</i>	<i>Sept</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
365.68	364.52	362.57	360.24	360.83	362.49	364.34

Chapitre 5

Lissages exponentiels

Dans l'industrie il est courant que l'on doive effectuer des prévisions de vente de production ou de stock pour une centaine de produits. Ces prévisions sont effectuées à partir de données qui peuvent être par exemple journalières, hebdomadaires ou mensuelles et l'horizon de prévision est court. Un certain nombre de techniques sont regroupées sous le nom de lissage exponentiel. Ces approches construisent des prévisions assez précises. Elles sont peu coûteuses et les calculs peuvent être rendus très rapides par l'utilisation de formules récursives de mise à jour. Elles ont beaucoup de succès dans les milieux dans lesquels un grand nombre de chroniques unidimensionnelles doivent être analysées séparément dans un but de prévision.

Ces techniques sont introduites par Holt en 1958, Winters en 1960 et popularisées par le livre de Brown en 1963, les méthodes de lissage constituent l'ensemble des techniques empiriques de prévision qui accordent plus ou moins très très d'importance aux valeurs du passé d'une série temporelle. Les trois modèles ci-dessous seront traités dans ce chapitre :

- $X_t = C_t + \varepsilon_t$;
- $X_t = C_t + S_t + \varepsilon_t$;
- $X_t = C_t S_t + \varepsilon_t$;

avec C_t une série constante ou linéaire. La composante stochastique ne sera pas nécessairement un bruit blanc.

5.1 lissage exponentiel simple (LS)

Modèle considéré : $X_t = a + \varepsilon_t$.

Etant donné un réel α tel que $0 < \alpha < 1$ et les observations X_1, X_2, \dots, X_T comme la tendance est constante, on cherche une prévision sous la forme de la constante qui s'ajuste le mieux au sens des moindres carrés pondérés au voisinage de T , c'est à dire la solution du problème de minimisation

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

5.1.1 Définition

La prévision de la série X_t à l'horizon h , $\widehat{X}_T(h)$, fournie par la méthode de lissage exponentiel simple est donnée par

$$\widehat{X}_T(h) = \alpha \sum_{j=0}^{T-1} (1 - \alpha)^j X_{T-j},$$

où α est la constante de lissage.

Démonstration. Soit

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

On commence par chercher les points critiques. Nous avons :

$$S'(a) = -2 \sum_{j=0}^{T-1} (1 - \alpha)^j (X_{T-j} - a).$$

Donc S' s'annule en

$$\hat{a} = \frac{\sum_{j=0}^{T-1} (1 - \alpha)^j X_{T-j}}{\sum_{j=0}^{T-1} (1 - \alpha)^j} = \frac{\alpha \sum_{j=0}^{T-1} (1 - \alpha)^j X_{T-j}}{1 - (1 - \alpha)^T} \underset{T \rightarrow +\infty}{\sim} \alpha \sum_{j=0}^{T-1} (1 - \alpha)^j X_{T-j}$$

selon les variations de S , \hat{a} est le minimum absolu. □

Remarque 5.1.1. • si α est indépendant de h , on note X_T au lieu de $\widehat{X}_T(h)$.

- Cette méthode prend en compte tout le passé de la série chronologique mais en accordant de moins en moins d'importance aux observations les plus éloignées de l'instant T .
- La valeur de α permet de nuancer la remarque précédente
 - α proche de 1, prévision souple i.e. fortement influencée par les observations les plus récentes.
 - $\alpha = 1$, la prévision est alors égale à la dernière valeur observée.
 - α est proche de 0, prévision rigide i.e. influence des observations passées est d'autant plus importante et remonte loin dans le passé.
 - $\alpha = 0$, alors toutes les prévisions sont identiques.

En pratique, on prend $\alpha \in]0, 1[$.

5.1.2 Formule de mise à jour

À partir de la définition, on obtient facilement la formule de mise à jour suivante :

$$\widehat{X}_T(h) = (1 - \alpha)\widehat{X}_{T-1}(h) + \alpha X_T. \quad (5.1)$$

Par conséquent, la connaissance de la prévision $\widehat{X}_{T-1}(h)$, et l'observation X_T suffisent pour calculer immédiatement la prévision suivante $\widehat{X}_T(h)$. Cependant, l'utilisation des formules récursives nécessite d'initialiser la récurrence en prenant par exemple $\widehat{X}_1(h) = X_1$, ou la moyenne de la série observée, on remarque que pour T assez grand la valeur initiale a peu d'influence. Remarquons aussi que la formule (2) permet d'interpréter $\widehat{X}_T(h)$ comme le barycentre de X_T et de $\widehat{X}_{T-1}(h)$ affectés respectivement des masses α et $1 - \alpha$. L'influence de la dernière observation est d'autant plus grande que α est proche de 1.

5.1.3 Choix du α

Un problème important en pratique est le choix du α qui est en général très subjectif et varie selon le contexte de l'étude et/ou le type de prévision souhaité. En pratique,

- Si on veut une prévision souple, on choisira $\alpha \in [0.7; 0.99]$.
- Si on veut une prévision rigide, on choisira $\alpha \in [0.01; 0.3]$.
- Sinon, une autre solution, dictée par les données, consiste à choisir α minimisant la somme des carrés des erreurs de prévision aux dates $1, \dots, T$ à h .

$$\sum_{h=1}^{T-h} (X_{t+h} - \widehat{X}_t(h))^2.$$

La Figure 5.1 illustre le comportement du lissage exponentiel simple en prenant la constante de lissage égale à 0.9 (lissage rigide) ou 0.2 (lissage souple).

La Figure 5.2 illustre l'influence de l'initialisation de l'équation de mise à jour. Sur la figure, la série initiale apparaît en noire. La courbe rouge correspond au lissage en prenant $\widehat{X}_1(h) = (X_1 + X_2)/2$ et la courbe verte correspond au lissage en prenant $\widehat{X}_1(h) = \bar{X}$. On s'aperçoit que quelle que soit la valeur initiale, la prévision est la même au bout d'un certain temps. En revanche, ce temps au bout duquel les prévisions coïncident est d'autant plus petit que α est proche de 0.

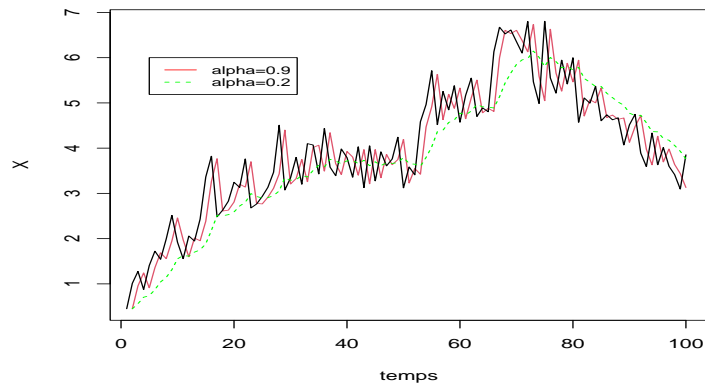


Figure 5.1

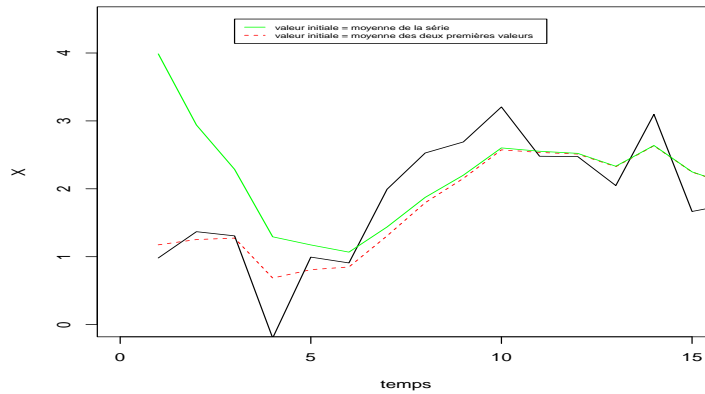


Figure 5.2

5.2 Lissage exponentiel double

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

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

où (\hat{a}_1, \hat{a}_2) sont solution de

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

Lemme 5.2.1. Les solutions du problème de minimisation ci-dessus sont (asymptotiquement)

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

où

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

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

Démonstration. Notons

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

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

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

Nous cherchons donc (a_1, a_2) solutions du système

$$\begin{cases} -\sum_{j=0}^{T-1} (1-\alpha)^j X_{T-j} + \sum_{j=0}^{T-1} (1-\alpha)^j a_1 + \sum_{j=0}^{T-1} (1-\alpha)^j j a_2 = 0 \\ -\sum_{j=0}^{T-1} (1-\alpha)^j j X_{T-j} + \sum_{j=0}^{T-1} (1-\alpha)^j j a_1 + \sum_{j=0}^{T-1} (1-\alpha)^j j^2 a_2 = 0. \end{cases} \quad (5.2)$$

Rappelons les formules

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

On obtient donc, en remplaçant certaines sommes partielles de séries par leurs limites,

$$\begin{cases} -\alpha \sum_{j=0}^{T-1} (1-\alpha)^j X_{T-j} + a_1 + \frac{1-\alpha}{\alpha} a_2 = 0 \\ -\alpha^2 \sum_{j=0}^{T-1} (1-\alpha)^j j X_{T-j} + (1-\alpha)a_1 + \frac{(1-\alpha)(2-\alpha)}{\alpha} a_2 = 0. \end{cases} \quad (5.3)$$

Posons (pour tout T)

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

Nous remarquons que

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

Le système devient alors

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

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

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

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

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

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

Les formules de mise à jour sont

$$\begin{aligned} L_1(T) &= (1 - \alpha)X_T + \alpha L_1(T - 1) \\ L_2(T) &= (1 - \alpha)L_1(T - 1) + \alpha L_2(T - 1) \\ &= (1 - \alpha)^2 X_T + \alpha(1 - \alpha)L_1(T - 1) + \alpha L_2(T - 1). \end{aligned}$$

En utilisant les equations précédentes, on obtient :

$$\begin{aligned} \hat{a}_1(T) &= 2[(1 - \alpha)X_T + \alpha L_1(T - 1)][(1 - \alpha)^2 X_T + \alpha(1 - \alpha)L_1(T - 1) + \alpha L_2(T - 1)] \\ &= (1 - \alpha^2)X_T + \alpha(1 + \alpha)L_1(T - 1) - \alpha L_2(T - 1), \end{aligned}$$

en tenant compte du fait que :

$$\begin{aligned} L_1(T) &= \hat{a}_1(T) \frac{\alpha}{1 - \alpha} \hat{a}_2(T) \\ L_2(T) &= \hat{a}_1(T) \frac{2\alpha}{1 - \alpha} \hat{a}_1(T) \end{aligned}$$

on trouve

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

Les formules analogues pour $\hat{a}_2(T)$ sont :

$$\begin{aligned} \hat{a}_2(T) &= \frac{1 - \alpha}{\alpha} [\alpha(1 - \alpha)X_T + \alpha^2 L_1(T - 1) - \alpha L_2(T - 1)] \\ &= (1 - \alpha)^2 X_T + \alpha(1 - \alpha)L_1(T - 1) - (1 - \alpha)L_2(T - 1) \\ &= (1 - \alpha)^2 X_T - (1 - \alpha)^2 \hat{a}_1(T - 1) + (2\alpha - \alpha^2) \hat{a}_2(T - 1). \end{aligned}$$

Pour calculer récursivement les coefficients \hat{a}_1 et \hat{a}_2 on utilise alors les formules suivantes.

$$\begin{cases} \hat{a}_1(T) = \hat{a}_1(T - 1) + \hat{a}_2(T - 1) + (2\alpha - \alpha^2)(X_T - \hat{X}_{T-1}(1)) \\ \hat{a}_2(T) = \hat{a}_2(T - 1) + \alpha^2(X_T - \hat{X}_{T-1}(1)), \end{cases}$$

avec l'initialisation

$$\begin{cases} \hat{a}_1(0) &= X_1 \\ \hat{a}_2(0) &= X_2 - X_1. \end{cases}$$

Exemple 5.2.1. Calcul d'un lissage exponentiel double ($\alpha = 0.6$).

Dates	Données	Lissage1	Lissage2	a1	a2	Prédiction
Janvier	52					
Février	50	52				
Mars	54	51.2	52	50.40	0.53	50.93
Avril	55	52.32	51.68	52.96	-0.43	52.53
Mai	58	53.39	51.94	54.84	-0.97	53.87
Juin	57	55.24	52.52	57.96	-1.81	56.15
Juillet	60	55.94	53.61	58.27	-1.55	56.72
Aout	61	57.56	54.54	60.58	-2.01	58.57
Septembre	64	58.94	55.75	62.13	-2.13	60.00
Octobre	62	60.96	57.02	64.90	-2.63	62.27
Novembre	64	61.38	58.60	64.16	-1.85	62.31
Décembre	65	62.43	59.71	65.15	-1.81	63.34

5.3 Méthode de Holt-Winters

5.3.1 Méthode non saisonnière

On cherche, au voisinage de l'instant T , à ajuster une droite d'équation

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

La prévision en $T + h$ sera

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

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

$$\begin{cases} \hat{a}_1(T) = \alpha X_T + (1 - \alpha)(\hat{a}_1(T - 1) + \hat{a}_2(T - 1)) \\ \hat{a}_2(T) = \beta(\hat{a}_1(T) - \hat{a}_1(T - 1)) + (1 - \beta)\hat{a}_2(T - 1). \end{cases} \quad (5.4)$$

Remarque 5.3.1. Cette méthode est plus souple que la précédente. Le α joue un rôle dans l'estimée de l'ordonnée en T 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 5.3.1. *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*

$$\begin{cases} \alpha &= 1 - (\alpha')^2 \\ \beta &= \frac{1 - \alpha'}{1 + \alpha'} \end{cases} \quad (5.5)$$

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 (5.4) sous la forme (on remarque au préalable que $(1 - \alpha')(2 - (1 - \alpha')) = 1 - (\alpha')^2$,

$$\begin{aligned} \hat{a}_1(T) &= (\alpha')^2(\hat{a}_1(T-1) + \hat{a}_2(T-1)) + (1 - (\alpha')^2)(X_T - \hat{X}_{T-1}(1) + \hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ &= (\alpha')^2(\hat{a}_1(T-1) + \hat{a}_2(T-1)) + (1 - (\alpha')^2)X_T \\ \hat{a}_2(T) &= \hat{a}_2(T-1) + \frac{(1 - \alpha')^2}{1 - (\alpha')^2}(\hat{a}_1(T) - (\alpha')^2(\hat{a}_1(T-1) + \hat{a}_2(T-1))) \\ &\quad - (1 - (\alpha'))^2(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ &= \hat{a}_2(T-1) + \left(\frac{1 - \alpha'}{1 + \alpha'}\right)\hat{a}_1(T) \\ &\quad + (\hat{a}_1(T-1) + \hat{a}_2(T-1))\frac{(1 - \alpha')}{1 + \alpha'}(-(\alpha')^2 - (1 - \alpha')(1 + \alpha')) \\ &= \hat{a}_2(T-1) + \left(\frac{1 - \alpha'}{1 + \alpha'}\right)\hat{a}_1(T) - (\hat{a}_1(T-1) + \hat{a}_2(T-1))\left(\frac{1 - \alpha'}{1 + \alpha'}\right) \\ &= (1 - \beta)\hat{a}_2(T-1) + \beta(\hat{a}_1(T) - \hat{a}_1(T-1)) \end{aligned}$$

On retrouve bien la formule de mise à jour (5.4) avec $\alpha = 1 - (\alpha')^2$, $\beta = \frac{1 - \alpha'}{1 + \alpha'}$. \square

5.3.2 Méthode saisonnière additive

On cherche, au voisinage de n , à ajuster une courbe d'équation

$$y_t = a_1 + a_2(t - T) + S_t$$

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

$$\begin{cases} \hat{a}_1(T) = \alpha(X_T - \hat{S}_{T-p}) + (1 - \alpha)(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ \hat{a}_2(T) = \beta(\hat{a}_1(T) - \hat{a}_1(T-1)) + (1 - \beta)\hat{a}_2(T-1) \\ \hat{S}_T = \delta(X_T - \hat{a}_1(T)) + (1 - \delta)\hat{S}_{T-p} \end{cases}$$

La prévision prend la forme

$$\begin{cases} \hat{X}_T(h) = \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-p}, 1 \leq h \leq p \\ \hat{X}_T(h) = \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-2p}, p+1 \leq h \leq 2p \\ \dots \qquad \qquad \qquad \dots \end{cases}$$

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

$$\begin{cases} \hat{a}_1(p+1) = & X_{p+1} \\ \hat{a}_2(p+1) = & \frac{X_{p+1} - X_1}{p} \\ \hat{S}_j = & X_j - (X_1 + (p-1)\hat{a}_2(p+1)), 1 \leq j \leq p. \end{cases}$$

5.3.3 Méthode saisonnière multiplicative

On cherche, au voisinage de n , à ajuster une courbe d'équation

$$y_t = (a_1 + a_2(t - T)) \times S_t$$

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

$$\begin{cases} \hat{a}_1(T) = (1 - \alpha) \frac{X_T}{\hat{S}_{T-p}} + \alpha(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ \hat{a}_2(T) = (1 - \beta)(\hat{a}_1(T) - \hat{a}_1(T-1)) + \beta\hat{a}_2(T-1) \\ \hat{S}_T = (1 - \delta) \frac{X_T}{\hat{a}_1(T)} + \delta\hat{S}_{T-p}. \end{cases}$$

La prévision prend la forme

$$\begin{cases} \hat{X}_T(h) = (\hat{a}_1(T) + h\hat{a}_2(T)) \times \hat{S}_{T+h-T}, 1 \leq h \leq p \\ \hat{X}_T(h) = (\hat{a}_1(T) + h\hat{a}_2(T)) \times \hat{S}_{T+h-2T}, p+1 \leq h \leq 2p \\ \dots \qquad \qquad \qquad \dots \end{cases}$$

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 5.3.2. 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.

5.4 Mise en œuvre sous R

Les méthodes de lissages exponentiels pour une série temporelle X sont disponibles sous R , grâce à la fonction `HoltWinters`.

- Pour faire un lissage exponentiel simple :

```
xlisse <- HoltWinters(x, alpha=alpha, beta=FALSE, gamma=FALSE),
```

- un lissage exponentiel double paramètre $1 - \alpha'$:
`xlisse <- HoltWinters(X, alpha=alpha, beta=beta, gamma=FALSE)`
- Pour un lissage de Holt-Winters sans composante saisonnière :
`xlisse <- HoltWinters(x, alpha=alpha, beta=beta, gamma=FALSE),`
- un lissage Holt-Winters additif :
`xlisse <- HoltWinters(x, alpha=alpha, beta=beta, gamma=gamma, seasonal="add"),`
- un lissage Holt-Winters multiplicatif :
`xlisse <- HoltWinters(x, alpha=alpha, beta=beta, gamma=gamma, seasonal="mul").`
 La visualisation et l'affichage des résultats peuvent être réalisés à l'aide des commandes :
 - `summary(xlisse)` : description de l'objet `xlisse` obtenu par la fonction `HoltWinters`,
 - `plot(xlisse)` : représentation des valeurs observées et des valeurs lissées,
 - `(xlisse$fitted[,1]); plot(xlisse,p)` : représentation des valeurs observées et des valeurs lissées,
 - `plot(xlisse)` : Représentation des valeurs observées + valeurs lissées.
- `(predict(xlisse, n.ahead=h)` : Représentation des prévisions à horizon h .

Notons que X doit être un objet de type série temporelle,

Remarque 5.4.1. • Notons que X doit être un objet de type série temporelle,

- à la place de "add" et "mul", on peut entrer n'importe quel début des mots "additive" et "multiplicative" (par exemple "addit" et "multip").

5.5 Exercices et feuille de travaux pratiques 4

Feuille de travaux pratique 4

Exercice 14. En utilisant le langage R :

1. Calculer le lissage exponentiel simple de la série simple ($\alpha = 0.4$) suivante :
 $X_t : 25, 29, 24, 21, 26, 23, 27, 25, 21, 24, 26, 29, 25.$
2. Tracer le graphe de la série et superposer son lissage exponentiel avec couleur différente.

Correction de l'exercice 14

```

1 #1)
2 x1=c(25, 29 ,24, 21, 26, 23, 27, 25, 21, 24, 26, 29,25)
3 x2=ts(x1)
4 ll=HoltWinters(x2, alpha=0.4, beta=F, gamma=F)
5 ll

```



```
6 fitted(l1)[,1]
7
```

Time Series : Start = 2 End = 13 Frequency = 1 [1] 25.00000 26.60000 25.56000 23.73600
24.64160 23.98496 25.19098 25.11459 23.46875 23.68125 [11] 24.60875 26.36525

```
1 # 2)
2 X11()
3 plot(fitted(liss1))
4 K=fitted(liss1)[,1]
5 K
6 X11()
7 plot(K)
8 p1=predict(liss1 ,n.ahead=12)
9 p1
10 X11()
11 plot(p1)
12 X11()
13 $x_p <- ts(c(x1,p1))$
14 $x_p$
15 X11()
16 plot($x_p$,xlab="Lissage et pr\{e}vision par lissage exponentiel
17 simple")
18 lines(K,col="red")
19 lines(p1,col="green")
```

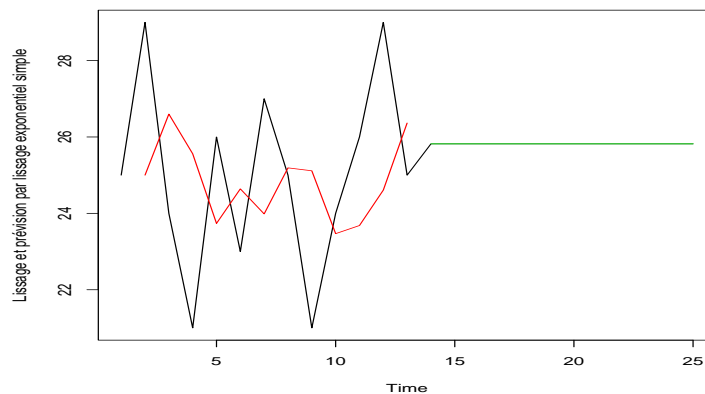


Figure 5.3

Exercice 15. Lissage et prévision de données simulées :

1. Pour chaque série, du feuille d'exercice 1 effectuer la prévision $\hat{X}_{t,1}$ ($1 \leq t \leq 99$) par lissage exponentiel simple, double et HoltWinters saisonnier additif et multiplicatif. Tester

différentes valeurs des paramètres de lissage α , β et γ , et représenter graphiquement la série ainsi que la prévision. Essayer de déterminer graphiquement le lissage le plus adaptée pour chaque série.

2. Calculer pour chaque prévision effectuée la racine des moyennes d'erreurs quadratique (RMSE).
3. Sélectionner le meilleur paramètre à l'aide RMSE pour chaque modèle.

Corection de l'exercice 15

```

1  par(mfrow=c(2,3))
2
3  BRUIT=rnorm(100)
4  BRUIT=ts(BRUIT,1930)
5  X1=ts(BRUIT[1:70],1930)
6  liss1=HoltWinters(X1, beta=F, gamma=F)
7  K=fitted(liss1)[,1]
8  p1=predict(liss1,n.ahead=30)
9  p1=ts(p1)
10 x_p <- ts(c(K,p1), start=1930)
11 Err1=sqrt(sqrt(sum((p1-BRUIT[71:100])^2)))
12 plot(BRUIT)
13 lines(x_p, col="red", xlab="LES, alpha=r", ylim=c(-3,3))
14
15
16 -----
17
18 liss1=HoltWinters(X1, alpha=0.9, beta=F, gamma=F)
19 liss1
20 K=fitted(liss1)[,1]
21 p1=predict(liss1,n.ahead=30)
22 p1=ts(p1)
23 x_p <- ts(c(K,p1), start=1930)
24 plot(BRUIT)
25 lines(x_p, col="red", xlab="LES ", ylim=c(-3,3))
26 Err2=sqrt(sqrt(sum((p1-BRUIT[71:100])^2)))
27 Err2
28 -----
29
30 liss2=HoltWinters(X1, alpha=0.9, beta=0.1, gamma=F)
31 liss2
32 K=fitted(liss2)[,1]
33 p1=predict(liss2,n.ahead=30)
34 x_p <- ts(c(K,p1), start=1930)

```

```

35 lines(x_p, col="red", xlab="LED")
36 plot(BRUIT)
37 Err4=sqrt(sum((p1-BRUIT[71:100])^2))
38 Err4
39 -----
40 -----
41 liss2=HoltWinters(X1, gamma=F)
42 liss2
43 K=fitted(liss2)[,1]
44 p1=predict(liss2, n.ahead=30)
45 x_p <- ts(c(K, p1), start=1930)
46 plot(BRUIT)
47 lines(x_p, col="red", xlab="LED")
48 Err6=sqrt(sqrt(sum((p1-BRUIT[71:100])^2)))
49 Err6
50 -----
51 -----
52 BRUIT=ts(BRUIT, 1930, f=12)
53 X1=ts(BRUIT[1:70], 1930, f=12)
54 -----
55 -----
56 liss3=HoltWinters(X1, alpha=0.1, beta=0.9, gamma=0.2)
57 liss3
58 K=fitted(liss3)[,1]
59 p1=predict(liss3, n.ahead=30)
60 x_p <- ts(c(K, p1), start=1930, f=12)
61 plot(BRUIT)
62 lines(x_p, col="red", xlab="LED")
63 -----
64 Err7=sqrt(sqrt(sum((p1-BRUIT[71:100])^2)))
65 -----
66 -----
67 liss3=HoltWinters(X1, seasonal="add")
68 liss3
69 K=fitted(liss3)[,1]
70 p1=predict(liss3, n.ahead=30)
71 x_p <- ts(c(K, p1), start=1930)
72 plot(BRUIT)
73 lines(x_p, col="red", xlab="LED")
74 Err10=sqrt(sum(p1-BRUIT[71:100]))
75 -----
76 -----

```

```

77
78 liss4=HoltWinters(X1, alpha=0.5,seasonal="mul")
79 liss4
80 K=fitted(liss4)[,1]
81 p1=predict(liss4,n.ahead=30)
82 x_p <- ts(c(K,p1), start=1930, frequency=12)
83 plot(BRUIT)
84 lines(x_p, col="red",xlab="LED")
85
86 Err11=sqrt(sqrt(sum((p1-BRUIT[71:100])^2)))
87 -----

88
89 er=c(Err1, Err2, Err3, Err4, Err5, Err6, Err7, Err8, Err9, Err10, Err11,
90      Err12, Err13, Err14)
91 min(er); which.min(er)
92 -----

92 t=1:100
93 X=0.5*t+2*BRUIT
94 X1=ts(X[1:70],1930,f=12)
95 par(mfrow=c(5,2))
96
97 liss1=HoltWinters(X1, beta=F, gamma=F)
98 K=fitted(liss1)[,1]
99 p1=predict(liss1,n.ahead=30)
100 p1=ts(p1)
101 x_p <- ts(c(K,p1), start=1930)
102 Err1=sqrt(sqrt(sum((p1-X1[71:100])^2)))
103
104 plot(x_p, col="red",xlab="LES, alpha=r", ylim=c(-3,3))
105 plot(X)
106
107 -----

108
109 liss1=HoltWinters(X1, alpha=0.8, beta=F, gamma=F)
110 liss1
111 K=fitted(liss1)[,1]
112 p1=predict(liss1,n.ahead=30)
113 p1=ts(p1)
114 x_p <- ts(c(K,p1), start=1930)
115 plot(X)
116 lines(x_p, col="red",xlab="LES",ylim=c(-3,3))
117 Err2=sqrt(sqrt(sum((p1-X1[71:100])^2)))

```

```

118 -----
119 liss2=HoltWinters(X1,alpha=0.9,beta=0.1,gamma=F)
120 liss2
121 K=fitted(liss2)[,1]
122 p1=predict(liss2,n.ahead=30)
123 x_p <- ts(c(K,p1),start=1930)
124 lines(x_p,col="red",xlab="LED")
125 plot(X)
126 Err4=sqrt(sum((p1-X1[71:100])^2))
127 -----

128 liss2=HoltWinters(X1,gamma=F)
129 liss2
130 K=fitted(liss2)[,1]
131 p1=predict(liss2,n.ahead=30)
132 x_p <- ts(c(K,p1),start=1930)
133 plot(X)
134 lines(x_p,col="red",xlab="LED")
135 Err6=sqrt(sqrt(sum((p1-X1[71:100])^2)))
136 -----

137 liss3=HoltWinters(X1,alpha=0.9,beta=0.1,gamma=0.5)
138 liss3
139 K=fitted(liss2)[,1]
140 p1=predict(liss2,n.ahead=30)
141 x_p <- ts(c(K,p1),start=1930)
142 plot(X)
143 lines(x_p,col="red",xlab="LED")
144
145 Err9=sqrt(sqrt(sum((p1-X1[71:100])^2)))
146 -----

147
148 liss3=HoltWinters(X1,seasonal="add")
149 liss3
150 K=fitted(liss3)[,1]
151 p1=predict(liss3,n.ahead=30)
152 x_p <- ts(c(K,p1),start=1930)
153 plot(X)
154 lines(x_p,col="red",xlab="LED")
155 Err10=sqrt(sum(p1-X1[71:100]))
156
157 -----
158 liss4=HoltWinters(X1,alpha=0.5,seasonal="mul")
159 liss4

```

```

160 K=fitted(liss4)[,1]
161 p1=predict(liss4,n.ahead=30)
162 x_p <- ts(c(K,p1), start=1930, frequency=12)
163 plot(X)
164 lines(x_p, col="red", xlab="LED")
165
166 Err11=sqrt(sqrt(sum((p1-X1[71:100])^2)))
167
168 -----
169
170 liss2=HoltWinters(X1, alpha= 0.9,seasonal="add")
171 liss2
172 K=fitted(liss2)[,1]
173 p1=predict(liss2,n.ahead=30)
174 x_p <- ts(c(K,p1), start=1930, frequency=12)
175 plot(X)
176 lines(x_p, col="red", xlab="LED")
177
178 Err12=sqrt(sqrt(sum((p1-X1[71:100])^2)))
179 #en ferra la même chose pour toutes les séries
180 -----

```

Exercice 16. Lissage et prévision de données réelles :

- Données uspop : nombre annuel de tâches solaires de 1790 à 1970.
 - Données co2 contient les concentrations en co2 à proximité du volcan Mauna Loa de 1959 à 1997.
 - Données AirPassengers représente l'évolution du nombre mensuel de personne ayant pris l'avion dans le monde \mathbb{R}^1 .
- Effectuer les démarches suivantes :
- Après avoir représenté graphiquement ces données, quel modèle de lissage exponentiel vous semble le mieux approprié ?.
 - Effectuer le lissage exponentiel le mieux appropriés pour chaque série.
 - tester la prédiction des données de la dernière années.

```

1
2 plot(co2)
3 donnees=window(co2, start=1959,end=c(1989,12),freq=12)
4 r=co2[((31*12)+1):468]
5 lisse=HoltWinters(donnees,seasonal="add")
6 K=fitted(lisse)[,1]

```

1. Les trois données précédentes sont contenu dans R est dans l'espace de travail dès que vous lancez R

```

7
8 p1=predict(lisse ,n.ahead=12*8)
9 p1
10 x_p <- ts(c(K,p1) , start=1959, frequency=12)
11 plot(co2)
12 lines(x_p, col="blue")
13 Err15=(sqrt(sum((p1-r)^2)))
14
15 -Prévision pour les années de 1998 à 2007 :
16 liss=HoltWinters(co2 , seasonal="add")
17 liss
18 p2=predict(liss ,n.ahead=10*12)
19 p2
20 Serie=ts(c(liss ,p2) , start=1959, freq=12)
21 plot(Serie)
22 lines(co2 , col="red")

```

Remarque 5.5.1. Lorsqu'aucune valeur n'est précisée pour les constantes de lissage, un algorithme interne à la procédure HoltWinters se charge d'estimer la meilleure constante possible à partir de la série des observations.

Ces remarques permettent de nuancer certains choix automatiques. Par exemple, R effectue par défaut un lissage exponentiel simple avec une constante de lissage égale à 0.8. Il peut être pertinent de choisir une autre valeur d'autant plus que la série est courte.

Conclusion. L'analyse des séries chronologiques est un des objectifs fondamentaux de la statistique. Nous insistons sur le fait, que quelle que soit la méthode utilisée, il faut être vigilant sur les prévisions effectuées qui peuvent être dans certains cas totalement aberrantes (cf. Bensaber et Bleuse-Trillon, p. 123, à propos du modèle de Holt et Winters). Nous conseillons fortement aux praticiens de se limiter aux méthodes qu'ils connaissent lorsqu'ils effectuent leurs prévisions. Le logiciel qu'ils utilisent doit être sûr, contrôlé : il nous est arrivé d'obtenir des résultats différents sur les mêmes données et par la même méthode en employant deux logiciels différents, ou d'aboutir des prévisions manifestement fausses puisque différentes des valeurs données dans les ouvrages de référence.

Chapitre 6

Appendice

6.1 Indices descriptifs d'une série temporelle

6.1.1 Indice de tendance centrale

Moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$$

6.1.2 Indices de dispersion

Variance empirique :

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

6.1.3 Indices de dépendance

Auto-covariance empirique d'ordre h (h dans \mathbb{N}^*) :

$$\hat{\sigma}_n(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n).$$

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 6.1.1. 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.

6.2 Commandes usuelles de R

Aide et fonctions de base.

help(topic), ?topic affiche l'aide relative à topic.

6.2.1 Création de données

Les vecteur

c(...) : fonction combinant les arguments pour former un vecteur.

c(nom1=, nom2=, ...) : fonction générique combinant une suite d'éléments en un vecteur (possibilité d'attribuer des noms).

seq(from, to) : génère une séquence; *by=* spécifie l'incrément; *length=* spécifie la longueur.

seq(1, 20, by=0.5) : séquence par pas.

seq(1, 20, length=13) : séquence par longueur finale.

seq(along=x) : génère une suite 1, 2, ...

length(x) : utile pour les boucles *for*.

rep(x, times) : répète *times* fois la valeur *x*.

Les tableaux et les matrices

array(x, dim=) : crée un tableau multidimensionnel *x*; les éléments de *x* sont répétés si la taille ne correspond pas aux dimensions spécifiées.

matrix(x, nrow=, ncol=) : crée une matrice; les éléments de *x* sont répétés si la taille ne convient pas.

matrix(x, nrow=, ncol=, byrow=T) : crée une matrice est la remplire par ligne.

Data-frame

Les *data-frames* sont des objets hétérogènes composés d'une collection d'objets homogènes de longueur identique.

data.frame(...) : crée un *data frame* avec les arguments (nommés ou non) par exemple :

textttdata.frame(v=1:4, ch=c("a", "b", "c", "d"), lettre="A");

un *data.frame* est un tableau dont les colonnes peuvent être de types différents.

Les listes

Ce sont des objets hétérogènes composés d'une collection d'objets homogènes de longueur pas obligatoirement identiques.

list(nom=x,y) : listes a 2 éléments (ou composantes) x de nom x et y sans nom.

6.2.2 Extraction de données et fonctions de base

Indexer des vecteurs

x[n] : n ième élément.

x[-n] : tous sauf le n ième élément.

x[1:n] : les n premier éléments.

x[-(1:n)] : les éléments de n+1 à la fin.

x[c(1,4,2)] : les éléments 1,4,2.

x["name"] : élément(s) du nom "name".

x[x > 3] : tous les

Fonctions utiles pour les vecteurs

sum(x) : somme des éléments de x.

cumsum(x) : somme cumulée de x.

diff(x) : différence entre chaque élément de x et son prédécesseur.

prod(x) : produit des éléments de x.

mean(x) : moyenne des éléments de x.

median(x) : médiane des éléments de x.

quantile(x, probs=) : quantiles correspondant aux probabilités données.

rank(x) : rang des éléments de x.

var(x) ou *cov(x)* : variance des éléments de x (calculé avec n-1 au dénominateur); si x est une matrice ou un data.frame, la matrice de variance-covariance est calculée.

sd(x) : écart-type.

cor(x) : matrice de corrélation de x (pour une matrice ou une data.frame).

var(x, y) ou *cov(x, y)* : covariance entre x et y, ou entre les colonnes de x et celles de y si ce sont des matrices ou des data frames.

cor(x, y) : coefficient de corrélation linéaire entre x et y, ou matrice de corrélation si ce sont des matrices ou des data frames.

round(x, n) : arrondit les éléments de x à n décimales.

union(x,y), *intersect(x,y)*, *setdiff(x,y)* : union et intersection d'ensembles; *setdiff(x,y)* trouve les éléments de x qui ne sont pas dans y.

abs(x) : valeur absolue.

indexer des matrices

$x[c(6, 1, 9),]$: affichage des lignes 6, 1 et 9.
 $x[, c(6, 1, 9)]$: affichage des colonnes 6, 1 et 9.
 $x[1, -2]$: affichage de la ligne 1 sans la colonne 2.
 $x["name",]$: lignes intitulées "name".
 $x[rowSums(x) > 10,]$: lignes dont la somme est sup

fonctions utiles pour les matrices

$t(x)$: transposée.
 $diag(x)$: diagonale.
 $diag(1:10)$: matrice diagonale d'ordre 10 dont les valeurs sont 1 :10.
 $diag(10)$: identité d'ordre 10.
 $solve(a, b)$: trouve x tel que $a x = b$.
 $solve(a)$: matrice inverse de a .
 $rowsum(x)$: somme par ligne d'une matrice ou d'un objet similaire.
 $colsum(x)$: somme par colonne.
 $rowMeans(x)$: moyenne des lignes d'une matrice.
 $colMeans(x)$: idem pour les colonnes.
 $length(x)$: longueur de x , i.e. son nombre d'éléments.
 $diag(x)$: extraction de la diagonale de x .
 $nrow(x)$: nombre de lignes de x .
 $ncol(x)$: nombre de colonnes de x .
 $dim(x)$: dimension de x .
 $cbind(x, y)$: concaténation par colonne de x et y .
 $rbind(x, y)$: concaténation par ligne de x et y .
 $qr(x)$: décomposition QR de x .
 $chol(x)$: décomposition de Cholesky de x .
 $svd(x)$: décomposition en valeur singulière.
 $eigen(x)$: valeurs propres et vecteurs propres de x .
 $det(x)$: déterminant de x .
 $apply(x, 1, sum)$: somme par ligne de x .
 $apply(x, 2, var)$: variance par colonne de x .
 $data.frame(x)$: onverti x en data-frame.
 $rbind(...)$: combine les arguments par ligne (row).
 $cbind(...)$: combine les arguments par colonne.

Indexer des listes

$x[i]$: le ou les éléments i de la liste (renvoyé(s) sous forme de liste, à la différence des cas suivants.

$x[[n]]$: n ième élément de la liste.

$x[["nom"]]$: l'élément nommé "nom".

$x\$nom$: l'élément nommé "nom".

$li[[2]]$: extraction de la com2 de la liste x .

$li\$nomx$: extraction de la compnomx de la liste x .

$dimnames(x) <- list(nomli, nomco)$: noms des lignes et des colonnes de la matrice x .

$lapply(li, f)$: application de la fonction f a chacune des composantes de la liste.

$unlist(li)$: transformation en vecteur de la liste.

Indexer des data frame

$cbind.data.frame(df, df2)$: Concaténation par colonne.

$names(df)$: nom des colonnes

$df[, 1]$: extraction de la colonne 1.

$df[[1]]$: extraction de la composante 1 (colonne 1 en général).

$df[, "nomcol"]$: extraction de la colonne $nomcol$.

$summary(df)$: résumé colonne par colonne (min,max,moyenne etc...).

$summary(a)$: propose un résumé de a , la plupart du temps un résumé statistique.

6.2.3 Importer et exporter des données

$data(x)$: charge le jeu de données x

$read.table(file)$: lit un fichier stocké sous la forme d'un tableau et crée un objet `data.frame` ; le séparateur par défaut est le caractère espace pour `read.table`, utiliser l'option `header=TRUE` pour que la première ligne soit considérée

`sep=""` : désigne n'importe quel espacement.

`header=TRUE` : pour prendre la première ligne comme titre.

`skip=n` : pour sauter n lignes.

6.2.4 Variables et attributs

`as.array(x)`, `as.data.frame(x)`, `as.numeric(x)`, `as.logical(x)`, `as.character(x)`,
... : conversion de type.

`is.na(x)`, `is.null(x)`, `is.array(x)`, `is.data.frame(x)`, `is.numeric(x)`, `is.character(x)`,
... : teste le type d'un objet.

6.2.5 Graphiques

`plot(x)` : graphique de x (fonction générique ayant des effets différents selon l'objet).

`plot(x, y)` : nuage de points.

`hist(x)` : histogramme des fréquences de x .

`barplot(x)` : diagramme en barre.

`pie(x)` : diagramme circulaire.

`boxplot(x)` : diagramme en boîte [boîte à moustaches].

`plot.ts(x)` : pour une ou des série(s) temporelle(s) (classe "ts"), valeurs de x en fonction du temps.

`qqnorm(x)` : nuage des quantiles observés contre quantiles théoriques ; si x suit une loi normale, une droite ; comparer `qqnorm(rnorm(100))` et `qqnorm(1:100)`

`qqplot(x, y)` : quantiles de y en fonction des quantiles de x

Les paramètres suivants sont communs à de nombreuses fonctions graphiques :

`add=TRUE` : ajoute sur le graphique précédent.

`axes=FALSE` : ne trace pas les axes.

`type="p"` : type de représentation des coordonnées ; "p" : points, "l" : lignes, "b" (both) points et lignes, "o" : idem mais lignes sur les points, "h" : bâtons, "s".

`xlim=`, `ylim=` : limites des zones du graphique, é.g. `xlim=c(1,5)`.

`xlab=`, `ylab=` : titre des axes (caractères).

`main=` : titre du graphique (caractères).

`sub=` : sous-titre du graphique (caractères).

Commandes graphiques de bas niveau

Permettent de compléter un graphique existant (éventuellement vide avec `plot(...,type="n")`).

`points(x, y)` : ajoute des points (`type=` peut être utilisé).

`lines(x, y)` : ajoute des lignes.

`text(x, y, labels, ...)` : ajoute du texte (labels).

`segments(x0, y0, x1, y1)` : trace des segments de (x_0, y_0) à (x_1, y_1) .

`abline(a, b)` : trace une droite (de forme $y=a+b*x$).

`legend(x, y, legend)` : ajoute une légende au point (x, y) avec les symboles donnés par `legend`.

locator(n) : renvoie les coordonnées des clics de la souris après *n* clics sur le graphique.

Paramètres graphiques

col : couleur(s) des symboles et lignes par exemple *col="red", "blue"* .

lwd : largeur des lignes.

pch : type de symboles pour les points (code entier de 1 à 25, ou caractère entre " ").

6.2.6 Modèles et analyses statistiques

lm(formula) : estimation d'un modèle linéaire *formula*.

summary(lm(...)) : donne des informations utiles.

predict(fit, ...) : fait une prédiction à partir du modèle estimé *fit* et de nouvelles données.

coef(fit) : coefficients du modèle estimé.

residuals(fit) : résidus du modèle.

fitted(fit) : valeurs prédites par le modèle.

Pour avoir un quantile utiliser le préfixe *q*. Pour avoir une probabilité utiliser le préfixe *p*, pour avoir une densité utiliser le préfixe *d* et pour générer des nombres aléatoire utiliser le préfixe *r*.

Lois disponibles : *norm, poiss, binom, logis,...*

qt(0.95, 12) : quantile a 95 d'une loi de student (*t*) a 12 ddl.

ppoiss(3, 2) : probabilité cumulée jusqu' a 3 d'une loi de poisson (*poiss*).

df(3, 6, 30) : densité au point 3 d'une loi de sher (*f*) a (6,30) ddl.

rnorm(10, 1, 2) : génération de 10 nombres aléatoires selon une loi normale de moyenne 1 d'écart type 2.

rnorm(n, mean=0, sd=1) : distribution gaussienne (normale).

rt(n, df) : distribution de Student (*t*).

rf(n, df1, df2) : distribution de Fisher–Snedecor (*F*). *idem*

6.2.7 Programmation

Fonctions permettant d'enchaîner des opérations de manière structurée. Pour avoir de l'aide sur ces fonctions, saisir leur nom entre guillemets.

function(arglist) expr : définition de fonction ; *arglist* est une liste d'arguments, *expr* est une expression exécutée.

if(cond) expr : si *cond* est vrai (TRUE), évaluer *expr*.

if(cond) cons.expr else alt.expr : si *cond* est vrai évaluer *cons.expr* sinon évaluer *alt.expr*.

for(var in seq) expr : exécute l'expression pour chaque valeur de *var* prises dans une se-

quence.

while(cond) expr : exécute l'expression tant que la condition est vraie.

repeat expr : répète *expr* en boucle; penser à l'arrêter avec *if(...)* *break*.

break : arrête une boucle *for*, *while* ou *repeat*.

next : arrête l'itération en cours et reprend la boucle (dans le cas de *for*, avec la valeur suivante de la séquence).

ifelse(test, yes, no) : pour chaque ligne/cellule de test, renvoie la valeur *yes* si le test est TRUE, *no* s'il est FALSE, NA s'il est indéterminé.

sample(10, 1:30) : tirage sans remise de 10 nombres parmi les coordonnées du vecteur 1 :30.

function(arglist) expr return(result) : définition de fonction.

Pour l'affichage on utilise *print(a)* : fonction d'affichage d'un objet *a* s'adaptant au type de l'objet.

Bibliographie

- [1] *Abadie et Meslier. Etude de l'utilisation des modèles ARIMA pour la prévision à très court terme de l'énergie journalière produite par Electricité de France. RAIRO Recherche opérationnelle Operations research n1, vol.13 p 37-54, février 1979.*
- [2] *Arthur. Charpentier. Cours de séries temporelles, Volume1 : Introduction à la théorie des processus en temps discret. [3] Axe. Sectoriel. Analyses des Chroniques, 2007-2008.*
- [4] *Bernard. Rapacchi. Les séries chronologiques, Centre Interuniversitaire de Calcul de Grenoble, 18 août 1993.*
- [5] *C.Do. Séries Temporelles Linéaires, Année 2004-2005.*
- [6] *C.Do. Prévisions dans les ARMA et les ARIMA, Mars 2003.*
- [7] *Céline Lévy-Leduc. Introduction à l'Etude des Séries Chronologiques, 13 mai 2008.*
- [8] *Corinne. Perraudin. Séries Chronologiques, Année 2004-2005.*
- [9] *Didier. Delignières. Séries temporelles-Modèles ARIMA, Mars 2000.*
- [10] *Eric. Moulines et François Roueff. Analyse des Séries Temporelles et Applications, 15 septembre 2010.*
- [11] *M. Gevers et L. Vandendorpe. Processus stochastiques, estimation et prédiction, Université catholique de Louvain. Faculté des Sciences Appliquées.*
- [12] *Martin. Filtrage de Kalman d'une série temporelle saisonnière, Application à la prévision de consommation d'électricité. Revue de statistique appliquée v. XLVII p 69 – 86, 1999.*
- [13] *Yves. Aragon. Introduction aux Séries temporelles, Septembre 2004.*