



MI0B702T
Partie : Séries Chronologiques

Polycopié de cours

Agnès Lagnoux

lagnoux@univ-tlse2.fr

Table des matières

1	Introduction	1
1.1	Série chronologique : vocabulaire et exemples	1
1.1.1	Définition	1
1.1.2	Description d'une série chronologique	2
1.1.3	Objectifs principaux	4
1.2	Description schématique de l'étude complète d'une série chronologique	5
1.2.1	Correction des données	6
1.2.2	Observation de la série	6
1.2.3	Modélisation	7
1.2.4	Analyse de la série à partir de ses composantes	8
1.2.5	Diagnostic du modèle	8
1.2.6	Prédiction	8
2	Modélisation déterministe	9
2.1	Le modèle additif	9
2.2	Le modèle multiplicatif	10
2.3	Les modèles mixtes	11
2.4	Choix du modèle	11
3	Moyennes mobiles	15
3.1	Exemple introductif : la moyenne mobile arithmétique	15
3.2	Définitions des moyennes mobiles	16
3.3	Propriétés d'un lissage par moyenne mobile	18
3.3.1	Effet d'une moyenne mobile sur une tendance	18
3.3.2	Effet d'une moyenne mobile sur une composante saisonnière	20
3.3.3	Effet d'une moyenne mobile sur les fluctuations irrégulières	21
3.3.4	Choix pratique de l'ordre d'une moyenne mobile	22
4	Décomposition d'une série chronologique	25
4.1	Un exemple introductif : Cas d'une tendance sensiblement constante	25
4.2	L'exemple étudié	27
4.3	La série lissée par moyenne mobile	27
4.4	Estimation de la saisonnalité	29
4.5	Estimation de la tendance	30
4.5.1	Quelques méthodes d'ajustement tendanciels	30
4.5.2	Retour à l'exemple	32
4.6	Prévision des valeurs futures	32
4.7	Remarque : cas du modèle multiplicatif	33
4.8	Analyse des résidus	34
4.9	Un autre exemple	36

5	POUR ALLER PLUS LOIN - Prévision par lissage exponentiel	41
5.1	Les lissages exponentiels	41
5.1.1	Le lissage exponentiel simple	41
5.1.2	Le lissage exponentiel double	44
5.2	La méthode de Holt-Winters	46
5.2.1	La méthode non saisonnière	47
5.2.2	La méthode saisonnière additive	47
5.2.3	La méthode saisonnière multiplicative	49
6	Annexes	51

Chapitre 1

Introduction

1.1 Série chronologique : vocabulaire et exemples

1.1.1 Définition

La théorie des séries chronologiques (ou temporelles) abordée dans ce cours est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie. On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but de **décrire**, **expliquer** puis **prévoir** ce phénomène dans le futur. On dispose ainsi d'observations à des dates différentes, c'est à dire d'une suite de valeurs numériques indicées par le temps.

Exemple : On peut songer par exemple à l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel (comme le nombre de taches solaires).

Cette suite d'observations d'une famille de variables aléatoires réelles notées $(X_t)_{t \in \Theta}$ est appelée **série chronologique** (ou temporelle). Dans la suite de ce cours, nous la noterons

$$(X_t)_{t \in \Theta} \text{ ou } (X_t, t \in \Theta),$$

où l'ensemble Θ est appelé **espace des temps** qui peut être

- **discret** (nombre de voyageurs SNCF quotidien, 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 issues de la famille $(X_t)_{t \in \Theta}$ où $\Theta \subset \mathbb{Z}$ (le plus souvent $\Theta = \mathbb{Z}$). Ainsi si h est l'intervalle de temps séparant deux observations et t_0 l'instant de la première observation, on a le schéma suivant

$$\begin{array}{cccc} t_0 & t_0 + h & \dots & t_0 + (T-1)h \\ \downarrow & \downarrow & \dots & \downarrow \\ X_{t_0} & X_{t_0+h} & \dots & X_{t_0+(T-1)h} \\ \downarrow & \downarrow & \dots & \downarrow \\ X_1 & X_2 & \dots & X_T \end{array}$$

- **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. Les méthodes présentées dans ce cadre sont différentes de celles pour les séries chronologiques à temps discret et présentées dans la suite.

Dans ce cours, nous considérerons uniquement des **processus stochastiques** $(X_t)_{t \in \Theta}$ à temps **discret** et **unidimensionnels** : chaque observation X_t est un réel. On peut également s'intéresser à des séries chronologiques multidimensionnelles, c'est à dire telles que X_t soit un vecteur de \mathbb{R}^d .

Les Figures 1.1 et 1.2 présentent différents exemples de séries chronologiques.

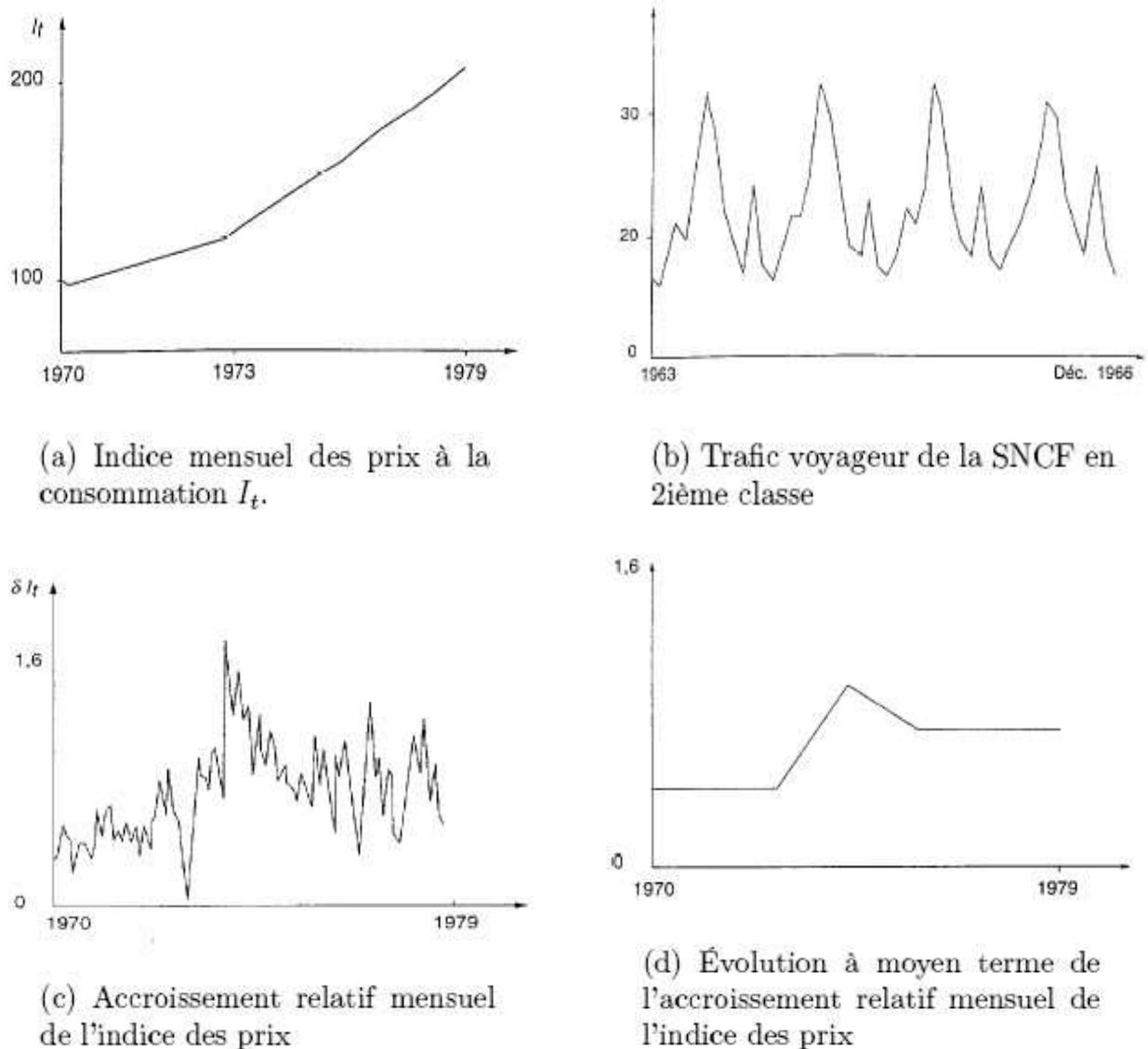


FIGURE 1.1 – Exemples de séries chronologiques(1)

1.1.2 Description d'une série chronologique

On considère qu'une série chronologique (X_t) est la résultante de différentes composantes fondamentales :

- la **tendance** (ou trend) (Z_t) représente l'évolution à long terme de la série étudiée. Elle traduit le

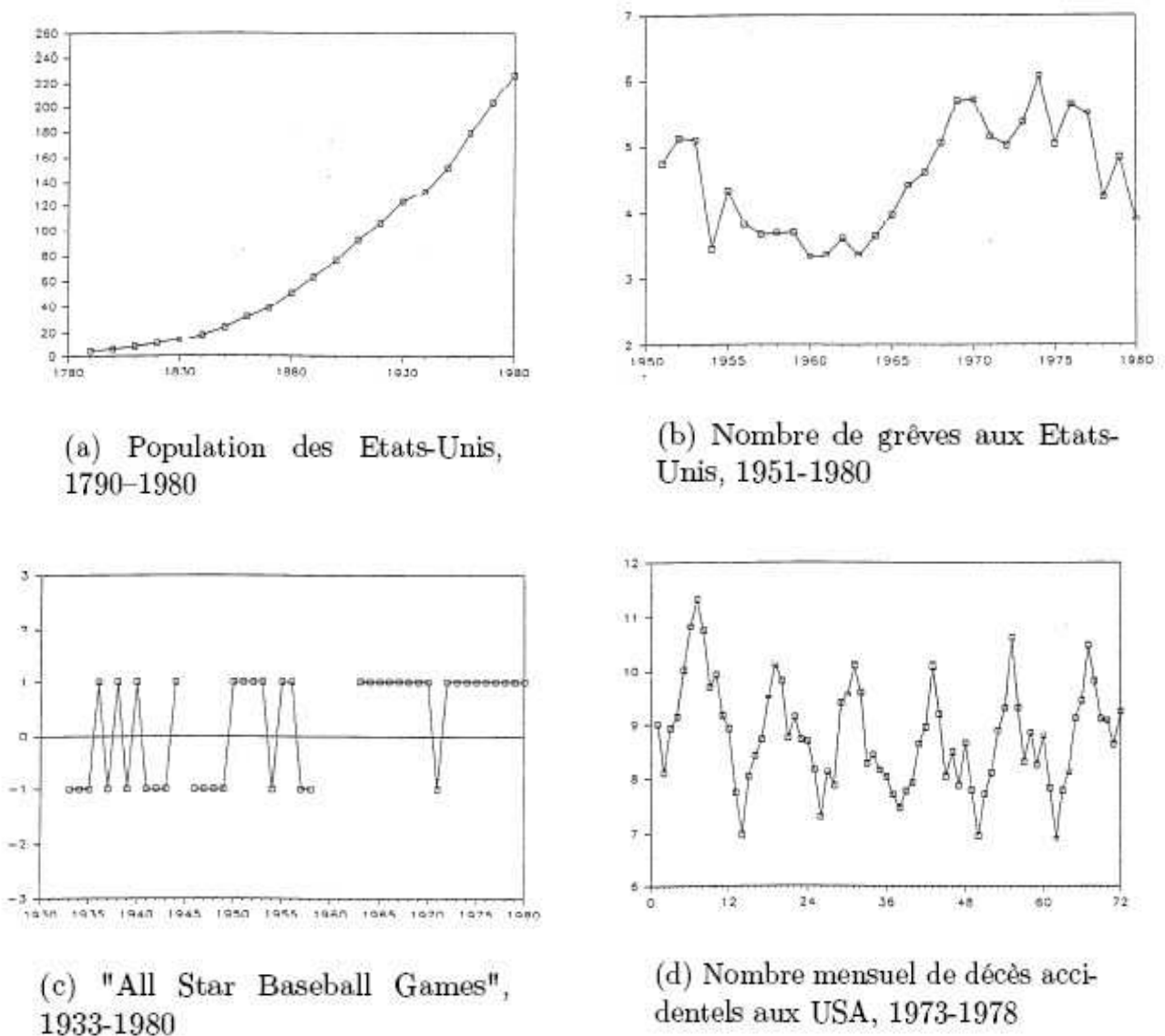


FIGURE 1.2 – Exemples de séries chronologiques (2)

comportement "moyen" de la série.

Par exemple, la série a) de la Figure 1. a tendance à augmenter de façon linéaire.

- 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.

Par exemple, la série b) de la Figure 1. présente des cycles réguliers au cours du temps et de même amplitude.

L'étendue de cet intervalle qui est constante est appelée **période** et sera notée P dans la suite. La plupart du temps, on suppose que la composante saisonnière est constante sur chaque période P , c'est-à-dire

$$S_{t+P} = S_t, \quad \forall t.$$

Cela revient à dire que l'effet net du saisonnier sur une période est nul; ce qui est naturel puisqu'il

est repris dans la tendance générale de la série chronologique. Il s'agit là du modèle le plus simple dans lequel le saisonnier est caractérisé par P coefficients c_1, \dots, c_P . Lorsque $P = 4$, la série est trimestrielle; lorsque $P = 12$, la série est mensuelle...

- la **composante résiduelle** (ou bruit ou résidu) (ϵ_t) correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. On parle aussi d'aléas. Ces **erreurs** sont des variables aléatoires centrées. On considère le plus souvent un bruit blanc, c'est-à-dire une suite de v.a.r. telles que

$$\mathbb{E}(\epsilon_t) = 0 \quad \text{et} \quad \mathbb{E}(\epsilon_t \epsilon_{t'}) = \sigma^2 \delta_{tt'}.$$

Les v.a.r. sont alors non corrélées et lorsque le bruit blanc est **gaussien** c'est-à-dire que

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2),$$

on a de plus l'indépendance des ϵ_t .

Par exemple, la série c) de la Figure 1. a un comportement assez irrégulier : il y a comme une sorte de bruit de faible amplitude qui perturbe les données.

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

- Des **phénomènes accidentels** (grèves, conditions météorologiques exceptionnelles, crash financier) peuvent notamment intervenir.

Par exemple, la série d) de la Figure 1. présente deux cassures.

- Une autre composante parfois étudiée de manière spécifique a trait au **phénomène cyclique** : c'est souvent le cas en climatologie et en économie (exemple : récession et expansion...). Il s'agit d'un phénomène se répétant mais contrairement à la saisonnalité sur des durées qui ne sont pas fixes et généralement plus longues. Sans informations spécifiques, il est généralement très difficile de dissocier tendance et cycle.

Dans le cadre de ce cours, la composante correspondant aux phénomènes accidentels sera intégrée aux fluctuations irrégulières de la série et la composante tendance regroupera à la fois la tendance et le cycle.

1.1.3 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... Plus généralement, nous pouvons déjà poser quelques problèmes lorsqu'on étudie une série chronologique.

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 chronologique à partir de ses valeurs observées jusqu'au temps T : X_1, X_2, \dots, X_T . La prédiction de la série chronologique au temps $t + h$ est notée $\hat{X}_T(h)$ et, en général, est différente de la valeur réelle X_{T+h} que prend la série au temps $T + h$. Pour mesurer cette différence, on définira l'**erreur de prédiction** par la différence

$\hat{X}_T(h) - X_{T+h}$ "en moyenne" avec l'idée que plus h est grand, plus grande est l'erreur. L'intervalle de précision, défini par les valeurs $\hat{X}_T^{(1)}(h)$ et $\hat{X}_T^{(2)}(h)$, est susceptible de contenir la valeur inconnue X_{T+h} . La qualité de la prédiction pourra être mesurée en se basant sur 80% des observations, puis en simulant une prédiction sur les 20% d'observations restantes. Cette technique est aussi utile pour :

- les séries qui contiennent des "trous"
- mesurer l'effet d'un phénomène accidentel (erreur,...)

Un autre problème intéressant est la **détection de ruptures résultantes**, par exemple, d'un changement de politique (économique). Ces ruptures peuvent être de deux ordres : une rupture de niveau (par exemple, le cours du PNB espagnol a été fortement modifié en raison de la crise pétrolière de 1973) ou une rupture de pente. La prévision de ces dates de rupture est bien évidemment très importante.

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 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).

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

Comme nous venons de le voir, l'un des objectifs principaux de l'étude d'une série chronologique est la prévision des valeurs futures de cette série. Pour cela, on a besoin de connaître ou tout au moins de **modéliser** le mécanisme de production de la série chronologique.

Notons que les variables X_t ne sont le plus souvent ni indépendantes (on peut s'attendre en effet à ce que des observations relativement proches dans le temps soient liées) ni identiquement distribuées (dans la plupart des cas, le phénomène évolue, se modifie au cours du temps ce qui entraîne que les variables le décrivant ne sont pas équidistribuées). Cela nécessite des méthodes statistiques de traitement et de modélisation spécifiques puisqu'en particulier dans un cadre standard (celui de la description d'un échantillon) les méthodes statistiques classiques sont basées sur des hypothèses d'indépendance.

Schématiquement, les principales étapes de traitement d'une série chronologique sont les suivantes :

1. correction des données
2. observation de la série
3. modélisation (avec un nombre fini de paramètres)
4. analyse de la série à partir de ses composantes

5. diagnostic du modèle - ajustement au modèle
6. prédiction (= prévision)

1.2.1 Correction des données

Avant de se lancer dans l'étude d'une série chronologique, il est souvent nécessaire de traiter, modifier les données brutes. Par exemple,

- évaluation de données manquantes, remplacement de données accidentelles,...
- découpage en sous-séries;
- standardisation afin de se ramener à des intervalles de longueur fixe. Par exemple, pour des données mensuelles, on se ramène au mois standard en calculant la moyenne journalière sur le mois (total des observations sur le mois divisé par le nombre de jours du mois);
- transformation des données : pour des raisons diverses, on peut être parfois amenés à utiliser des données transformées. Par exemple en économie, on utilise la famille de transformations de Box-Cox :

$$Y_t = \frac{1}{\lambda} \left[(X_t)^\lambda - 1 \right], \quad \lambda \in \mathbb{R}^* .$$

1.2.2 Observation de la série

Une règle générale en Statistique Descriptive consiste à commencer par regarder les données avant d'effectuer le moindre calcul. Ainsi, une fois la série corrigée et prétraitée, on trace son graphique c'est à dire la courbe de coordonnées (t, X_t) (cf. Figure 1.3 représentant le trafic SNCF sur différentes années). L'observation de ce graphique est souvent une aide à la modélisation de la série chronologique et permet de se faire une idée des différentes composantes de la série chronologique que nous avons rapidement mentionnées en Section 1.1.2.

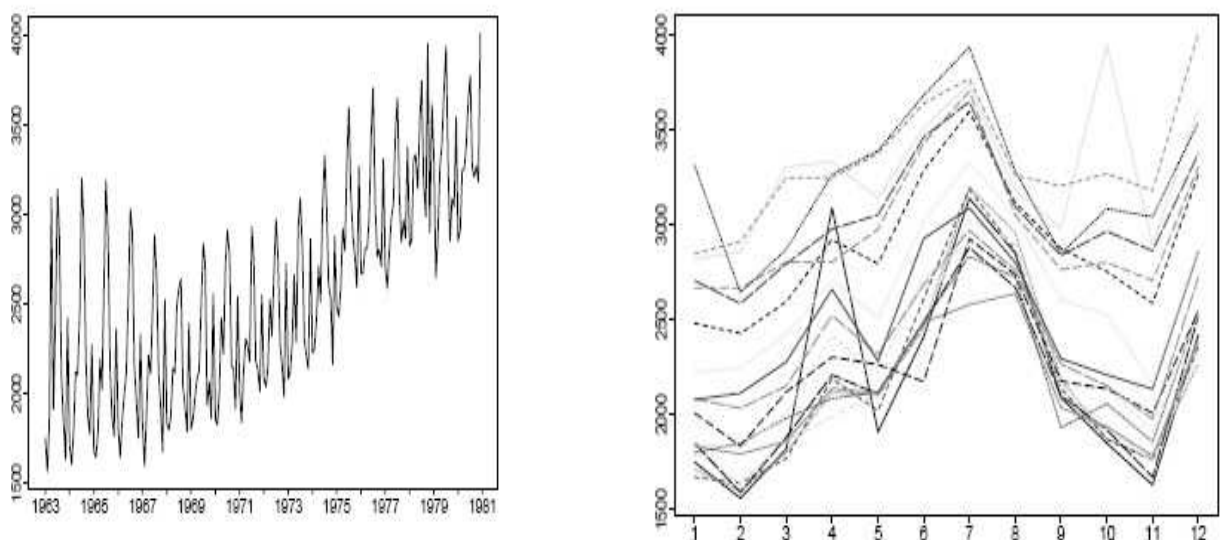


FIGURE 1.3 – Évolution du trafic voyageur SNCF de 1960 à 1980 (à gauche) et évolution annuelle (à droite)

- L'observation du graphique de gauche de la Figure 1.3 indique par exemple que le nombre de voyageurs SNCF a augmenté de manière régulière au cours du temps. De manière générale, la courbe peut indiquer un "mouvement" à moyen terme de croissance ou décroissance (linéaire,

quadratique...) révélant la présence d'une composante déterministe dans la série appelée **tendance** (ou *trend*) qui exprime donc l'évolution générale à moyen ou long terme de la série, du phénomène étudié. Par exemple, si on admet le scénario d'un réchauffement de la planète, la courbe des températures moyennes indique un mouvement de croissance à moyen terme.

- Le graphe de la série peut encore faire apparaître une périodicité dans les valeurs observées révélant la présence d'un phénomène dit **saisonnier**. Les variations saisonnières sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz, vente de bois avant l'hiver...) ou encore par des activités économiques et sociales (fêtes, vacances, soldes,...). Elles sont de nature périodique c'est à dire qu'il existe un entier p , appelé période, tel que $S_t = S_{t+p}$, pour tout t et cette composante est donc entièrement déterminée par ses p premières valeurs S_1, S_2, \dots, S_p . 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 (cf. Figure 1.3 à droite). Sur ce graphique, on voit bien une similarité des différentes courbes annuelles liée aux saisons météorologiques : on constate par exemple un pic au mois de juin...

- Bien entendu, on constate sur les deux figures des fluctuations plus ou moins importantes que l'on appelle **irrégularités** ou **mouvements résiduels**. Ces fluctuations irrégulières sont dues à des facteurs exceptionnels pour la plupart imprévisibles (exemple : grève, risque de guerre...), ont souvent un effet de courte durée et de faible intensité et sont de nature aléatoire (ce qui signifie ici dans un cadre purement descriptif qu'elles ne sont pas expliquées). On regroupe donc généralement ces variations dans une composante aléatoire représentant les effets non expliqués ou encore l'erreur au modèle.

- Nous remarquons aussi un phénomène accidentel : sur l'une des courbes de la Figure 1.3 de droite (il s'agit de l'année 1963), on voit un pic "anormalement" élevé au mois d'avril. On peut également s'intéresser à l'impact de mai 1968 sur le nombre de voyageurs.

Les modèles présentés dans la section suivante tiennent compte uniquement des trois premières composantes (tendance, saisonnalité et fluctuations irrégulières); les phénomènes accidentels étant intégrés au terme de fluctuations irrégulières.

1.2.3 Modélisation

Un **modèle** est une image simplifiée de la réalité qui vise à traduire les mécanismes de fonctionnement du phénomène étudié et permet de mieux les comprendre. Un modèle peut être meilleur qu'un autre pour décrire la réalité et bien sûr, plusieurs questions se posent alors : comment mesurer cette qualité? comment diagnostiquer un modèle? Nous présentons dans cette section une petite liste qui sert à résumer et classifier les différents modèles envisagés dans ce cours.

On distingue principalement deux types de modèles :

- les **modèles déterministes**. Ces modèles relèvent de la Statistique Descriptive. Ils ne font intervenir que de manière sous-jacente le calcul des probabilités et consistent à supposer que l'observation de la série à la date t est une fonction du temps t et d'une variable ϵ_t centrée faisant office d'erreur au modèle, représentant la différence entre la réalité et le modèle proposé :

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

On suppose de plus que les ϵ_t sont décorréliées.

- 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 (\pm lointaines suivant le modèle) et d'un terme d'erreur η_t

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

La classe des modèles de ce type la plus fréquemment utilisée est la classes des modèles SARIMA (et de ses sous-modèles ARIMA, ARMA,...). Comme vu plus haut, la série chronologique est l'observation d'un processus stochastique : la modélisation porte ici sur la forme du processus (ϵ_t).

Le cas particulier où la relation fonctionnelle g est linéaire est très important et très usité. Il mène aux modèles autorégressifs linéaires, par exemple un modèle d'ordre 2 avec des coefficients autorégressifs a_1, a_2 est donné par

$$\epsilon_t = a_1 X_{t-1} + a_2 X_{t-2} + \eta_t,$$

où (η_t) est un bruit blanc c'est à dire une variable aléatoire de moyenne nulle non corrélée.

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

⚠ Attention. Dans le cadre de ce cours, nous n'étudierons que les modèles déterministes. Les modèles stochastiques seront abordés dans l'UE de Renforcement Statistique.

📌 A retenir. La tendance et la saisonnalité sont, comme nous l'avons dit, déterministes c'est-à-dire non aléatoires. Elles sont constantes par rapport à l'aléa. Par contre, elles dépendent de t et varient donc au cours du temps.

1.2.4 Analyse de la série à partir de ses composantes

Une fois l'étape de modélisation effectuée, on étudie les composantes du modèle les unes après les autres comme nous le verrons au chapitre 4.

1.2.5 Diagnostic du modèle

Une fois le modèle construit et ses paramètres estimés, on vérifie que le modèle proposé est bon c'est-à-dire l'ajustement au modèle :

- en étudiant les résidus
- en faisant des tests
- ...

1.2.6 Prédiction

Enfin, une fois ces différentes étapes réalisées, nous serons en mesure de faire de la prédiction.

⚠ Attention. Dans le cadre de ce cours, nous ne traiterons pas les étapes de correction des données et d'ajustement au modèle mais seulement les étapes d'observation, modélisation, analyse de la série à partir de ses composantes et prédiction.

Chapitre 2

Modélisation déterministe

2.1 Le modèle additif

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

$$X_t = Z_t + S_t + \epsilon_t, \quad t = 1 \dots T,$$

où Z_t est la composante tendancielle, S_t la composante saisonnière et ϵ_t représente l'erreur ou l'écart au modèle.

Comme nous l'avons dit en introduction,

- la **tendance** Z_t exprime un mouvement à moyen terme de la série.
- la **composante saisonnière** exprime un phénomène qui se reproduit de manière analogue sur chaque intervalle de temps successif. En plus de la notion de périodicité, on suppose par ailleurs que l'effet du saisonnier est en moyenne nul sur une période, ce qui signifie que

$$\sum_{i=1}^P c_i = 0.$$

Remarque 2.1. *A 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*

$$X_t = Z_t + S_t + \epsilon_t, \quad t \in \mathbb{Z}.$$

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

$$X_t = Z'_t + S'_t + \epsilon_t, \quad t \in \mathbb{Z},$$

où pour tout $t \in \mathbb{Z}$, $Z'_t = Z_t + \frac{c}{p}$ et $S'_t = S_t - \frac{c}{p}$. On a donc trouvé une autre décomposition du signal X_t . On remarquera que 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.

Toutefois, il existe un lien entre la notion de périodicité et celle de somme nulle sur une période :

Propriété 2.2. Toute composante de somme nulle sur une période p est périodique de période p .

Exercice 1. Démontrer la propriété précédente.

— les **erreurs** sont des variables aléatoires.

☞ **A retenir.** Dans ce modèle, l'amplitude de la série reste constante au cours du temps. Ceci se traduit graphiquement par des fluctuations autour de la tendance Z_t constantes au bruit près.

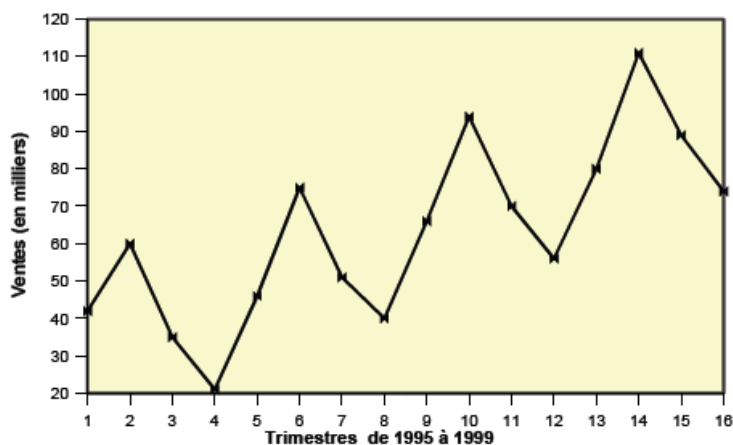


FIGURE 2.1 – Modèle additif. Amplitude constante autour de la tendance

Que peut-on dire des composantes présentes sur l'exemple précédent ?

- la série $(Z_t)_t$ représente la tendance générale (hausse des ventes? linéaire?).
- les données étant trimestrielles, la période est donc un an et $p = 4$.
- des valeurs $c_2 = +20$ et $c_4 = -10$ signifient que les ventes au deuxième trimestre sont supérieures de 20 milliers par rapport à l'ensemble de l'année alors qu'au quatrième trimestre, elles ont été inférieures de 10 milliers.
- une fluctuation irrégulière $\epsilon_{14} = -2$ signifie que les ventes ont été inférieures de 2 milliers au deuxième trimestre de 1998 par rapport à ce qui était prévu (c'est-à-dire ce que nous laissaient prévoir la tendance et l'effet saisonnier pour le deuxième trimestre de 1998).

2.2 Le modèle multiplicatif

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

$$X_t = Z_t(1 + S_t)(1 + \epsilon_t), \quad t = 1 \dots T,$$

où Z_t est la composante tendancielle, S_t la composante saisonnière et ϵ_t représente l'erreur ou l'écart au modèle.

Là encore, la composante saisonnière vérifie

$$\sum_{i=1}^P c_i = 0.$$

☛ **A retenir.** L'amplitude de la série n'est plus constante au cours du temps : elle varie au cours du temps proportionnellement à la tendance Z_t au bruit près. Dans ce modèle, on considère que les amplitudes des fluctuations dépendent du niveau.

Reprenons l'exemple des ventes trimestrielles.

- des valeurs $c_2 = +0.8$ et $c_4 = -0.5$ signifient ici que les ventes augmentent de 80% au second semestre et diminuent de 50% au quatrième trimestre par rapport à l'ensemble de l'année.
- une valeur $\epsilon_9 = +0.2$ signifie que les ventes ont été supérieures de 20% au premier semestre de 1998 par rapport au chiffre attendu pour ce semestre là.

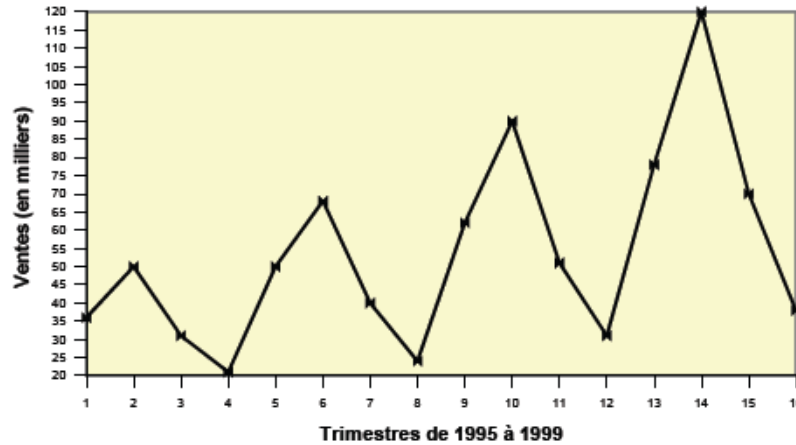


FIGURE 2.2 – Modèle multiplicatif. Amplitude proportionnelle à la tendance

Remarque 2.3. Le modèle multiplicatif est généralement utilisé pour des données de type économique.

2.3 Les modèles mixtes

Il s'agit là 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 :

$$X_t = Z_t \tilde{S}_t + \epsilon_t, \quad t = 1 \dots T,$$

avec l'hypothèse ici que $\sum_{i=1}^P c_i = P$.

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'une série dans laquelle pour une observation X donnée

- la variation saisonnière S s'ajoute simplement à la tendance Z ; c'est le modèle additif.
- la variation saisonnière S multiplie la tendance Z ; c'est le modèle multiplicatif.

Afin de faire cette distinction, on peut se baser sur une méthode graphique ou utiliser une méthode analytique. Nous étudions ces méthodes sur un exemple concret en nous basant sur la série chronologique "Nouvelles immatriculations de voitures particulières, commerciales et utilitaires

neuves selon le mois" :

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août
1996	2006	3224	3789	4153	3100	2527	3015	1504
1997	2247	3862	3586	4047	2838	2727	2730	1648
1998	2433	3723	4325	4493	3399	3083	3247	1928
1999	3127	4437	5478	4384	3552	3678	3611	2260
2000	3016	4671	5218	4746	4814	3545	3341	2439

	Sept.	Oct.	Nov.	Déc.
1996	1847	2314	1673	1602
1997	2007	2450	1966	1695
1998	2377	2831	2388	2126
1999	2699	3071	2510	2182
2000	2637	3085	2737	2055

• Méthode du 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.

Sur le graphique de notre exemple Figure 2.3, les courbes de profil semblent parallèles pour toutes les saisons. On peut donc supposer que le modèle est additif.

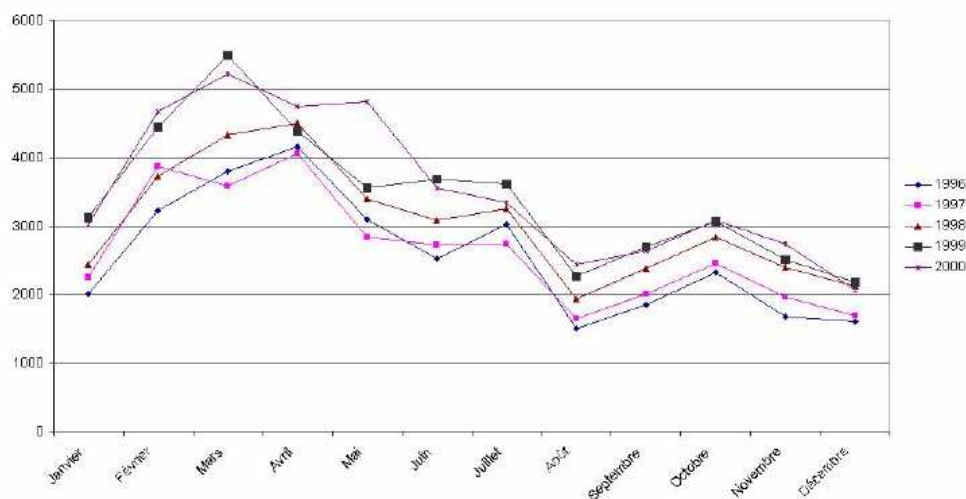


FIGURE 2.3 – Nouvelles immatriculations de voitures particulières, commerciales et utilitaires neuves au Luxembourg selon le mois

• Méthode de la bande

On fait un graphique représentant la série chronologique (cf. Figure 2.3), 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. Dans le cas contraire, c'est un modèle multiplicatif.

Sur notre exemple, nous constatons que ces deux droites ne sont pas parallèles, ce qui laisse plutôt penser que le modèle est multiplicatif.

Nous ne sommes donc pas en mesure de conclure.

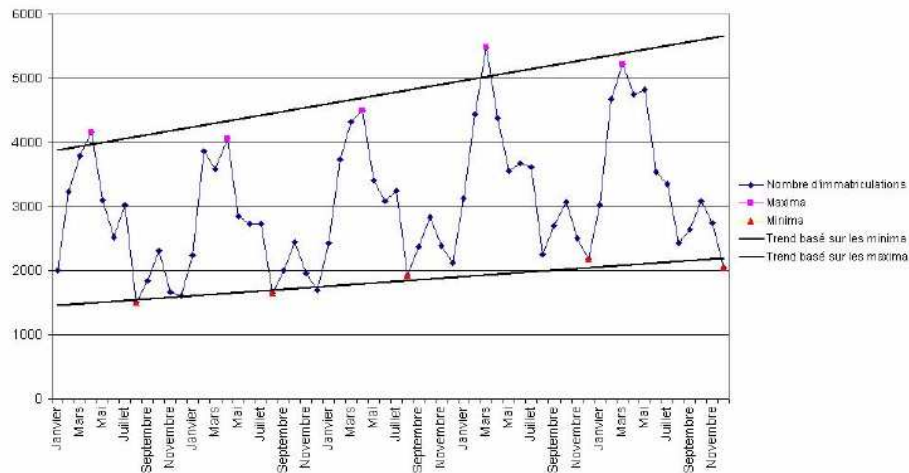


FIGURE 2.4 – Nouvelles immatriculations de voitures particulières, commerciales et utilitaires neuves de 1996 à 2000 au Luxembourg

• Méthode analytique

On calcule les moyennes \bar{x} et les écarts-types σ pour chacune des périodes considérées puis la droite des moindres carrés $\sigma = a\bar{x} + b$. Pour des rappels sur la droite des moindres carrés voir le chapitre suivant.

Si a est nul, c 'est le modèle additif, sinon c 'est le modèle multiplicatif.

Exemple 2.4. Vérifier grâce à la méthode analytique sur l'exemple des nouvelles immatriculations de voitures particulières que le modèle est multiplicatif.

☞ **A retenir.** Il faut bien tester avec les trois méthodes pour décider du modèle!

Chapitre 3

Moyennes mobiles

Dans le présent chapitre, nous nous intéressons à un ensemble d'outils, les moyennes mobiles ou filtres linéaires, transformations de séries chronologiques. Le but sera de lisser une série temporelle, en gardant la tendance et en supprimant la saisonnalité pour ensuite procéder à l'estimation de ces deux composantes. Nous présentons dans les Sections 3.2 et 3.3 les moyennes mobiles et leurs propriétés. Dans le prochain chapitre, la mise en oeuvre de la méthode en pratique et l'estimation des composantes déterministes seront détaillées dans le cadre des modèles déterministes.

3.1 Exemple introductif : la moyenne mobile arithmétique

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

$$\frac{t_1 + \dots + t_k}{k} \quad \text{puis} \quad \frac{t_2 + \dots + t_{k+1}}{k} \dots \quad \text{jusqu'à} \quad \frac{t_{T-k+1} + \dots + t_T}{k}$$

et pour la variable d'intérêt

$$\frac{X_1 + \dots + X_k}{k} \quad \text{puis} \quad \frac{X_2 + \dots + X_{k+1}}{k} \dots \quad \text{jusqu'à} \quad \frac{X_{T-k+1} + \dots + X_T}{k}.$$

Exemple 3.1. *Calcul d'une moyenne mobile arithmétique d'ordre 3.*

Date t	Série X_t	Date $M_3(t)$ de la MM	$M_3(X_t)$
1	5		
2	3		
3	4		
4	5		
5	4		
6	4		

Rappelons que par définition k est impair, $k = 2m + 1$. Généralisons cette procédure à tout k . Lorsque k est impair, $k = 2m + 1$, la série moyenne mobile est calculée aux mêmes instants que les observations initiales. En revanche, lorsque k est pair, $k = 2m$, 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

$$\frac{1}{k} \left(\frac{1}{2} X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1} + \frac{1}{2} X_{t+m} \right) = \frac{1}{2} M_k^1 X_t + \frac{1}{2} M_k^2 X_t,$$

combinaison linéaire des moyennes mobiles arithmétiques sur $2m$ valeurs

$$M_k^1 X_t = \frac{1}{k} (X_{t-m} + \dots + X_{t+m-1}) = M_{m_1+m_2+1} X_t \quad \text{avec} \quad m_1 = m, \quad m_2 = m-1$$

et

$$M_k^2 X_t = \frac{1}{k} (X_{t-m+1} + \dots + X_{t+m}) = M_{m_1+m_2+1} X_t \quad \text{avec} \quad m_1 = m-1, \quad m_2 = m.$$

Exemple 3.2. Calcul des moyennes mobiles arithmétiques d'ordre 2 et 4.

Date t	Série X_t	Date $M_2(t)$ de la MM	$M_2(X_t)$
1	5		
2	3		
3	4		
4	5		
5	4		
6	4		

Date t	Série X_t	Date $M_4(t)$ de la MM	$M_4(X_t)$
1	5		
2	3		
3	4		
4	5		
5	4		
6	4		

3.2 Définitions des moyennes mobiles

Ces outils font partie des premières méthodes pour l'analyse des séries chronologiques. Il semble que le physicien Poynting soit le premier, en 1884, à avoir utilisé les moyennes mobiles pour éliminer les variations accidentelles ou périodiques d'une série. En 1904, Spencer introduit une moyenne mobile (symétrique d'ordre 15) permettant de conserver les polynômes de degré 3. Puis à partir de 1914, les grands personnages de la statistique tels que Student, Pearson et Yule par exemple, s'intéressent à ce genre de problèmes.

Définition 3.3. 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

$$M_{m_1+m_2+1} X_t = \sum_{i=-m_1}^{m_2} \theta_i X_{t+i} = \theta_{-m_1} X_{t-m_1} + \dots + \theta_{-1} X_{t-1} + \theta_0 X_t + \theta_1 X_{t+1} + \dots + \theta_{m_2} X_{t+m_2},$$

où $\theta_{-m_1}, \dots, \theta_{m_2}$ sont des réels et $m_1, m_2 \in \mathbb{N}$. On appelle **ordre** de la moyenne mobile la valeur $m_1 + m_2 + 1$.

Remarque 3.4. 1. L'ordre représente simplement le nombre de termes considérés dans la somme.

2. Il est clair que la définition ci-dessus n'a de sens que pour des instants t tels que

$$m_1 + 1 \leq t \leq T - m_2.$$

3. On note aussi $M_{m_1+m_2+1} X_t = M_{m_1+m_2+1}(X_t) = X_t^*$. La seconde notation masque le fait la moyenne mobile est un opérateur et non une application. La dernière notation est plus compacte mais son inconvénient est que l'ordre n'apparaît plus.

Une moyenne mobile en t étant une combinaison linéaire finie des valeurs de la série correspondant à des dates entourant t , elle réalise donc un lissage de la série, une moyennisation.

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

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

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

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

On peut itérer cette application et définir par récurrence

$$B^k X_t = X_{t-k}, \quad k \in \mathbb{N}.$$

Par convention, B^0 est l'opérateur identité I .

L'opérateur B est linéaire et inversible. Son inverse $B^{-1} = F$ est défini par

$$\forall t \in \mathbb{Z}, \quad FX_t = X_{t+1}$$

L'opérateur F est appelé **opérateur avance**.

On peut réécrire la moyenne mobile en termes d'opérateurs B et F :

$$M_{m_1+m_2+1} = \sum_{i=-m_1}^{m_2} \theta_i B^{-i} = \sum_{i=-m_1}^{m_2} \theta_i F^i.$$

En factorisant par B^{m_1} et en faisant $j = i + m_1$, on obtient la **forme canonique** suivante

$$M_{m_1+m_2+1} = B^{m_1} \sum_{j=0}^{m_1+m_2} \theta_{j-m_1} F^j = B^{m_1} \sum_{i=-m_1}^{m_2} \theta_i F^{m_1+i} =: B^{m_1} P(F)$$

Remarque 3.5. On notera que $-m_1$ est donc le plus petit exposant de F et m_2 le plus grand dans la définition de la moyenne mobile. De manière équivalente, m_1 est donc le plus grand exposant de B et $-m_2$ le plus petit.

Vocabulaire :

— Le polynôme P intervenant dans cette dernière expression de $M_{m_1+m_2+1}$ et défini par

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

est appelé **polynôme caractéristique** de la moyenne mobile $M_{m_1+m_2+1}$. On remarque que P peut aussi s'écrire sous la forme suivante

$$P(x) = \sum_{j=0}^{m_1+m_2} \theta_{j-m_1} x^j$$

— On dit que la moyenne mobile est **centrée** lorsque $m_1 = m_2 = m$.

- 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

$$\theta_i = \frac{1}{2m+1}, \quad \forall i = -m \dots m.$$

Une moyenne mobile arithmétique est donc centrée (par définition) et symétrique. On a en particulier dans ce cas, $\sum_{i=-m}^m \theta_i = 1$ et $M_{2m+1} X_t$ apparaît comme la moyenne des observations $X_{t-m}, \dots, X_t, \dots, X_{t+m}$.

Exercice 2. Calculer les séries des moyennes mobiles arithmétiques pour $k = 2, 3$ et 4 de la série initiale X_t suivante

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_t	30	15	5	30	36	18	9	36	45	15	10	60	48	16	8	72

Exercice 3. Les moyennes mobiles symétriques vérifient les propriétés suivantes à démontrer :

1. Si M_1 et M_2 sont deux moyennes mobiles centrées, alors il en est de même de $M_1 M_2$.
2. Une moyenne mobile centrée $M = B^m P(F)$ est symétrique si et seulement si $P(B) = B^{2m} P(F)$.
3. Si M_1 et M_2 sont deux moyennes mobiles symétriques, alors il en est de même de $M_1 M_2$.

3.3 Propriétés d'un lissage par moyenne mobile

3.3.1 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.

Plus précisément,

Propriété 3.6. 1. Une moyenne mobile conserve les constantes non nulles si et seulement si

$$\theta_{-m_1} + \dots + \theta_{m_2} = 1.$$

2. Une moyenne mobile symétrique conservant les constantes conserve les polynômes de degré 1.

Exercice 4. Démontrer les deux propriétés précédentes.

On vient de voir qu'une moyenne mobile arithmétique d'ordre $2m + 1$ conserve les polynômes de degré 1. Quelles sont les autres séries invariantes par cette moyenne ?

Exercice 5. Vérifier qu'une moyenne mobile arithmétique ne conserve pas les polynômes de degré 2.

Exercice 6. On peut trouver des séries invariantes par $M_1 M_2$ à partir de celles invariantes par M_1 et M_2 :

1. Si X_t est une série invariante par M_1 et M_2 , alors X_t est invariante par $M_1 M_2$.
2. Qu'en est-il de la réciproque ?

Remarque 3.7. Comme une moyenne mobile est un opérateur linéaire sur l'espace vectoriel des suites indexées par \mathbb{Z} , chercher les séries invariantes par une moyenne mobile, c'est-à-dire les séries qui vérifient

$$X_t^* = X_t,$$

revient donc à chercher les vecteurs propres associés à la valeur propre 1 de l'opérateur moyenne mobile considéré.

Pour trouver les séries invariantes par une moyenne mobile, on considère le polynôme

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

dont on cherche les racines. Dans le cas centré, on considère simplement

$$P(x) - x^m = \theta_{-m} + \theta_{-m+1}x + \dots + \theta_m x^{2m} - x^m$$

Propriété 3.8. La moyenne mobile d'ordre $m_1 + m_2 + 1$ conserve les polynômes de degré inférieur ou égal à p si et seulement si 1 est racine d'ordre $p + 1$ du polynôme $P(x) - x^{m_1}$.

Démonstration. Étudions le polynôme $Q(x) := P(x) - x^{m_1}$. Tout d'abord, nous avons pour tout k

$$\left\{ \begin{array}{l} P(x) = \sum_{i=-m_1}^{m_2} \theta_i x^{m_1+i} \\ P'(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i) x^{m_1+i-1} \\ P''(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i)(m_1+i-1) x^{m_1+i-2} \\ \vdots \\ P^{(k)}(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i)(m_1+i-1)\dots(m_1+i-k+1) x^{m_1+i-k} \end{array} \right.$$

et immédiatement

$$\left\{ \begin{array}{l} Q(1) = \sum_{i=-m_1}^{m_2} \theta_i - 1 \\ Q'(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i) - m_1 = \sum_{i=-m_1}^{m_2} \theta_i i + m_1 \left(\sum_{i=-m_1}^{m_2} \theta_i - 1 \right) \\ Q''(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i)(m_1+i-1) - m_1(m_1-1) \\ = \sum_{i=-m_1}^{m_2} \theta_i i^2 + (2m_1-1) \sum_{i=-m_1}^{m_2} \theta_i i + m_1(m_1-1) \left(\sum_{i=-m_1}^{m_2} \theta_i - 1 \right) \\ \vdots \\ Q^{(k)}(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1+i)\dots(m_1+i-k+1) - m_1(m_1-1)\dots(m_1-k+1) \\ = \sum_{l=1}^k \sum_{i=-m_1}^{m_2} \theta_i i^l + \dots + m_1 \dots (m_1-k+1) \left(\sum_{i=-m_1}^{m_2} \theta_i - 1 \right) \end{array} \right.$$

Par conséquent, 1 est racine d'ordre $p + 1$ de $P(x) - x^{m_1}$ si et seulement si

$$\left\{ \begin{array}{l} Q(1) = 0 \\ Q'(1) = 0 \\ Q''(1) = 0 \\ \vdots \\ Q^{(p)}(1) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \sum_{i=-m_1}^{m_2} \theta_i = 1 \\ \sum_{i=-m_1}^{m_2} \theta_i i = 0 \\ \sum_{i=-m_1}^{m_2} \theta_i i^2 = 0 \\ \vdots \\ \sum_{i=-m_1}^{m_2} \theta_i i^p = 0 \end{array} \right.$$

Soit maintenant la série $X_t = t^k$, pour $k \leq p$. Sa transformée par la moyenne mobile centrée d'ordre $2m + 1$ est donnée par

$$\begin{aligned} X_t^* &= \sum_{i=-m_1}^{m_2} \theta_i (t+i)^k = \sum_{i=-m_1}^{m_2} \theta_i \sum_{l=0}^k C_k^l t^{k-l} i^l \\ &= \sum_{l=0}^k C_k^l t^{k-l} \sum_{i=-m_1}^{m_2} \theta_i i^l = C_k^0 t^k = t^k \end{aligned}$$

si on suppose que 1 est racine d'ordre $p + 1$ de $P(x) - x^{m_1}$. Et on en déduit que la moyenne mobile d'ordre $m_1 + m_2 + 1$ conserve bien les polynômes de degré inférieur ou égal à p . On vérifie aisément que la réciproque est vraie. \square

Propriété 3.9. La moyenne mobile d'ordre $m_1 + m_2 + 1$ conserve les fonctions de la forme a^t ssi a est racine du polynôme $P(x) - x^{m_1}$.

Exercice 7. Démontrer la propriété précédente.

Exercice 8. Etudions la moyenne arithmétique d'ordre 5.

- Ecrire son polynôme associé.
- Vérifier que la moyenne arithmétique conserve les polynômes de degré 1.
- Déterminer les séries invariantes par la moyenne mobile arithmétique d'ordre 5.

3.3.2 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 arithmétique d'ordre P supprime cette saisonnalité. La série $(M_P X_t)_t$ ne possède plus de composante saisonnière de période P .

Plus précisément, cherchons les séries chronologiques qui sont arrêtées par le filtre moyenne mobile centrée d'ordre $2m + 1$. Ce sont les séries S_t telles que leur transformée S_t^* par la moyenne mobile vérifie $S_t^* = 0$. On cherche donc à déterminer le **noyau** d'une moyenne mobile arithmétique d'ordre $2m + 1$ ou encore les vecteurs propres associés à la valeur propre nulle.

Pour trouver les séries arrêtées par une moyenne mobile, on considère le polynôme

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

dont on cherche les racines. Dans le cas centré, on considère simplement

$$P(x) = \theta_{-m} + \theta_{-m+1}x + \dots + \theta_m x^{2m}$$

Propriété 3.10. La moyenne mobile d'ordre $m_1 + m_2 + 1$ arrête les fonctions de la forme a^t ssi a est racine du polynôme $P(x)$.

Exercice 9. Démontrer la propriété précédente.

Exercice 10. On peut trouver des séries arrêtées par $M_1 M_2$ à partir de celles arrêtées par M_1 et M_2 :

1. Si X_t est une série arrêtée par M_1 ou M_2 , alors X_t est arrêtée par $M_1 M_2$.
2. Qu'en est-il de la réciproque ?

La proposition ci-dessous montre que l'ensemble des séries absorbées par M : $\text{Ker}(M)$ est un sous-espace vectoriel.

Proposition 3.11. Soit une moyenne mobile M définie par

$$M = \sum_{i=-m_1}^{m_2} \theta_i B^{-i}.$$

Alors

$$\text{Ker}(M) = \text{Vect} \left\{ (t^k r_j^t, \quad t \in \mathbb{Z}; \quad \forall k = 0 \dots l_j - 1, \quad \forall j = 1 \dots m) \right\},$$

où r_1, \dots, r_m sont les racines d'ordres de multiplicité respectives l_1, \dots, l_m de l'équation suivante

$$\forall t \in \mathbb{Z}, \quad \sum_{i=-m_1}^{m_2} \theta_i X_{t+i} = 0.$$

Démonstration. En utilisant le polynôme caractéristique, il est immédiat de voir que toute suite (X_t) absorbée par la moyenne mobile M est solution de l'équation de récurrence ci-dessous

$$\forall t \in \mathbb{Z}, \quad \sum_{i=-m_1}^{m_2} \theta_i X_{t+i} = 0.$$

Il s'agit d'une équation de récurrence linéaire à coefficients constants d'ordre $m_1 + m_2$. L'ensemble des solutions de cette équation forme un espace vectoriel de dimension $m_1 + m_2$ dont une base est $(t^k r_j^t, \quad t \in \mathbb{Z})$ où r_1, \dots, r_m sont les racines d'ordres de multiplicité respectives l_1, \dots, l_m de l'équation ci-dessus. \square

Exercice 11. Rechercher les séries arrêtées par le filtre moyenne mobile arithmétique.

Propriété 3.12. Une moyenne mobile absorbe les composantes saisonnières de période p si et seulement si son polynôme caractéristique P est divisible par $1 + x + \dots + x^{p-1}$.

Exercice 12. Démontrer la proposition précédente.

3.3.3 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). Nous allons étudier maintenant l'effet d'une moyenne mobile sur le résidu lorsque celui-ci est un bruit blanc. Par construction, une moyenne mobile consiste à faire des moyennes partielles de proche en proche. On obtient donc un lissage de la série. L'effet de la composante irrégulière est d'autant plus atténué que l'ordre de la moyenne mobile est grand.

Plus précisément,

Propriété 3.13. Les moyennes mobiles arithmétiques d'ordre $2m + 1$ sont les moyennes mobiles minimisant la variance d'un bruit blanc parmi les moyennes mobiles centrées telles que $\sum_{i=-m}^m \theta_i = 1$.

Exercice 13. Nous nous proposons de démontrer la propriété précédente.

- Calculer la transformée d'un bruit blanc par une moyenne mobile centrée d'ordre $2m + 1$ telle que $\sum_{i=-m}^m \theta_i = 1$ et sa variance.

— Vérifier que le problème revient à résoudre le problème de minimisation

$$\min_{\theta_i, i=-m \dots m} \sum_{i=-m}^m \theta_i^2 \quad \text{sous la contrainte} \quad \sum_{i=-m}^m \theta_i = 1.$$

— En appliquant la méthode des multiplicateurs de Lagrange, déterminer le $(2m + 1)$ -uplet $(\theta_{-m}, \dots, \theta_m)$ solution de ce problème de minimisation.

— Conclure.

Ainsi les moyennes arithmétiques d'ordre $2m + 1$ transforment un bruit blanc en un processus centré (inchangé) et de variance $\frac{\sigma^2}{2m+1}$ réduite. Notons cependant que les variables ϵ_t^* sont à présent corrélées alors que les ϵ_t ne l'étaient pas. On a en effet

$$\begin{aligned} \gamma(h) &= \text{Cov}(\epsilon_t^*, \epsilon_{t+h}^*) = \mathbb{E}(\epsilon_t^* \epsilon_{t+h}^*) \\ &= \mathbb{E} \left(\sum_{i=-m}^m \theta_i \epsilon_{t+i} \sum_{j=-m}^m \theta_j \epsilon_{t+h+j} \right) \\ &= \sum_{i=-m}^m \sum_{j=-m}^m \theta_i \theta_j \mathbb{E}(\epsilon_{t+i} \epsilon_{t+h+j}) \\ &= \begin{cases} \sigma^2 \sum_{i=h-m}^m \theta_i \theta_{i-h} & \text{pour } h \leq 2m \\ 0 & \text{pour } h > 2m \end{cases} \\ &= \begin{cases} \sigma^2 \frac{2m+1-h}{(2m+1)^2} & \text{pour } h \leq 2m \\ 0 & \text{pour } h > 2m \end{cases} \end{aligned}$$

On voit ainsi que la covariance entre ϵ_t^* et ϵ_{t+h}^* ne dépend que de σ^2 (variance du bruit blanc et de h et non de t et qu'elle s'annule dès que $h > 2m$. Nous verrons plus loin la notion qui généralise les processus vérifiant cette propriété d'indépendance en t de la covariance. En divisant $\gamma(h)$ par $\text{Var}(\epsilon_t^*) = \gamma(0)$, on obtient la corrélation linéaire entre ϵ_t^* et ϵ_{t+h}^* :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} \frac{2m+1-h}{2m+1} & \text{pour } h \leq 2m \\ 0 & \text{pour } h > 2m \end{cases}$$

La corrélation est donc indépendante de la variance du bruit blanc.

3.3.4 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. Il s'agit donc de faire disparaître la saisonnalité et de réduire au maximum le bruit blanc. Nous avons vu précédemment que

- on fait disparaître une composante saisonnière de période p avec une moyenne mobile d'ordre p .
- on gomme d'autant plus le bruit que l'ordre de la moyenne mobile est grand.
- en revanche, on perd les caractéristiques de la tendance avec une moyenne mobile d'ordre trop élevé (jusqu'à obtenir une tendance constante) d'après un exercice du TD.

En pratique, on doit donc trouver le meilleur compromis pour le choix de l'ordre de lissage optimal.

Cas particulier des moyennes mobiles arithmétiques

On a vu qu'une moyenne mobile arithmétique d'ordre $2m + 1$

- laisse invariants les polynômes de degré 1,
- arrête les fonctions périodiques de période $2m + 1$,
- réduit la variance d'un bruit blanc de manière optimale (c'est-à-dire d'un facteur $\frac{1}{2m+1}$),
- laisse invariants certaines suites géométriques de la forme a^t (si a est racine de $P(x) - x^m$),
- absorbe certaines suites géométriques de la forme a^t (si a est racine de P).

C'est pourquoi ce sont celles que nous utiliserons le plus fréquemment en pratique.

Remarque 3.14. *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*

$$\frac{1}{2m} \left(\frac{1}{2} X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1} + \frac{1}{2} X_{t+m} \right).$$

En reprenant des calculs analogues à ceux effectués précédemment dans cette section, on peut montrer que les moyennes mobiles ainsi définies

- laissent invariants les polynômes de degré 1;
- arrêtent fonctions périodiques de période $2m$ dont la somme des coefficients est nulle;
- réduisent la variance d'un bruit blanc par un facteur $\frac{2m-1/2}{4m^2}$.

Remarque 3.15. *1) Les moyennes mobiles permettent de lisser directement la série sans hypothèse a priori sur le modèle sous-jacent. La méthode est donc valable quel que soit le modèle de décomposition. Pour cette raison, on peut classer ce type de lissage dans les méthodes non-paramétriques (par opposition aux méthodes paramétriques abordées plus loin). C'est un outil simple à mettre en oeuvre qui met en évidence l'allure de la tendance en supprimant la composante saisonnière et en atténuant le bruit.*

*2) Dans tout ce qui précède, on peut remplacer la moyenne par la médiane et on obtient alors un lissage par **médiane mobile**. Ce procédé a alors l'avantage d'être moins sensible aux valeurs aberrantes.*

Chapitre 4

Décomposition d'une série chronologique

Nous disposons maintenant des outils de base permettant de faire l'étude complète d'une série chronologique. Afin de faire des estimations précises en vue de la prédiction, il est important de disposer des observations de la série statistique sur une longue période. Le premier travail consiste à effectuer une représentation graphique adéquate permettant d'avoir une vue globale du phénomène en question : tendance, saisonnalité, période, bruit, modèle additif ou multiplicatif.

Ensuite, la décomposition et l'étude de la série statistique (X_t) en vue de la prédiction se font selon les étapes suivantes

1. application d'une moyenne mobile d'ordre judicieusement choisi afin de
 - conserver la tendance
 - supprimer la saisonnalité
 - atténuer l'effet du bruit
2. isolement et estimation de la saisonnalité.
3. isolement et estimation de la tendance.
4. itération éventuelle de la procédure.
5. prévision des valeurs futures.
6. analyse des résidus.

4.1 Un exemple introductif : Cas d'une tendance sensiblement constante

Commençons par traiter un premier exemple simple : dans la Figure 1.2 (d), on observe que la tendance ne varie pas beaucoup et il semble naturel de la supposer constante. Un estimateur naturel de la tendance est donc

$$\bar{X} = \frac{1}{72} \sum_{i=1}^{72} X_i.$$

Les résidus $(X_i - \bar{X})_{i=1, \dots, 72}$ (représentés à la Figure 4.1) comportent toujours la composante saisonnière qui par définition ne dépend que du mois et non de l'année. En conséquence, un estimateur de cette composante S_j est donné par la moyenne empirique sur le même mois de chaque année des résidus :

$$\tilde{c}_i = \frac{1}{6} \sum_{j=1}^6 (X_{12(j-1)+i} - \bar{X}), \quad i = 1, \dots, 12.$$

De façon à avoir une estimation de la saisonnalité nulle en moyenne sur une période, on calcule

$$\sum_{i=1}^{12} \tilde{c}_i$$

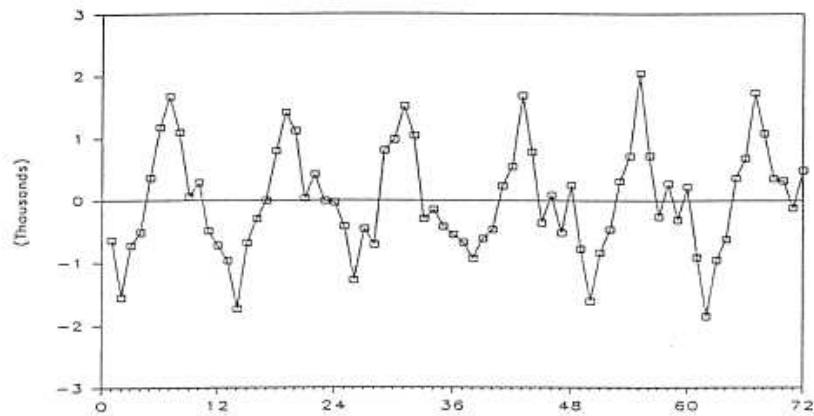


FIGURE 4.1 – Résidus de l'exemple 1.2 (d) après élimination de la tendance

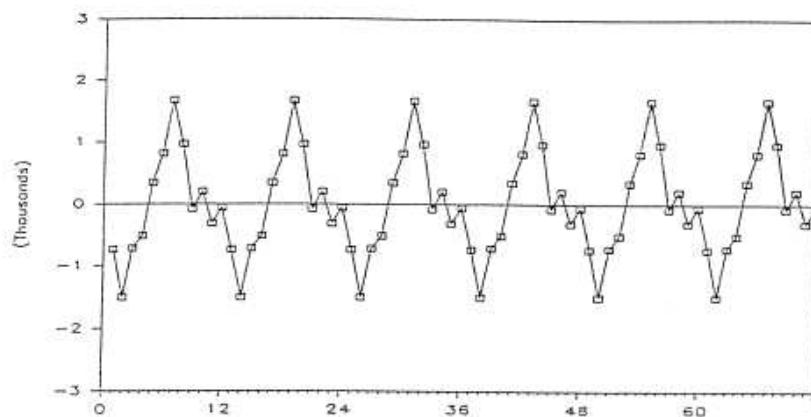


FIGURE 4.2 – Estimation de la composante saisonnière de l'exemple 1.2 (d)

et on définit

$$\hat{c}_i = \bar{c}_i - \frac{1}{12} \sum_{i=1}^{12} \bar{c}_i, \quad i = 1, \dots, 12.$$

La Figure 4.2 montre l'estimation \hat{c}_i de la saisonnalité et la Figure 4.3 représente la série chronologique après élimination de la tendance \bar{X} et de la saisonnalité \hat{S}_i :

$$X_{12(j-1)+i} - \bar{X} - \hat{c}_i.$$

On obtient donc la composante aléatoire de la série que nous analyserons à la fin du chapitre et au second semestre en renforcement. La méthode que nous venons d'appliquer se généralise à des périodes non nécessairement égales à 12 mais présuppose quand même un modèle additif pour lequel la tendance varie faiblement.

Remarque 4.1. Afin d'être plus précis, nous aurions aussi pu supposer la tendance constante sur une année (et non sur toute la durée de l'étude). Un estimateur naturel de la tendance pour l'année j aurait donc été

$$\frac{1}{12} \sum_{i=1}^{12} X_{12(j-1)+i}, \quad j = 1, \dots, 6.$$

La suite de la décomposition et de l'estimation aurait été inchangée.

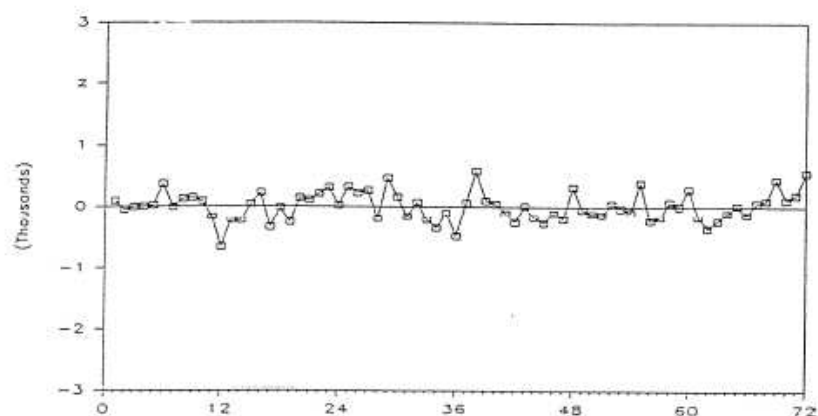


FIGURE 4.3 – Résidus de l'exemple 1.2 (d) après élimination de la tendance et de la saisonnalité

4.2 L'exemple étudié

Nous allons illustrer cette décomposition par un exemple que nous étudierons pas à pas tout au long de ce chapitre.

Exemple 4.2. *Étudions par exemple les ventes d'un produit P sur trois années. Les résultats sont consignés dans le tableau suivant*

Année	Numéro du trimestre	Ventes
2005	1	860
	2	794
	3	1338
	4	1148
2006	1	1096
	2	1021
	3	1705
	4	1505
2007	1	1436
	2	1363
	3	2319
	4	2047

Commençons par représenter cette série chronologique. Voir Figure 4.4.
Quel est votre diagnostic ?

1. Modèle additif ? multiplicatif ?
2. Tendance ?
3. Saisonnalité ? Période ?
4. Bruit ?

4.3 La série lissée par moyenne mobile

Afin d'éliminer les mouvements cycliques, saisonniers et accidentels, on utilise la technique des moyennes mobiles et on procède ainsi au lissage de la courbe.

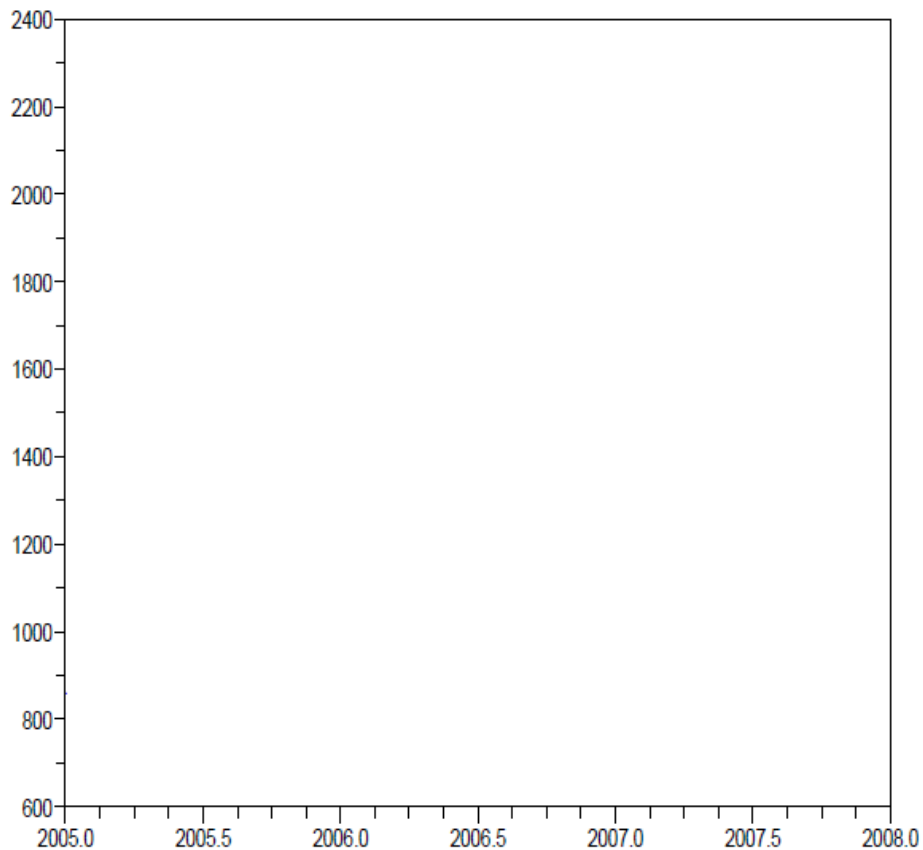


FIGURE 4.4 – Représentation graphique des ventes trimestrielles d'un produit P de l'exemple 4.2

D'après ce que nous venons de voir, la méthode des moyennes mobiles arithmétiques peut être utilisée pour tout type de modèle. Cependant, d'après ses propriétés, elle est particulièrement adaptée pour le modèle déterministe additif lorsque

- la tendance est sensiblement linéaire,
- la composante saisonnière est périodique,
- le bruit est de variance faible.

Remarque 4.3. *Dans le cas de polynôme de degré supérieur à 1, on peut aussi utiliser des moyennes mobiles autres qu'arithmétiques laissant invariants les polynômes de degré supérieur à 1.*

Certains auteurs préconisent également d'utiliser la méthode des moyennes mobiles comme technique de lissage de la série quelle que soit la forme de la tendance. On peut cependant utiliser d'autres techniques plus adaptées pour lisser la série (telle la méthode de loess).

Par conséquent, on applique une moyenne mobile arithmétique d'ordre $2m + 1$ dans le cas d'une saisonnalité d'ordre impair $2m + 1$ ou une moyenne mobile arithmétique modifiée d'ordre $2m + 1$ dans le cas d'une saisonnalité de période paire $2m$. Si le modèle est bon, la série transformée ne contient plus aucun mouvement saisonnier. Notons que la série ainsi obtenue est de longueur $T - 2m$.

Exemple 4.2 (suite)

Ici la période est . On applique donc une moyenne mobile Ainsi

$$\left\{ \begin{array}{l} \text{pour 2005, } X_3^* = \quad, X_4^* = \\ \text{pour 2006, } X_5^* = \quad, X_6^* = \quad, X_7^* = \quad, X_8^* = \\ \text{pour 2007, } X_9^* = \quad, X_{10}^* = \end{array} \right.$$

4.4 Estimation de la saisonnalité

Afin d'isoler la saisonnalité, on calcule la série diminuée de sa tendance \tilde{S}_t en retranchant à la série de départ la série transformée X_t^* . On estime ensuite les P coefficients du saisonnier par la moyenne des valeurs de \tilde{S}_t correspondant à chaque temps de la période. On affine la méthode en retranchant ensuite à chaque coefficient estimé la moyenne des coefficients afin que la condition de nullité de la moyenne des coefficients sur la période soit respectée.

Plus précisément, supposons que les observations sont périodiques de période $P = 2m$ paire et réalisées sur N périodes. A l'issue du premier filtrage, on dispose de la série

$$\tilde{S}_t = X_t - X_t^*, \quad t = m+1, \dots, PN - m.$$

Cette série est appelée **série corrigée de la tendance**. Disposons ces valeurs dans un tableau

inter-périodes	période	1	...	m	...	j	...	P-m+1	...	P
1		***	***	***						
⋮										
i						$\tilde{S}_{j+P(i-1)}$				
⋮										
N								***	***	***

L'estimation du coefficient saisonnier c_j est donc

$$\tilde{c}_j = \begin{cases} \frac{1}{N-1} \sum_{i=2}^N \tilde{S}_{j+P(i-1)}, & 1 \leq j \leq m \\ \frac{1}{N-1} \sum_{i=1}^{N-1} \tilde{S}_{j+P(i-1)}, & 1 \leq j \leq P \end{cases}$$

Cette opération consiste à appliquer à la série $(\tilde{S}_t)_{t=m+1, \dots, PN-m}$ à laquelle on ajoute P observations nulles pour $t = 1, \dots, m$ et $t = PN - m + 1, \dots, PN$, une moyenne mobile d'ordre $P(N-1) + 1$ et de coefficients

$$\frac{1}{N-1} \left(\underbrace{1, 0, \dots, 0}_{P-1}, 1, 0, \dots, 0, 1, 0, \dots, 0, \underbrace{1}_{P-1} \right).$$

Les estimateurs des P coefficients saisonniers sont ensuite

$$\hat{c}_j = \tilde{c}_j - \frac{1}{P} \sum_{j'=1}^P \tilde{c}_{j'},$$

de façon à bien vérifier la contrainte d'une somme nulle sur une période.

Exemple 4.2 (suite)

Calculons les estimations des coefficients saisonniers pour l'exemple 4.2.

A l'issue du premier filtrage, on dispose de la série

$$\tilde{S}_t = X_t - X_t^*, \quad t = 3, \dots, 10.$$

On obtient alors

$$\begin{cases} \text{pour 2005, } \tilde{S}_3 = & , \tilde{S}_4 = \\ \text{pour 2006, } \tilde{S}_5 = & , \tilde{S}_6 = & , \tilde{S}_7 = & , \tilde{S}_8 = \\ \text{pour 2007, } \tilde{S}_9 = & , \tilde{S}_{10} = \end{cases}$$

L'estimation du coefficient saisonnier c_j est donc

$$\begin{cases} \tilde{c}_1 = \\ \tilde{c}_2 = \\ \tilde{c}_3 = \\ \tilde{c}_4 = \end{cases}$$

Les estimateurs des 4 coefficients saisonniers sont ensuite

$$\begin{cases} \hat{c}_1 = \\ \hat{c}_2 = \\ \hat{c}_3 = \\ \hat{c}_4 = \end{cases}$$

de façon à bien vérifier la contrainte d'une somme nulle sur une période.

Enfin on obtient

$$\begin{cases} \text{pour 2005, } \hat{S}_1 = & , \hat{S}_2 = & , \hat{S}_3 = & , \hat{S}_4 = \\ \text{pour 2006, } \hat{S}_5 = & , \hat{S}_6 = & , \hat{S}_7 = & , \hat{S}_8 = \\ \text{pour 2007, } \hat{S}_9 = & , \hat{S}_{10} = & , \hat{S}_{11} = & , \hat{S}_{12} = \end{cases}$$

4.5 Estimation de la tendance

Une fois les coefficients saisonniers estimés à l'étape précédente, on retranche l'estimation du saisonnier à la série X_t . La série ainsi obtenue est appelée **série corrigée des valeurs saisonnières** :

$$X_{CVS,t} = X_t - \hat{S}_t, \quad t = 1, \dots, T$$

avec $\hat{S}_t = \hat{c}_j, t \equiv j[P]$.

Remarque 4.4. Notons que si l'on avait soustrait \tilde{S}_t à X_t au lieu de \hat{S}_t , nous aurions simplement obtenu

$$X_t - \tilde{S}_t = X_t^*, \quad t = m+1, \dots, T-m,$$

au lieu de $X_{CVS,t}$. Mais ce résultat aurait été moins précis puisque nous n'aurions eu des valeurs t seulement de $m+1$ à T .

On procède ensuite à l'estimation du terme représentant la tendance par une méthode de régression : on modélise le plus souvent la tendance par un polynôme $Q(t)$ (en général un polynôme de degré 1 pour être cohérent avec le choix fait lors de la première étape). Puis on ajuste au sens des moindres carrés un polynôme $\hat{Q}(t)$ à la série corrigée des valeurs saisonnières $X_{CVS,t}$.

4.5.1 Quelques méthodes d'ajustement tendanciels

Ajustement tendanciel linéaire par moindres carrés C'est ce qui sera utilisé le plus souvent en pratique. On supposera que la tendance est linéaire et on appliquera la méthode des moindres carrés pour prédire les coefficients a et b . Cf. Annexe pour les rappels concernant cette méthode.

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_α, Z_α) et (t_β, Z_β) et on fait passer la droite par ces deux points. Les coefficients (\tilde{a}, \tilde{b}) vérifient

$$\begin{cases} Z_\alpha = \tilde{a}t_\alpha + \tilde{b} \\ Z_\beta = \tilde{a}t_\beta + \tilde{b} \end{cases} \text{ soit } \begin{cases} \tilde{a} = \frac{Z_\alpha - Z_\beta}{t_\alpha - t_\beta} \\ \tilde{b} = Z_\alpha - \frac{Z_\alpha - Z_\beta}{t_\alpha - t_\beta} t_\alpha \end{cases}$$

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.

Exercice 14. : Illustration des deux méthodes

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

t	1999	2000	2001	2002	2003	2004	2005	2006	2007
Z_t	32	38	48	52	61	73	80	84	95

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

Ajustements tendanciels non linéaires Dans certaines situations, la modélisation linéaire de la tendance peut être trop simplificatrice. Dans le cas de la Figure 1.2(a), 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.

Exemple 4.5. — 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 $Z_t = b \exp(at)$, 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 .

Exemple 4.6. Dans le cas de l'exemple de la Figure 1.2 (a), on peut aussi mener des calculs analogues à ceux effectués précédemment et approcher la tendance par

$$\hat{Z}_t = \hat{a}t^2 + \hat{b}t + \hat{c} \text{ avec } \begin{cases} \hat{a} = 6.499 \cdot 10^3 \\ \hat{b} = -2.335 \cdot 10^7 \\ \hat{c} = 2.098 \cdot 10^{10} \end{cases}$$

Estimation non paramétrique Dans certaines situations, il n'est pas facile de trouver le degré du polynôme d'ajustement pour Z_t ou de changement de variable adéquat. Par exemple, dans la Figure 1.2 (b), il n'est pas possible d'utiliser la méthode des moindres carrés car le polynôme utilisé au départ n'est ni linéaire, ni quadratique. On pourrait utiliser un polynôme avec un degré élevé mais le nombre de paramètres à estimer serait important et rendrait les calculs fastidieux. Par

ailleurs, on ne sait pas non plus déterminer l'allure de cette fonction.

Dans cette situation, on a recours à la théorie non paramétrique de l'estimation de la tendance qui ne suppose rien sur celle-ci a priori et on approxime la tendance par la moyenne mobile arithmétique d'ordre $2m + 1$,

$$X_t^* = M_{2m+1}(t).$$

Voir le Chapitre 3 pour la définition des moyennes mobiles et leurs propriétés.

Remarque 4.7. *Compte tenu des propriétés des moyennes mobiles vues au Chapitre 3, les hypothèses préalables à l'emploi de cet algorithme sont*

- *une saisonnalité à coefficients constants dont la période est égale à l'ordre de la moyenne mobile (période impaire) ou à l'ordre de la moyenne mobile moins 1 (période paire).*
- *une erreur de faible variance.*

Remarque 4.8. *Par cette technique, qu'on utilise une moyenne mobile arithmétique, on ne dispose pas d'estimation de la tendance pour les m premières et m dernières valeurs observées. Afin d'avoir des estimées de la tendance pour toutes les dates observées, on peut, si les points $(X_t^*)_{t=m+1, \dots, T-m}$ paraissent alignés par exemple, faire un ajustement linéaire comme expliqué précédemment. On fait ensuite une interpolation linéaire afin de calculer des estimées de la tendance pour les points $t = 1, \dots, m$ et $t = T - m + 1, \dots, T$.*

4.5.2 Retour à l'exemple

Exemple 4.2 (suite)

Au vu de la Figure 4.4, nous choisissons d'estimer la tendance par une droite. On obtient alors pour X_{CVS}

$$\begin{cases} \text{pour 2005, } X_{CVS}(1) = X_{CVS}(2) = X_{CVS}(3) = X_{CVS}(4) = \\ \text{pour 2006, } X_{CVS}(5) = X_{CVS}(6) = X_{CVS}(7) = X_{CVS}(8) = \\ \text{pour 2007, } X_{CVS}(9) = X_{CVS}(10) = X_{CVS}(11) = X_{CVS}(12) = \end{cases}$$

Par la méthode des moindres carrés, on trouve

$$\begin{aligned} \hat{a} &= \\ \hat{b} &= \end{aligned}$$

De plus, la qualité de la régression est bonne puisque

$$r =$$

Remarque 4.9. *On peut aussi faire une translation des temps et étudier plutôt $t' = [1 : 12]$. On trouve alors*

$$\begin{aligned} \hat{a} &= \frac{\text{Cov}(t', X_{CVS,t})}{\text{Var}(t')} \approx \frac{1127.9}{13} \approx 94.564 \\ \hat{b} &= \overline{X_{CVS,t}} - \hat{a} \bar{t}' \approx 771.33 \end{aligned}$$

La qualité de la régression est la même. Et on retrouve les mêmes prédictions à arrondis près.

4.6 Prévision des valeurs futures

Afin de prévoir les valeurs futures de la série, on utilise l'estimation de la tendance et celle de la composante saisonnière. Plus précisément, si on souhaite prévoir une valeur de la série à l'instant

$T + h$, où $h \geq 1$, c'est-à-dire à l'**horizon** h , on utilise les estimations de la tendance \hat{Q} et de la saisonnalité \hat{c} et on pose

$$\hat{X}_T(h) = \hat{Q}_{T+h} + \hat{c}_j, \quad T + h \equiv j[P].$$

Exemple 4.2 (suite)

Utilisons les estimées de la tendance et de la composante saisonnière. Ainsi à l'horizon 3 par exemple, on prévoit

$$\hat{X}_T(3) =$$

Pour juger de la qualité de nos estimations, on peut aussi faire des prévisions aux horizons -1 et 0 et les comparer aux vraies valeurs :

— à l'horizon -1, on prévoit

$$\hat{X}_T(-1) =$$

— à l'horizon 0 par exemple, on prévoit

$$\hat{X}_T(0) =$$

Remarque 4.10. Cette méthode de prévision a un inconvénient majeur qui est celui d'une prévision ne tenant pas compte des valeurs les plus récentes de la série : celles-ci ont été "éliminées" par application de moyennes mobiles. Pour cette raison, on utilise plus souvent la méthode comme méthode de désaisonnalisation que comme méthode de prévision.

4.7 Remarque : cas du modèle multiplicatif

On utilise parfois la méthode des moyennes mobiles dans le cadre du modèle multiplicatif ou du modèle mixte bien qu'elle ne soit pas particulièrement bien adaptée à ces contextes. Dans ce cas, l'algorithme est identique à celui décrit ci-dessus avec quelques adaptations selon le cas.

Dans le cas du modèle multiplicatif du type $X_t = Z_t(1 + S_t)(1 + \epsilon_t)$, les différences sont les suivantes :

— à l'étape 2, on pose

$$\tilde{S}_t = \frac{X_t}{X_t^*} - 1.$$

Puis on estime les coefficients saisonniers et on obtient les estimations \tilde{c}_j , $j = 1, \dots, P$ de la même manière que pour le modèle additif. On pose ensuite

$$\hat{c}_j = \frac{\tilde{c}_j}{\frac{1}{P} \sum_{j=1}^P \tilde{c}_j} - 1, \quad j = 1, \dots, P$$

de façon à respecter la contrainte de nullité en moyenne sur la période

$$\sum_{j=1}^P c_j = 0.$$

— à l'étape 3, on pose

$$X_{CVS,t} = \frac{X_t}{1 + \hat{S}_t}, \quad t = m + 1, \dots, T - m.$$

Dans le cas du modèle mixte du type $X_t = Z_t S_t + \epsilon_t$, l'algorithme est identique à celui décrit ci-dessus avec quelques adaptations selon le cas. Par exemple, pour un modèle multiplicatif du type $X_t = Z_t S_t + \epsilon_t$, les différences sont les suivantes :

— à l'étape 2, on pose

$$\tilde{S}_t = \frac{X_t}{X_t^*}.$$

Puis on estime les coefficients saisonniers et on obtient les estimations \tilde{c}_j , $j = 1, \dots, P$ de la même manière que pour le modèle additif. On pose ensuite

$$\hat{c}_j = \frac{\tilde{c}_j}{\frac{1}{P} \sum_{j=1}^P \tilde{c}_j}, \quad j = 1, \dots, P$$

de façon à respecter la contrainte de constance en moyenne sur la période

$$\sum_{j=1}^P c_j = P.$$

— à l'étape 3, on pose

$$X_{CVS,t} = \frac{X_t}{\hat{S}_t}, \quad t = m+1, \dots, T-m.$$

4.8 Analyse des résidus

Une fois estimées les composantes du modèle, on peut contrôler la pertinence du modèle par une analyse des résidus. Ceux-ci sont définis par

$$\hat{\varepsilon}_t = X_t - \hat{S}_t - \hat{Q}_t = \hat{X}_{CVS,t} - \hat{Q}_t.$$

Si le modèle est bon, il ne doit rester dans les résidus aucune trace du saisonnier. Pour le vérifier, on trace le corrélogramme des résidus c'est-à-dire le graphe d'un estimateur de la fonction d'autocorrélation. Pour la définition, les propriétés et un estimateur de la fonction d'autocorrélation, le lecteur est renvoyé au cours du second semestre.

Comme nous le verrons plus loin, le corrélogramme n'est tracé en théorie que dans le cas où la série est stationnaire, ce qui implique en particulier qu'il n'y a dans cette série ni tendance ni saisonnalité. En pratique, on s'en sert (dans le cas de l'analyse des résidus) pour vérifier justement l'absence de saisonnalité dans les résidus.

Si c'est le cas et si le modèle est bon, le corrélogramme ne doit présenter que des valeurs faibles, indiquant une faible corrélation entre les erreurs.

Si le corrélogramme présente une persistance, cela indique que la tendance n'a pas été complètement éliminée et c'est donc le signe que le modèle proposé a échoué. On peut alors réitérer la procédure ci-dessus ou proposer un autre modèle.

Si le corrélogramme présente des pics régulièrement espacés, cela indique que le saisonnier n'a pas été complètement éliminé et c'est donc le signe que le modèle proposé a échoué. On peut alors réitérer la procédure ci-dessus ou proposer un autre modèle.

Dans le cas où le corrélogramme (ou le périodogramme cf. Remarque 4.11) des résidus n'indique pas la présence d'un mouvement saisonnier, on trace le graphe des résidus $(t, \hat{\varepsilon}_t)$ qui sert à repérer d'éventuelles observations exceptionnelles, un mouvement tendanciel...

Dans le cas de l'hypothèse d'erreurs gaussiennes, on vérifie celle-ci en traçant l'histogramme des résidus, un qq-plot ou encore en effectuant un test de normalité... (cf. TP)

Exemple 4.2 (suite)

Dans les Figures 4.5, 4.6 et 4.7 sont représentés le corrélogramme, l'histogramme et le QQ-plot des résidus pour l'exemple 4.2.

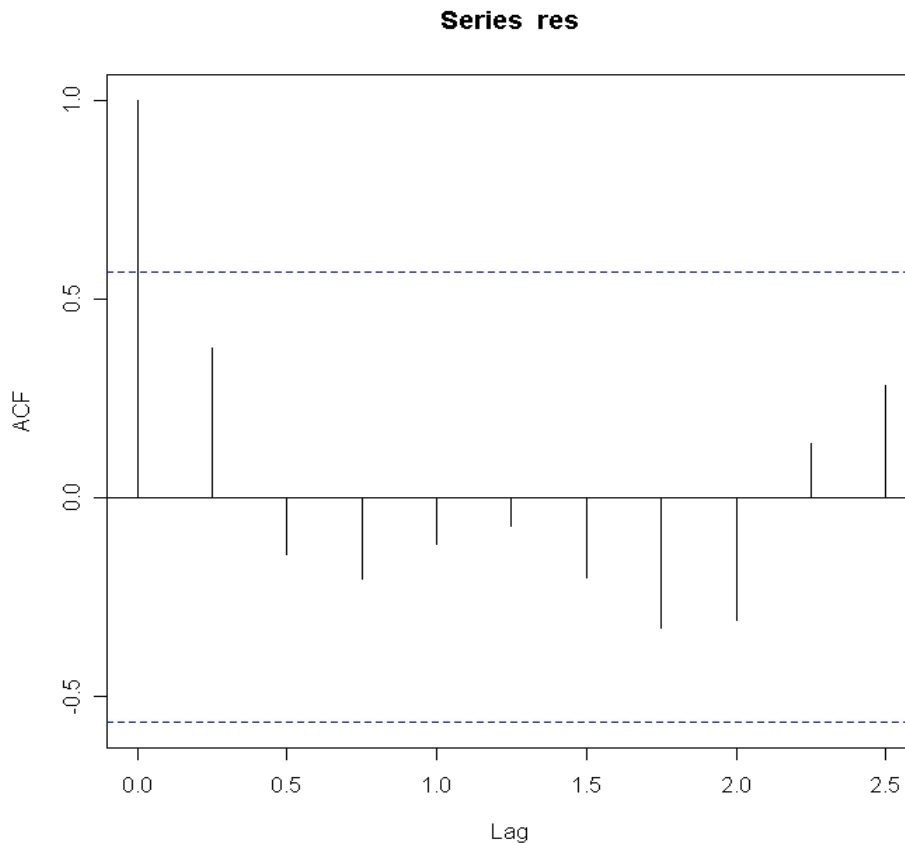


FIGURE 4.5 – Corrélogramme des résidus de l'exemple 4.2

Remarque 4.11. Rappelons que dans le cas où l'erreur du modèle est un bruit blanc, l'utilisation d'une moyenne mobile a introduit des corrélations dans le processus transformé.

Remarque 4.12. Pour mettre en évidence une éventuelle périodicité dans les résidus, on peut également tracer le **périodogramme**, qui est le graphe de la projection du vecteur $X = {}^t(X_1, \dots, X_T)$ sur le sous-espace vectoriel de \mathbb{R}^T engendré par les deux vecteurs

$$C = \begin{pmatrix} \cos(\omega \times 1) \\ \vdots \\ \cos(\omega \times T) \end{pmatrix} \quad \text{et} \quad S = \begin{pmatrix} \sin(\omega \times 1) \\ \vdots \\ \sin(\omega \times T) \end{pmatrix},$$

où ω est la **fréquence**. On cherche ainsi à ajuster à la série X_t le $t^{\text{ème}}$ terme d'une série trigonométrique, dite de Fourier, appelée une **harmonique**, et s'écrivant

$$s(t) = a \cos(\omega \times t) + b \sin(\omega \times t).$$

Dans le cas où la série X_t est périodique de période P , on a $P = 2\pi / \omega$.

Une approximation de la norme de cette projection est

$$Q(\omega) = \frac{T}{2} (\hat{a}^2 + \hat{b}^2),$$

où

$$\hat{a}^2 = \frac{2}{T} \sum_{t=1}^T X_t \cos(\omega t) \quad \text{et} \quad \hat{b}^2 = \frac{2}{T} \sum_{t=1}^T X_t \sin(\omega t).$$

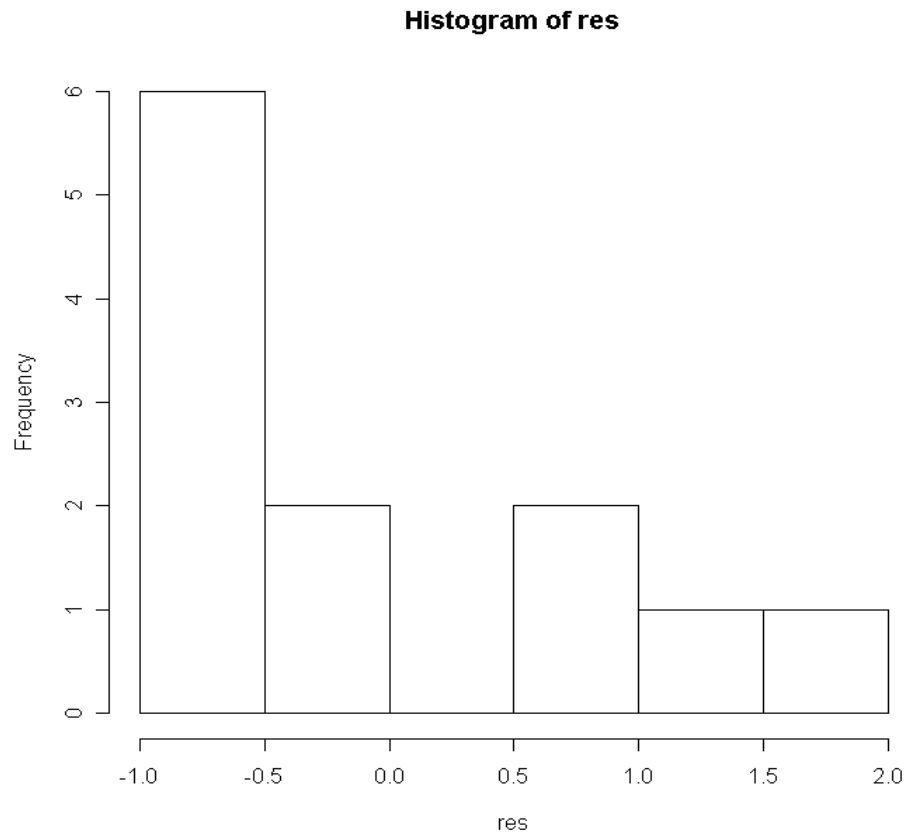


FIGURE 4.6 – Histogramme des résidus de l'exemple 4.2

Le périodogramme est donc le graphe de la fonction $(\omega, Q(\omega))$. On cherche alors la valeur ω^* qui maximise $Q(\omega)$ (c'est-à-dire qui minimise la norme de l'erreur de projection). Si le périodogramme présente un pic en ω^* , on lit cette valeur sur son graphe et on retient une période $T = \frac{2\pi}{\omega^*}$ ou plus exactement la valeur entière la plus proche de cette expression.

4.9 Un autre exemple

Prenons comme exemple les ventes trimestrielles de crèmes solaires d'un certain fabricant de 1994 à 1999. On suppose ici les composantes du modèle connues et que la série des ventes $(X_i)_{i=1\dots 24}$ est issue d'un modèle de type multiplicatif :

$$X_i = Z_i(1 + S_i)(1 + \epsilon_i).$$

Remarque 4.13. Nous verrons dans les chapitres suivants comment déterminer ces différentes composantes à partir de la seule série observée.

- **Tendance, facteurs saisonniers et fluctuations irrégulières**

- L'examen de la tendance nous montre que les ventes étaient stables jusqu'en 1996, ont augmenté en 1996 et 1997 (suite à une campagne de publicité), puis se sont stabilisées à partir de 1998 (la campagne de publicité ayant atteint ses limites).
- Les facteurs saisonniers (+0.3, -0.4, +0.8, -0.7) nous indiquent une augmentation des ventes de 30% au premier trimestre (vacances d'hiver) et de 80% au troisième trimestre (vacances d'été).

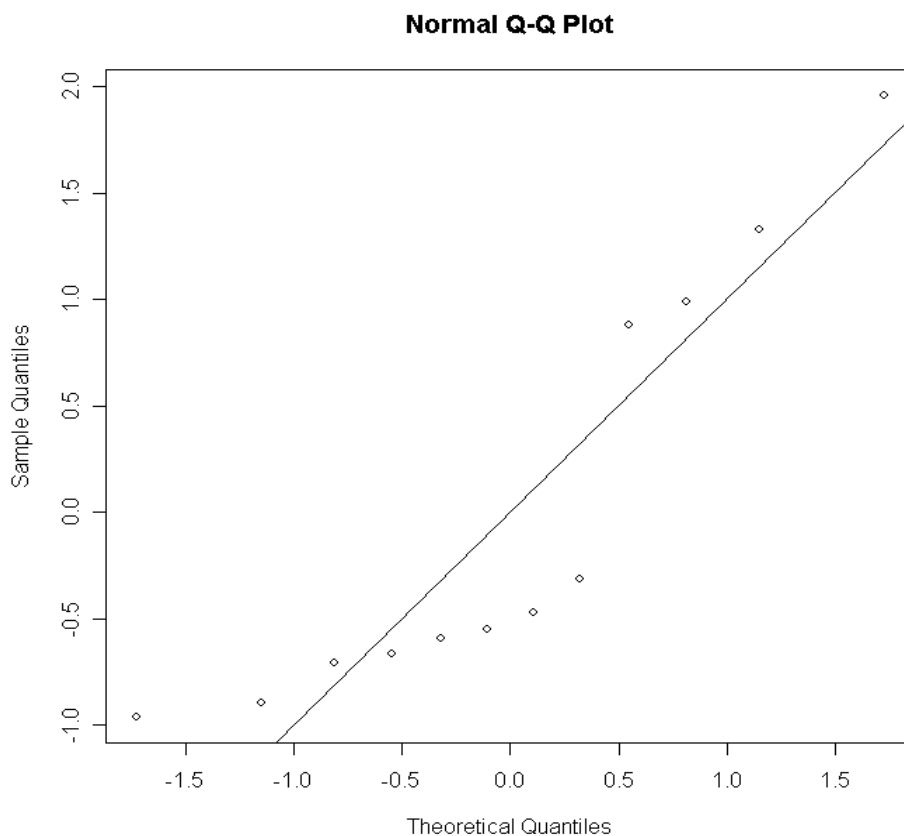


FIGURE 4.7 – QQ-plot des résidus de l'exemple 4.2

- L'étude des fluctuations irrégulières permet d'analyser précisément ce qui s'est passé chaque trimestre indépendamment de la tendance et de la saison. On remarque ainsi un "accident" au premier trimestre 1995 (augmentation des ventes de 40% due à une promotion) ainsi qu'au premier trimestre 1997 (chute accidentelle des ventes de 50% due à une grève).

• La série corrigée de la tendance

La tendance agit comme une forte corrélation entre les variables X_t mais cette corrélation n'exprime aucune liaison à caractère explicatif. Il s'agit donc d'isoler cette tendance puis de l'étudier à part et enfin de l'éliminer de la série pour voir si des liaisons à caractère explicatif existent et étudier seulement ces corrélations sans tendance. On définit la **série corrigée de la tendance** $(X_{CST,t})_t$ en supprimant la tendance. La série détendancialisée est

- pour le modèle additif : $X_{CST,t} = S_t + \epsilon_t$.
- pour le modèle multiplicatif : $X_{CST,t} = S_t(1 + \epsilon_t)$.

Nous verrons dans le Chapitre ?? comment enlever cette tendance et l'estimer une fois isolée.

• La série corrigée des variations saisonnières

Dans le même ordre d'idée, nous corrigerons les éventuelles variations saisonnières qui résultent d'un comportement périodique dans la série observée. Par exemple, considérons la figure représentant les ventes trimestrielles de crèmes solaires (Figure 4.8 en haut à gauche) : on observe une relation directe entre la période de l'année (par exemple l'automne) et la vente de crèmes solaires. Supposons par exemple que l'entreprise ait mis en place une mesure économique en été afin de réduire l'augmentation des prix des crèmes. L'observation de ventes plus faibles en automne qu'en été permet-il d'en déduire un effet de cette mesure ? Il faut se garder de conclure trop rapidement ;

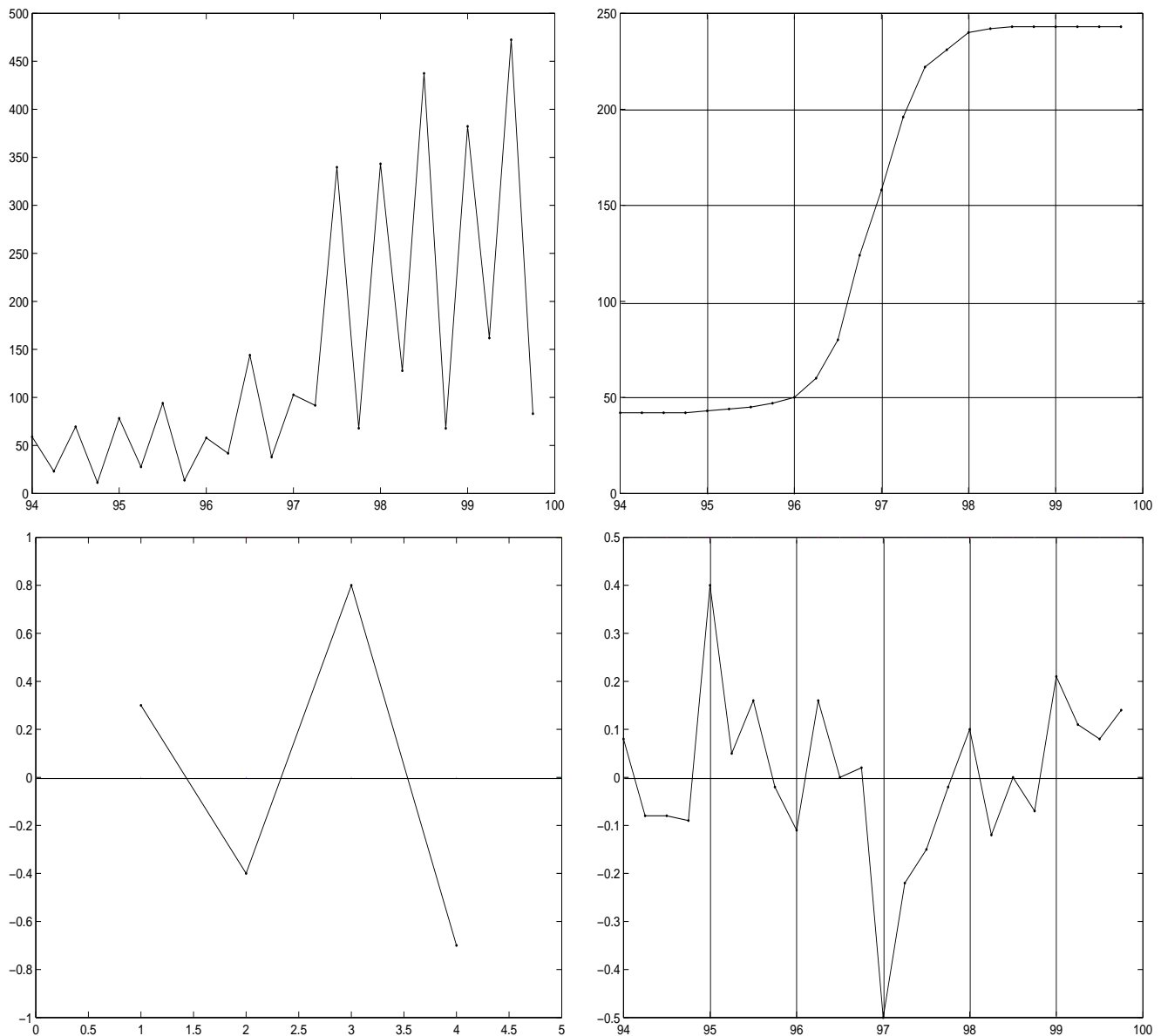


FIGURE 4.8 – Ventes trimestrielles de crèmes solaires (en ht à g), tendance (en ht à dte), facteurs saisonniers (en bas à g) et fluctuations irrégulières (en bas à dte)

l'observation de la série brute a montré que cette diminution du taux d'accroissement existait pour toutes les années; il faut donc savoir si elle est plus faible ou plus forte que d'habitude. L'observation des valeurs prises par la série corrigée des variations saisonnières, où l'effet de la saison est éliminé, permet de répondre directement à cette question. Toute théorie d'analyse et de prévision devra tenir compte de ce phénomène.

Pour pouvoir réellement comparer les ventes d'un trimestre à l'autre, on doit donc supprimer l'effet de la saisonnalité et on définit la **série corrigée des variations saisonnières** $(X_{CVS,t})_t$ en supprimant la composante saisonnière $(S_t)_t$ du modèle. La série désaisonnalisée est

- pour le modèle additif : $X_{CVS,t} = Z_t + \epsilon_t$.
- pour le modèle multiplicatif : $X_{CVS,t} = Z_t(1 + \epsilon_t)$.

Dans notre exemple, la série corrigée des variations saisonnières (Figure 4.9 à gauche) permet de mettre en évidence la progression des ventes entre 1995 et 1996, ainsi que les "accidents" survenus en 1995 et 1997.

Nous verrons dans le Chapitre 4 comment éliminer cette saisonnalité afin de nous concentrer sur les composantes aléatoires de la série chronologique puis l'estimer une fois isolée.

• **La série lissée des prédictions**

On définit la **série lissée des prédictions** $(X_{SLP,t})_t$ en supprimant les fluctuations irrégulières $(\epsilon_t)_t$ du modèle. C'est à partir de cette série que nous ferons les prédictions et en utilisant les modélisations et estimations de la tendance et de la saisonnalité. Par exemple, après avoir supprimé les fluctuations irrégulières, on obtient

- pour le modèle additif : $X_{SLP,t} = Z_t + S_t$.
- pour le modèle multiplicatif : $X_{SLP,t} = Z_t(1 + S_t)$.

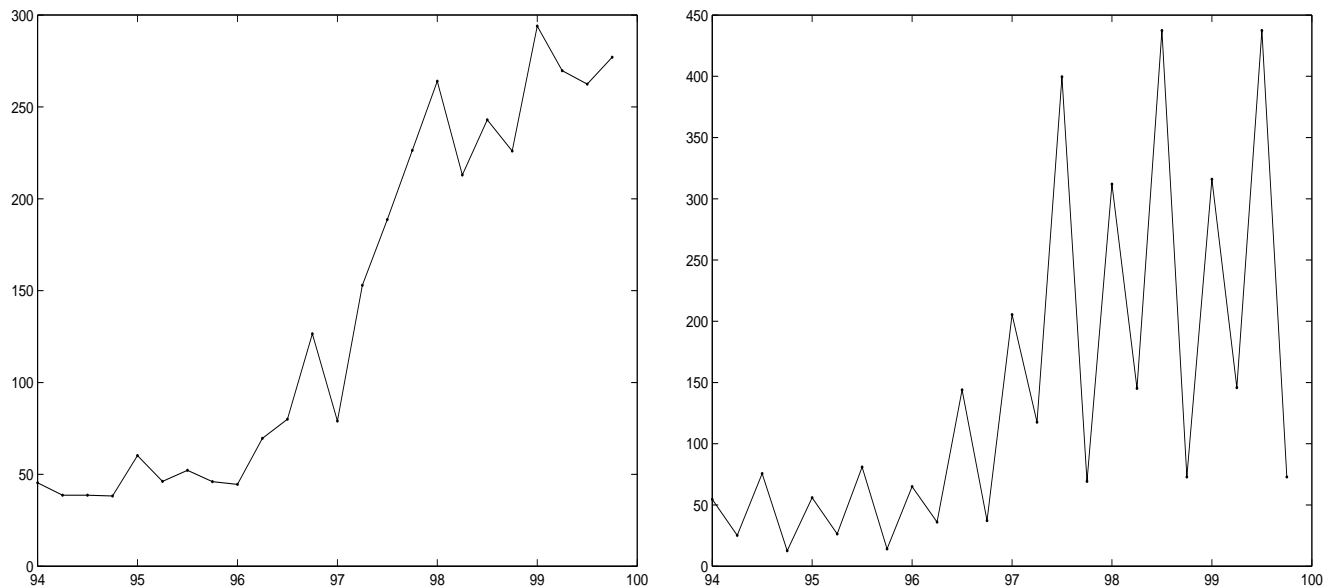


FIGURE 4.9 – Ventes trimestrielles de crèmes solaires : série corrigée des variations saisonnières (à gauche), et série lissée des prédictions (à droite)

Dans notre exemple, la série lissée des prédictions des ventes (Figure 4.9 à droite) permet de mettre en évidence la progression des ventes entre 1995 et 1996, ainsi que l'augmentation des ventes au 1er et 3èmes trimestres.

Chapitre 5

POUR ALLER PLUS LOIN - Prévision par lissage exponentiel

Introduites par Holt en 1958 ainsi que par 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 d'importance aux valeurs du passé d'une série temporelle. Les trois modèles ci-dessous seront traités dans ce chapitre :

- $\forall t \in \mathbb{Z}, X_t = Z_t + \epsilon_t;$
- $\forall t \in \mathbb{Z}, X_t = Z_t + S_t + \epsilon_t.$
- $\forall t \in \mathbb{Z}, X_t = Z_t S_t + \epsilon_t.$

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

Les techniques seront bien évidemment différentes selon que nous serons en présence de saisonnalité ou non.

5.1 Les lissages exponentiels

5.1.1 Le lissage exponentiel simple

Le lissage exponentiel simple permet d'effectuer des prévisions pour des séries chronologiques dont la tendance est constante et sans saisonnalité. Soit (X_t) une telle série dont on a observé les T premiers instants X_1, \dots, X_T . Pour $h \in \mathbb{N}^*$, on cherche à prévoir la valeur X_{T+h} , c'est-à-dire faire une prévision de la série à l'horizon h .

Étant donné un réel β tel que $0 < \beta < 1$, comme la tendance est constante, on cherche une prévision $\hat{X}_T(h)$ 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

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

Remarque 5.1. Notons que dans l'expression à minimiser l'influence des observations décroît lorsqu'on s'éloigne de la date T . Les dernières observations sont prises en compte ce qui constitue un avantage majeur par rapport à la méthode de prévision vue au chapitre précédent.

Définition 5.2. La prévision de la série à l'horizon h , $\hat{X}_T(h)$, fournie par la méthode de lissage exponentiel simple est donnée par

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

où β est la **constante de lissage**.

Exercice 15. Vérifier que la solution du problème de minimisation est bien donnée par

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

On fait ensuite l'approximation $\beta^T \approx 0$ pour obtenir le prédicteur du lissage exponentiel simple.

Démonstration. La solution de ce problème s'obtient facilement en annulant la dérivée de la fonction $g(a) = \sum_{j=0}^{T-1} \beta^j (X_{T-j} - a)^2$ et est donnée par

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

Lorsque T est assez grand, on peut négliger le terme $(1 - \beta^T)$ et poser

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

□

Remarque 5.3. 1. Dans le cas où β est indépendant de h , on notera simplement \hat{X}_T au lieu de $\hat{X}_T(h)$. Les prévisions sont dans ce cas identiques pour tout h .

2. Cette méthode de prévision prend en compte tout le passé d'une série temporelle, mais en accordant de moins en moins d'importance aux observations les plus éloignées de l'instant T (puisque β^j décroît avec j).
3. La valeur de la constante de lissage β permet de nuancer la remarque précédente.

Si β est proche de 0, la prévision est **souple**, c'est-à-dire fortement influencée par les observations les plus récentes (β^j devenant négligeable pour les grandes valeurs de j). Dans le cas extrême où $\beta = 0$, la prévision est alors égale à la dernière valeur observée.

En revanche, si β est proche de 1, l'influence des observations passées est d'autant plus importante et remonte loin dans le passé. On dit dans ce cas que la prévision est **rigide** en ce sens qu'elle est peu sensible aux fluctuations exceptionnelles (aussi appelées fluctuations conjoncturelles). Dans l'autre cas extrême où $\beta = 1$, alors toutes les prévisions sont identiques (et donc à la valeur choisie pour l'initialisation).

En pratique, on prend $\beta \in]0, 1[$ afin d'exclure ces deux cas extrêmes dégénérés.

A partir de la définition ci-dessus, on obtient directement

$$\hat{X}_T(h) = \beta \hat{X}_{T-1}(h) + (1 - \beta) X_T = \hat{X}_{T-1}(h) + (1 - \beta)(X_T - \hat{X}_{T-1}(h)).$$

Cette dernière égalité est appelée **formule de mise à jour** et permet de calculer directement (à partir de la prévision \hat{X}_{T-1} à la date $T - 1$) une nouvelle prévision \hat{X}_T lorsqu'une nouvelle observation X_T est effectuée.

Remarque 5.4. 1. Cette équation permet donc de mettre à jour les prévisions à l'horizon h à partir de la dernière prévision de manière extrêmement simple. L'initialisation de la récurrence est en général faite en prenant $\hat{X}_1(h) = X_1$. Un autre choix possible consiste à prendre la moyenne. Si T est assez grand, ce choix a en fait peu d'importance, comme le montre les quelques simulations de la Figure 5.2. Le choix de l'initialisation importe peu en réalité puisque cette initialisation est rapidement "oubliée". Cet oubli est d'autant plus rapide que la constante de lissage est proche de 1.

2. La première égalité fait apparaître $\hat{X}_T(h)$ comme le barycentre entre $\hat{X}_{T-1}(h)$, la valeur prédite à l'horizon h à partir des $T-1$ premières observations et X_T la dernière observation.
3. La seconde égalité fait elle intervenir $(X_T - \hat{X}_{T-1}(h))$ la dernière erreur de prévision.

Un problème important en pratique est celui du choix de la constante de lissage β 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 souhaite faire une prévision rigide, on choisira $\beta \in [0.7; 0.99]$ et si au contraire on souhaite une prévision souple, on choisira $\beta \in [0.01; 0.3]$. Une autre solution, dictée par les données, consiste à choisir β comme la solution du problème des moindres carrés ordinaires suivant

$$\sum_{t=1}^{T-h} (X_{t+h} - \hat{X}_t(h))^2 = (X_{1+h} - \hat{X}_1(h))^2 + (X_{2+h} - \hat{X}_2(h))^2 + \dots + (X_T - \hat{X}_{T-h}(h))^2,$$

c'est-à-dire de minimiser la somme des carrés des erreurs de prévision aux dates $1, \dots, T-h$.

On peut aussi ne considérer que les écarts obtenus sur la deuxième moitié de la série, afin de ne pas tenir compte de l'initialisation. On cherche alors le β qui minimise

$$\sum_{t=\lceil (T-h)/2 \rceil}^{T-h} (X_{t+h} - \hat{X}_t(h))^2.$$

La Figure 5.1 illustre le comportement du lissage exponentiel simple en prenant la constante de lissage égale à 0.8 (lissage rigide) ou 0.2 (lissage souple). Sur les quatre sous-figures, la courbe noire représente la série initiale, la courbe en pointillé rouge la série lissée avec $\beta = 0.8$ et la courbe en pointillé bleu la série lissée avec $\beta = 0.2$. Quatre cas sont présentés (de gauche à droite et de haut en bas) : tendance constante, tendance linéaire, tendance constante avec une valeur aberrante au milieu et tendance constante par morceaux (modèle avec rupture).

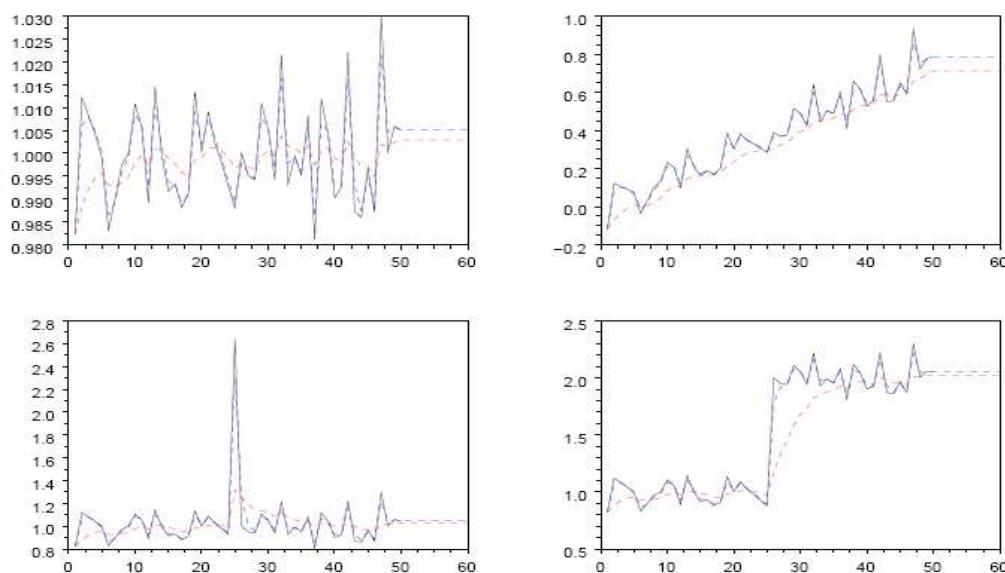


FIGURE 5.1 – Influence de la constante de lissage pour le lissage exponentiel simple : série temporelle initiale (trait noir), séries lissées avec $\beta = 0.8$ (trait pointillé rouge) et avec $\beta = 0.2$ (trait pointillé bleu).

La Figure 5.2 illustre l'influence de l'initialisation de l'équation de mise à jour; cette influence dépend aussi de la constante de lissage (à droite avec $\beta = 0.8$ et à gauche avec $\beta = 0.2$). Sur les deux

sous-figures, la série initiale apparaît en cercles noirs. La courbe rouge correspond au lissage en prenant $\hat{X}_1(h) = X_1$ et la courbe bleue correspond au lissage en prenant $\hat{X}_1(h) = \bar{X}$. On s'aperçoit que quelle que soit la valeur de la 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 β est proche de 1. 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.

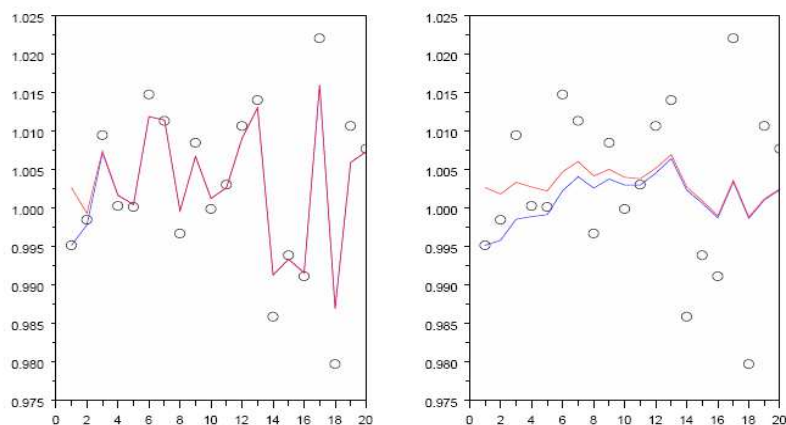


FIGURE 5.2 – Influence de l'initialisation pour le lissage exponentiel simple : série temporelle initiale (cercle noir), séries lissées avec initialisation en X_1 ou \bar{X} avec $\beta = 0.8$ (à droite) et avec $\beta = 0.2$ (à gauche).

5.1.2 Le lissage exponentiel double

Le lissage exponentiel double généralise l'idée du lissage exponentiel simple au cas où la série peut être ajustée par une droite au voisinage de T . On cherche dans ce cas une prévision à l'horizon h , $\hat{X}_T(h)$ de la forme

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

où le couple (\hat{a}_T, \hat{b}_T) minimise la fonction

$$\sum_{j=0}^{T-1} \beta^j (X_{T-j} - (aj + b))^2.$$

La solution de ce problème s'obtient en annulant les dérivées partielles de la fonction ci-dessus par rapport à a et b . En notant la série lissée

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

et la série doublement lissée

$$S_2(t) = (1 - \beta) \sum_{j=0}^{t-1} \beta^j S_1(t - j),$$

on déduit la définition suivante

Définition 5.5. La prévision de la série à l'horizon h , $\hat{X}_T(h)$, fournie par la méthode de lissage exponentiel double est donnée par

$$\hat{X}_T(h) = \hat{a}_T h + \hat{b}_T$$

où β est la **constante de lissage** et le couple (\hat{a}_T, \hat{b}_T) est donné par

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

Remarque 5.6. Les expressions de \hat{a} et \hat{b} dépendent de $S_1(T)$ et $S_2(T)$ qui sont respectivement le lissage exponentiel simple de la série initiale et le lissage exponentiel simple de la série lissée. On a donc effectué deux lissages consécutifs d'où le nom de lissage exponentiel double.

Exercice 16. En remplaçant, lorsque T est grand, $\sum_{j=0}^{T-1} \beta^j$ par $\frac{1}{1-\beta}$, $\sum_{j=0}^{T-1} \beta^j j$ par $\frac{\beta}{(1-\beta)^2}$ et $\sum_{j=0}^{T-1} \beta^j j^2$ par $\frac{\beta(1+\beta)}{(1-\beta)^3}$, vérifier que le couple (\hat{a}_T, \hat{b}_T) solution du problème de minimisation est bien donné par les relations de la définition précédente.

Démonstration. Notons g la fonction définie par

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

On doit donc résoudre le système

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

soit

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

avec

$$T_1(X) := \sum_{j=0}^{T-1} j \beta^j X_{T-j} \quad \text{et} \quad T_2(X) := \sum_{j=0}^{T-1} \beta^j X_{T-j}.$$

Mais lorsque T est assez grand, on peut négliger le terme $(1 - \beta^T)$ et remplacer $\sum_{j=0}^{T-1} \beta^j$ par $\frac{1}{1-\beta}$, $\sum_{j=0}^{T-1} \beta^j j$ par $\frac{\beta}{(1-\beta)^2}$ et $\sum_{j=0}^{T-1} \beta^j j^2$ par $\frac{\beta(1+\beta)}{(1-\beta)^3}$. Ainsi il s'agit de résoudre

$$\begin{cases} a \frac{\beta(1+\beta)}{1-\beta} + b\beta = (1-\beta)^2 T_2(X) \\ a \frac{\beta}{1-\beta} + b = (1-\beta) T_1(X) \end{cases}$$

D'où l'on déduit

$$\begin{cases} \hat{a}_T = \frac{(1-\beta)^2}{\beta} [(1-\beta) T_2(X) - \beta T_1(X)] \\ \hat{b}_T = (1-\beta^2) T_1(X) - (1-\beta)^2 T_2(X) \end{cases}$$

On déduit le résultat en développant $S_2(T)$ et en remarquant que

$$S_1(T) = (1-\beta) T_1(X) \quad \text{et} \quad S_2(T) = (1-\beta)^2 (T_1(X) + T_2(X)).$$

□

Exercice 17. Déterminer les formules de mise à jour du lissage exponentiel double.

Démonstration. Les formules de mise à jour s'obtiennent à partir de ces expressions

$$\hat{a}_T = \hat{a}_{T-1} + (1-\beta)^2 (X_T - \hat{X}_{T-1}(1)),$$

et

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

□

Pour utiliser ces formules de mise à jour, il faut avoir des valeurs initiales pour les suites \hat{a}_t et \hat{b}_t . On prend en général $\hat{a}_2 = X_2 - X_1$ et $\hat{b}_2 = X_2$. Notons enfin que pour sélectionner la constante de lissage, on peut utiliser un critère similaire à celui introduit à la section précédente.

La Figure 5.3 illustre le comportement du lissage exponentiel double en prenant comme précédemment la constante de lissage égale à 0.8 ou 0.2. Sur les quatre sous-figures, la courbe noire représente toujours la série initiale, la courbe en pointillé rouge la série lissée avec $\beta = 0.8$ et la courbe en pointillé bleu la série lissée avec $\beta = 0.2$. Quatre cas sont présentés (de gauche à droite et de haut en bas) : tendance linéaire, tendance quadratique, tendance linéaire avec une valeur aberrante au milieu et tendance linéaire par morceaux (modèle avec rupture).

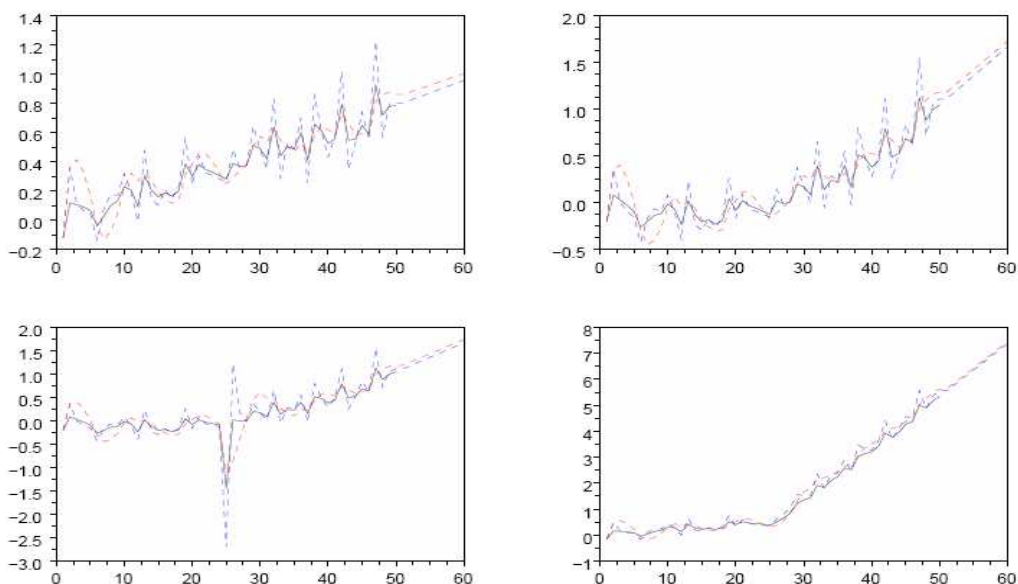


FIGURE 5.3 – Influence de la constante de lissage pour le lissage exponentiel double : série temporelle initiale (trait noir), séries lissées avec $\beta = 0.8$ (trait pointillé rouge) et avec $\beta = 0.2$ (trait pointillé bleu).

Remarque 5.7. On peut généraliser cette technique de lissage pour traiter des séries sans saisonnalité présentant une tendance polynomiale de degré supérieur à 2. Les résultats font intervenir, dans ce cas, les opérateurs de lissage d'ordre p $S_p(t)$, $p \in \mathbb{N}$, itérées d'ordre p de $S_1(t)$.

5.2 La méthode de Holt-Winters

La méthode de lissage exponentiel double permet de traiter des séries présentant une tendance linéaire mais sans saisonnalité. On peut également définir des lissages exponentiels généralisés sur le même principe que les techniques décrites dans les sections précédentes permettant de traiter des séries avec saisonnalité.

Une méthode un peu différente a été introduite par Holt et Winters. Il existe une version non saisonnière de cette méthode, c'est-à-dire adaptée aux séries sans saisonnalité pouvant être ajustées par une droite au voisinage de T (comme pour le lissage exponentiel double). La différence entre la méthode de Holt-Winters et le lissage exponentiel double porte sur les formules de mise à jour.

5.2.1 La méthode non saisonnière

La formule de mise à jour pour le lissage exponentiel double peut aussi s'écrire sous la forme

$$\hat{a}_T = \frac{1-\beta}{2} (\hat{X}_T(1) - \hat{X}_{T-1}(1)) + \frac{1+\beta}{2} \hat{a}_{T-1},$$

et

$$\hat{b}_T = (1-\beta^2)X_T + \beta^2 (\hat{X}_{T-1}(1) + \hat{X}_{T-1}(2)).$$

Ainsi écrits, \hat{a}_T et \hat{b}_T apparaissent comme des barycentres. Holt et Winters ont alors proposé une version de ces formules de mise à jour où les pondérations ne dépendent pas que d'un seul paramètre mais de deux paramètres :

$$\hat{a}_T = (1-\gamma) (\hat{X}_T(1) - \hat{X}_{T-1}(1)) + \gamma \hat{a}_{T-1},$$

et

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

où $|\alpha| < 1$ et $|\gamma| < 1$. L'initialisation peut se faire comme dans le cas du lissage exponentiel double. L'avantage de cette approche est d'avoir une plus grande flexibilité mais la contrepartie est de devoir régler deux paramètres. Si α et γ sont proches de 1 tous les deux, la prévision est lisse (fort poids du passé). La prévision par cette méthode est donnée par

$$\hat{X}_T(h) = \hat{a}_T h + \hat{b}_T.$$

Exercice 18. Pour quelles valeurs de α et γ retrouve-t-on le lissage exponentiel double ?

Démonstration. On retrouve le lissage exponentiel double de Brown si $\alpha = \beta^2$ et $\gamma = \frac{2\beta}{1-\beta^2}$. □

Étudions graphiquement l'effet induit par les constantes de lissage pour les mêmes quatre cas-tests que ceux vus pour le lissage exponentiel double. La Figure 5.4 illustre le comportement du lissage de Holt-Winters en prenant comme précédemment les constantes de lissage égale à 0.8 ou 0.2. Sur les quatre sous-figures, la courbe noire représente toujours la série initiale, la courbe en pointillé rouge la série lissée avec $\alpha = 0.8$ et $\gamma = 0.8$, la courbe en pointillé magenta la série lissée avec $\alpha = 0.8$ et $\gamma = 0.2$, la courbe en pointillé vert la série lissée avec $\alpha = 0.2$ et $\gamma = 0.8$ et la courbe en pointillé bleu la série lissée avec $\alpha = 0.2$ et $\gamma = 0.2$. Quatre cas sont présentés (de gauche à droite et de haut en bas) : tendance linéaire, tendance quadratique, tendance linéaire avec une valeur aberrante au milieu et tendance linéaire par morceaux (modèle avec rupture).

Le choix des constantes de lissage est encore plus subjectif et peut être réglé automatiquement en utilisant un critère des moindres carrés comme proposé précédemment.

5.2.2 La méthode saisonnière additive

On considère une série chronologique dont on a observé les T premiers instants X_1, \dots, X_T et on suppose que cette série peut être approchée, au voisinage de T , par le modèle

$$a(t-T) + b + S_t,$$

où S_t représente la saisonnalité de période P . La méthode de Holt-Winters propose pour l'estimation de a , b et S_t les formules de mise à jour suivantes

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

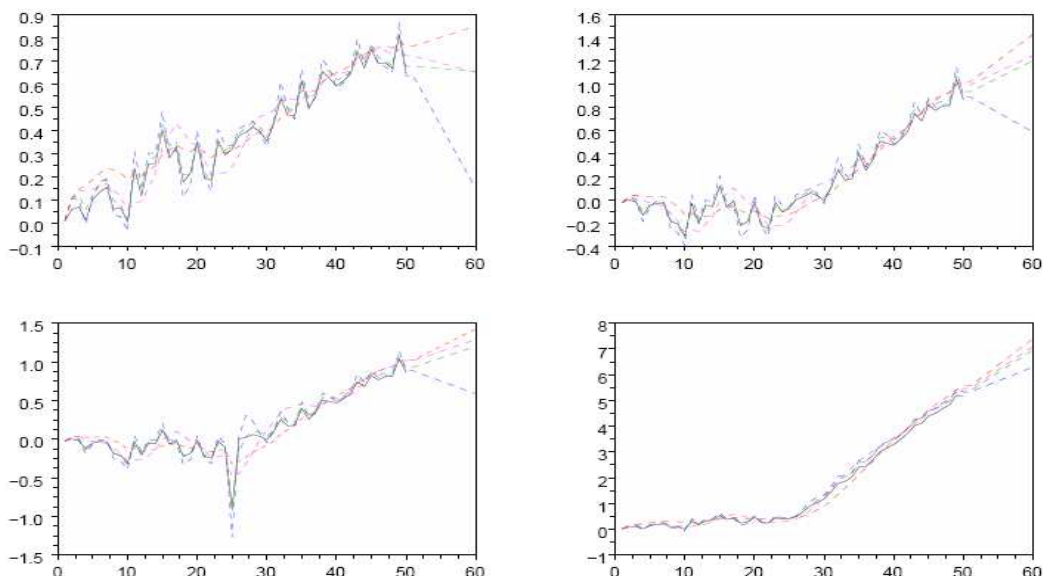


FIGURE 5.4 – Influence des constantes de lissage pour la méthode de Holt-Winters sans saisonnalité : série temporelle initiale (trait noir) et quatre séries lissées (traits pointillés).

où α , β et γ sont des constantes de lissage appartenant à $]0, 1[$. La première formule de mise à jour s'interprète comme une moyenne pondérée de la différence des niveaux estimés aux instants T et $T - 1$ et la pente estimée à l'instant $T - 1$; la deuxième comme une moyenne pondérée de l'observation X_T (à laquelle on a retranché la composante saisonnière estimée à l'étape précédente) et l'estimation de la tendance faite à l'instant $T - 1$; enfin, la troisième comme une moyenne pondérée de l'observation X_T (à laquelle on a retranché le niveau calculé à l'instant T) et de la composante saisonnière calculée à l'instant $T - P$.

On en déduit une prévision à l'horizon h

$$\hat{X}_T(h) = \begin{cases} \hat{a}_T h + \hat{b}_T + \hat{S}_{T+h-P}, & \text{si } 1 \leq h \leq P \\ \hat{a}_T h + \hat{b}_T + \hat{S}_{T+h-2P}, & \text{si } P+1 \leq h \leq 2P \\ \dots & \dots \end{cases}$$

Remarque 5.8. 1. Le choix des constantes de lissage dans la pratique peut s'effectuer de la même manière que précédemment, c'est-à-dire en minimisant la somme des carrés des erreurs de prévision.

2. On doit calculer les valeurs initiales pour utiliser les formules de mise à jour ci-dessus.

Exemple 5.9. Soit (X_t) une série chronologique observée jusqu'à l'instant T , de tendance linéaire et de composante saisonnière de période 4. On souhaite appliquer la méthode de Holt-Winters et approcher la série, au voisinage de T , par le modèle

$$a(t - T) + b + S_t + \epsilon_t.$$

Nous allons voir dans cet exemple comment initialiser les formules de mise à jour.

1. Appliquer une moyenne mobile adaptée à la série afin de supprimer la saisonnalité. On obtient ainsi une estimation de la tendance linéaire.

2. Appliquer un opérateur adapté à la série transformée afin de récupérer le coefficient a . L'estimer. En déduire une estimation du b .
3. En déduire une estimation du saisonnier.

Démonstration. Tout d'abord, appliquons

- d'une part la moyenne mobile arithmétique modifiée d'ordre 4 au temps 4, à la série; on obtient la tendance puisque cette moyenne mobile annule la saisonnalité et conserve les polynômes de degré 1. On déduit donc une estimation de $a(t - T) + b$:

$$M_4(4) = \frac{1}{4} \left(\frac{1}{2} X_2 + X_3 + X_4 + X_5 + \frac{1}{2} X_6 \right).$$

- d'autre part l'opérateur $(I - B)$, appelé opérateur de différenciation, à $M_4(4)$; on obtient le coefficient directeur de la tendance puisque l'opérateur $(I - B)$ transforme un polynôme de degré p en un polynôme de degré $p - 1$ (cf. feuille d'exercices). On déduit donc une estimation de a :

$$\begin{aligned} (I - B)M_4(4) &= M_4(4) - M_4(3) \\ &= \frac{1}{4} \left(\frac{1}{2} X_2 + X_3 + X_4 + X_5 + \frac{1}{2} X_6 \right) - \frac{1}{4} \left(\frac{1}{2} X_1 + X_2 + X_3 + X_4 + \frac{1}{2} X_5 \right). \end{aligned}$$

C'est pour cela que nous avons besoin de commencer par calculer $M_4(4)$ même s'il n'utilise pas X_1 .

Donc $\hat{a}_4 = M_4(4) - M_4(3)$. Puis comme $M_4(4)$ estime la tendance lorsqu'on a 4 observations et au point 4 i.e. $\hat{a}_4(4 - 4) + \hat{b}_4$, on en déduit une estimation de b ,

$$\hat{b}_4 = M_4(4) \quad \text{et} \quad \hat{b}_3 = M_4(3).$$

Ainsi

$$\hat{a}_4 = \hat{b}_4 - \hat{b}_3, \quad \hat{b}_3 = M_4(3) \quad \text{et} \quad \hat{b}_4 = M_4(4).$$

Puis on approche S_t par

$$X_t - a(t - T) - b$$

et on déduit les quatre valeurs initiales pour la saisonnalité :

$$\begin{cases} \hat{S}_4 &= X_4 - \hat{b}_4 - (4 - 4)\hat{a}_4 = X_4 - \hat{b}_4 \\ \hat{S}_3 &= X_3 - \hat{b}_4 - (3 - 4)\hat{a}_4 = X_3 - \hat{b}_3 \\ \hat{S}_2 &= X_2 - \hat{b}_4 - (2 - 4)\hat{a}_4 = X_2 - \hat{b}_3 + \hat{a}_4 \\ \hat{S}_1 &= X_1 - \hat{b}_4 - (1 - 4)\hat{a}_4 = X_1 - \hat{b}_3 + 2\hat{a}_4. \end{cases}$$

□

5.2.3 La méthode saisonnière multiplicative

On considère une série chronologique dont on a observé les T premiers instants X_1, \dots, X_T et on suppose que cette série peut être approchée, au voisinage de T , par le modèle

$$(a(t - T) + b)S_t,$$

où S_t représente la saisonnalité de période P . La méthode de Holt-Winters propose pour l'estimation de a , b , S_t les formules de mise à jour suivantes

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

où α , β et γ sont des constantes de lissage appartenant à $]0, 1[$. Les formules de mise à jour s'interprètent comme dans le cas de la méthode saisonnière additive.

On en déduit une prévision à l'horizon h

$$\hat{X}_T(h) = \begin{cases} (\hat{a}_T h + \hat{b}_T) \hat{S}_{T+h-P}, & \text{si } 1 \leq h \leq P \\ (\hat{a}_T h + \hat{b}_T) \hat{S}_{T+h-2P}, & \text{si } P+1 \leq h \leq 2P \\ \dots & \dots \end{cases}$$

Remarque 5.10. Pour déterminer les valeurs initiales afin d'utiliser les formules de mise à jour, dans le cas d'une période 4, on détermine \hat{b}_3 , \hat{a}_4 et \hat{b}_4 comme à la section précédente. De même les coefficients saisonniers initiaux sont obtenus comme précédemment mais en divisant les observations par la tendance linéaire.

Chapitre 6

Annexes

A. Rappels sur la régression par moindres carrés

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 :

$$Y \simeq aX + b,$$

où les coefficients a et b sont inconnus. Le problème est donc d'estimer ces coefficients grâce aux valeurs observées sur l'échantillon. Si les points du nuage sont parfaitement alignés (sur une même droite), il serait facile de donner des valeurs à a et b : il suffirait en effet de prendre pour a la pente de la droite sur laquelle se trouvent les points du nuage et pour b la valeur en $x = 0$ (la solution se trouve en résolvant un système de deux équations à deux inconnues à partir de deux points du nuage). Le problème est que les points du nuage sont rarement (parfaitement) alignés : ils sont "proches" d'une droite.

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 : $\min_{a,b} \sum_{i=1}^n |y_i - ax_i - b|$.
- la somme des erreurs au carré : $\min_{a,b} \sum_{i=1}^n (y_i - ax_i - b)^2$.

La méthode des **moindres carrés** minimisant le second critère est la plus usité et décrite à la section suivante.

A1. La méthode des moindres carrés

On démontre en minimisant la fonction de deux variables

$$g(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

que le couple solution (\hat{a}, \hat{b}) est donné par

$$\hat{a} = \frac{\frac{1}{n} \sum_{i=1}^n y_i x_i - \left(\frac{1}{n} \sum_{i=1}^n y_i\right) \left(\frac{1}{n} \sum_{i=1}^n x_i\right)}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad \text{et} \quad \hat{b} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{a} \frac{1}{n} \sum_{i=1}^n x_i = \bar{Y} - \hat{a} \bar{X}.$$

La droite d'équation $y = \hat{a}x + \hat{b}$ est appelée **droite de régression de Y en X** et est notée : $\Delta_{Y/X}$.

Propriétés :

- 1) Cette droite passe par le point moyen $M(\bar{X}; \bar{Y})$. Puisqu'il suffit de deux points pour tracer une droite, on pourra, pour tracer $\Delta_{Y/X}$, placer les points $B(0; b)$ et $M(\bar{X}; \bar{Y})$
- 2) Le coefficient directeur \hat{a} de $\Delta_{Y/X}$, $\text{Cov}(X, Y)$ et $r(X, Y)$ (voir la définition dans le paragraphe suivant) sont de même signe :
 - lorsqu'ils sont positifs, on parle de **corrélacion positive** (y augmente quand x augmente).
 - lorsqu'ils sont négatifs, on parle de **corrélacion négative** (y diminue quand x augmente).

Rappels :

- 1) On rappelle qu'une équation de droite donne la relation entre l'abscisse (lue horizontalement) et l'ordonnée (lue verticalement) d'un point de la droite. Ainsi pour une droite d'équation $y = ax + b$, le point de la droite d'abscisse x_i aura pour ordonnée $ax_i + b$.
- 2) Le nombre a est appelé la **pente de la droite** ou encore **coefficient directeur** de la droite car il détermine la direction (la pente) de la droite. Lorsque a est positif la droite est croissante, lorsque a est négatif la droite est décroissante.
Le nombre b s'appelle l'**ordonnée à l'origine** car c'est l'ordonnée du point de la droite d'abscisse 0 (intersection de la droite avec l'axe des ordonnées).
- 3) Deux points (et une règle) suffisent pour tracer une droite. Pour représenter une droite lorsqu'on connaît son équation, il suffit de placer deux points (par exemple les points de coordonnées $(0; b)$ et $(1; a + b)$) puis tracer la droite passant par ces 2 points.

A2. Propriétés et interprétation du coefficient de corrélacion linéaire

Afin de confirmer qu'il est raisonnable d'approximer le nuage de points par une droite, on calcule le **coefficient de corrélacion linéaire** encore appelé **coefficient de Bravais-Pearson** :

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

Propriétés :

- a) Le coefficient de corrélacion linéaire est symétrique :

$$r(X, Y) = r(Y, X).$$

- b) L'inégalité de Cauchy-Schwarz donne :

$$-1 \leq r(X, Y) \leq 1.$$

- c) **Transformation affine des données.** Soient a, b, c et d quatre nombres réels quelconques ($a \neq 0$ et $c \neq 0$). Posons $Z = aX + b$ et $T = cY + d$. On a alors :

$$r(Z, T) = \begin{cases} r(X, Y), & \text{si } a \text{ et } c \text{ sont de même signe,} \\ -r(X, Y), & \text{si } a \text{ et } c \text{ sont de signes opposés.} \end{cases}$$

En particulier, pour $a = c = 0$, on voit que le coefficient de corrélacion linéaire est invariant par translations et pour $b = d = 0$, il est invariant au signe près par homothéties.

On peut à nouveau utiliser ces relations pour simplifier les calculs du coefficient de corrélacion linéaire.

— **Si $r(X, Y) = 0$.**

Dans ce cas l'éloignement des points du nuage avec la droite de régression de Y en X est maximal. On dira alors que X et Y sont **linéairement indépendants**.

— **Si $r(X, Y) > 0$.**

Dans ce cas la droite de régression de Y en X est croissante; on parle alors de **corrélacion linéaire croissante** entre X et Y . Lorsque $r(X, Y)$ est proche de 1, les points du nuage sont donc presque alignés, on a donc une forte corrélation linéaire croissante (ou positive) entre X et Y .

Dans le cas extrême $r(X, Y) = 1$, les points du nuage sont alors parfaitement alignés, on peut donc parler de **corrélacion linéaire croissante totale** : pour un individu, sa donnée suivant X détermine entièrement sa donnée suivant Y .

Arbitrairement, on considèrera la corrélation linéaire croissante **faible** lorsque $0 < r(X, Y) < 0,3$, **moyenne** lorsque $0,3 \leq r(X, Y) \leq 0,7$ et **forte** lorsque $r > 0,7$.

— **Si $r(X, Y) < 0$.**

Dans ce cas la droite de régression de Y en X est décroissante; on parle alors de corrélation linéaire décroissante entre X et Y . Lorsque $r(X, Y)$ est proche de -1, les points du nuage sont donc presque alignés, on a donc une forte corrélation linéaire décroissante (ou négative) entre X et Y .

Dans le cas extrême $r(X, Y) = -1$, les points du nuage sont alors parfaitement alignés, on peut donc parler de **corrélacion linéaire décroissante totale** : pour un individu, sa donnée suivant X détermine entièrement sa donnée suivant Y .

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

Voici quelques exemples de nuages de points avec la valeur du coefficient de Bravais-Pearson qui permettent de mieux comprendre ce que traduit la valeur de $r(X, Y)$ au niveau de nuage de points : **En pratique, il faut commencer par tracer le nuage de points puis calculer $r(X, Y)$ et ce n'est que si la corrélation linéaire est assez forte que l'on cherchera la droite de régression de Y en X .**

A3. Ajustement tendanciel linéaire par moindres carrés

Supposons que l'on observe T valeurs d'une série dont la tendance semble être linéaire comme dans l'exemple (a) de la Figure 1.1. La méthode des moindres carrés décrite au paragraphe précédent consiste à estimer la tendance par une fonction linéaire

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

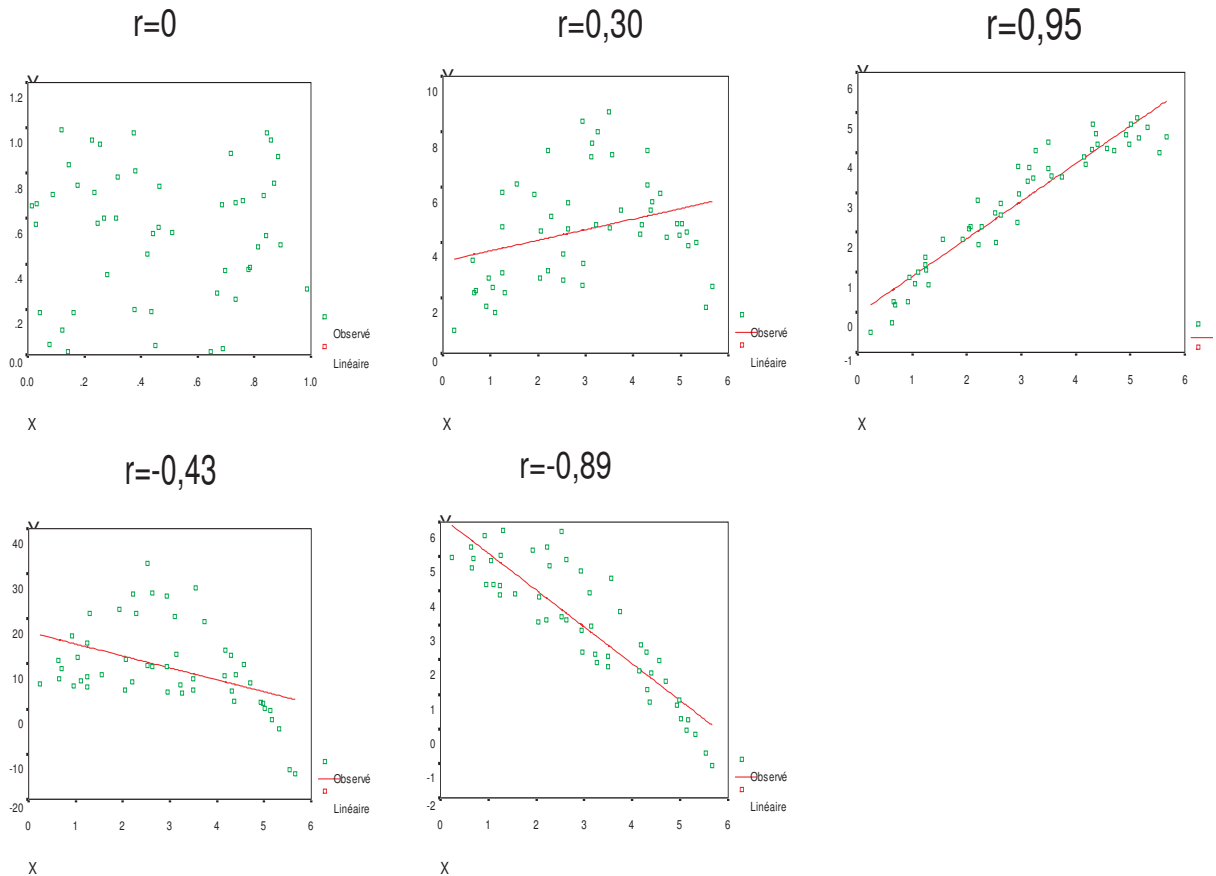
Le couple solution (\hat{a}, \hat{b}) est donné par

$$\hat{a} = \frac{\text{Cov}(t, Z_t)}{\text{Var}(t)} \quad \text{et} \quad \hat{b} = \bar{Z} - \hat{a}\bar{t},$$

en posant $\bar{t} = \frac{1}{T} \sum_{t=1}^T t$, $\bar{Z} = \frac{1}{T} \sum_{t=1}^T Z_t$ et

$$\begin{cases} \text{Cov}(t, Z_t) = \frac{1}{T} \sum_{t=1}^T (t - \bar{t})(Z_t - \bar{Z}) = \frac{1}{T} \sum_{t=1}^T t \cdot Z_t - \bar{t} \cdot \bar{Z} \\ \text{Var}(t) = \frac{1}{T} \sum_{t=1}^T (t - \bar{t})^2 \end{cases}$$

Remarquons que



- Le point moyen de coordonnées (\bar{t}, \bar{Z}) appartient à la droite des moindres carrés.
- Le coefficient de corrélation linéaire est défini par

$$r = \frac{\text{Cov}(t, Z_t)}{\sqrt{\text{Var}(t) \text{Var}(Z_t)}}.$$

La corrélation linéaire entre la date t et la variable Z_t est d'autant plus importante que $|r|$ est proche de 1.

B. Résumé de la procédure et des notations

La décomposition et l'étude de la série statistique (X_t) en vue de la prédiction se font selon les étapes suivantes

1. application d'une moyenne mobile :

- on applique une moyenne mobile d'ordre judicieusement choisi (généralement égal à la période de la saisonnalité).
- on récupère ainsi la série lissée

$$X_t^* = M_P(X_t).$$

2. estimation de la saisonnalité :

- on calcule la série diminuée de sa tendance

$$\tilde{S}_t = X_t - X_t^*.$$

- on estime chaque coefficient saisonnier c_j par la moyenne sur les périodes \tilde{c}_j :

$$\tilde{c}_j = \text{moyenne de } \tilde{S}_t \text{ sur les saisons } j.$$

- on retranche à chaque coefficient \tilde{c}_j la quantité $\frac{1}{P} \sum_{k=1}^P \tilde{c}_k$ de façon à satisfaire la condition de somme nulle sur une période :

$$\hat{c}_j = \tilde{c}_j - \frac{1}{P} \sum_{k=1}^P \tilde{c}_k.$$

3. estimation de la tendance :

- on calcule la série corrigée des variations saisonnières

$$X_{CVS,t} = X_t - \hat{S}_t.$$

- on estime cette série par des méthodes de régression.
— on récupère ainsi la série

$$\hat{Q}_t.$$

4. itération éventuelle de la procédure.

5. prévision des valeurs futures à l'horizon h par

$$\hat{X}_T(h) = \hat{Q}_{T+h} + \hat{S}_{T+h}.$$

6. analyse des résidus.

C. Petit lexique des moyennes mobiles

Dire qu'une série $(X_t)_t$

1. est **conservée/n'est pas modifiée/est invariante** par l'application d'une moyenne mobile M signifie que

$$MX_t = M(X_t) = X_t^* = X_t;$$

2. est **arrêtée/supprimée/annulée/absorbée** par l'application d'une moyenne mobile M signifie que

$$MX_t = M(X_t) = X_t^* = 0.$$

D. Rappels de probabilités

- On appelle **espérance** d'une variable aléatoire discrète X et on note $\mathbb{E}[X]$ le nombre défini par

$$\mathbb{E}[X] = \sum_{i \in \mathbb{N}} x_i \mathbb{P}(X = x_i) = x_1 p_1 + \dots + x_n p_n$$

où $\{x_0, x_1, \dots\}$ représente le support de X .

Remarque : $\mathbb{E}[X] = \frac{\sum_{i \in \mathbb{N}} x_i \mathbb{P}(X = x_i)}{\sum_{i \in \mathbb{N}} \mathbb{P}(X = x_i)}$.

$\mathbb{E}[X]$ apparaît ainsi comme le barycentre des x_i affectés des masses $\mathbb{P}(X = x_i)$; c'est donc la moyenne des valeurs prises par X pondérées par leur probabilité. Ainsi dans le cadre de n observations discrètes, l'espérance sera donnée par

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i.$$

Propriété : $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

- On appelle **variance** d'une variable aléatoire discrète X et on note $\text{Var}(X)$ le nombre défini par

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}[X])^2) = \sum_{i \in \mathbb{N}} (x_i - \mathbb{E}[X])^2 \mathbb{P}(X = x_i) = (x_1 - \mathbb{E}[X])^2 p_1 + \dots + (x_n - \mathbb{E}[X])^2 p_n$$

où $\{x_0, x_1, \dots\}$ représente le support de X .

Remarque : 1) La variance est une quantité positive.

2) Dans le cadre de n observations, la variance sera donnée par

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mathbb{E}[X])^2.$$

Propriété :

— $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

— Formule de Koenig-Huyghens

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

- On appelle **écart-type** d'une variable aléatoire discrète X et on note $\sigma(X)$ le nombre défini par

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

Remarque : La variance étant positive, l'écart-type est bien défini.

Propriété : $\sigma(aX + b) = |a| \sigma(X)$.

- On appelle **Covariance** entre deux variables aléatoires X et Y et on note $\text{Cov}(X, Y)$ le nombre défini par

$$\text{Cov}(X, Y) = \mathbb{E}[(XY - \mathbb{E}[X]\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Remarque : La covariance traduit la corrélation entre X et Y . Si les variables sont décorré-
lées, on aura $\text{Cov}(X, Y) = 0$.

Propriété : $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$.

E. Rappels d'algèbre

On dit que

- a est **racine d'ordre p** du polynôme $P \Leftrightarrow$ le polynôme P est divisible par $(x - a)^p$
- $\Leftrightarrow \exists Q$ polynôme tel que $P(x) = (x - a)^p Q(x)$
et $(x - a)$ ne divise Q
- \Leftrightarrow le reste de la division euclidienne de P par $(x - a)^p$ est nul

$$\Leftrightarrow \begin{cases} P(a) & = & 0 \\ P'(a) & = & 0 \\ P''(a) & = & 0 \\ & \vdots & \\ P^{(p-1)}(a) & = & 0 \\ P^{(p)}(a) & \neq & 0 \end{cases}$$