Statistiques Flux Portuaires

Master 1 Management Portuaire et Maritime

Université du Littoral - Côte d'Opale, Pôle Lamartine

Laurent SMOCH

Septembre 2013

Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville
Université du Littoral, zone universitaire de la Mi-Voix, bâtiment H. Poincaré
50, rue F. Buisson, BP 699, F-62228 Calais cedex

Table des matières

Chapitre 3

Séries chronologiques

3.1 Introduction

L'étude des **séries chronologiques** (ou **séries temporelles** ou **chroniques** - time series- en anglais) permet de décrire, expliquer, contrôler, prévoir des phénomènes évoluant au cours du temps.

Définition 3.1.1 *On appelle série chronologique la succession de valeurs que prend une variable (ou caractéristique) au cours du temps pour N périodes successives pour un même individu (ou cas).*

Donc, une série chronologique est une suite d'observations d'une variable statistique au cours du temps.

Exemple 3.1.1

- la production mensuelle d'une entreprise,
- le prix du pain au 1^{er} de chaque mois,
- le revenu national de 1953 à aujourd'hui.

Une série chronologique peut être identifiée à une série statistique à deux variables t et X_t, où t représente le temps. On la note (t, X_t).

Remarque 3.1.1 La variable t sert à ordonner X_t, ses valeurs sont séparées par une constante c'est-à-dire l'unité de temps (année, mois, semaine) qui caractérise la série chronlogique et en donne la périodicité. On peut ainsi parler de séries chronologiques annuelles, mensuelles ou trimestrielles.

Les observations sont supposées faites à intervalles de temps réguliers. Dans la pratique, ce n'est pas toujours réalisé. Par exemple, pour des séries mensuelles, le nombre de jours varie d'un mois à l'autre, on utilise alors des données corrigées.

3.2 Modèles de composition

Pour étudier (t, X_t), on décompose la série chronologique en différents mouvements.

3.2.1 Les composantes fondamentales

- La **tendance à long terme** ou "**trend**" : T_t
  Elle représente l'évolution à long terme de la série et traduit son aspect général (croissance, décroissance,...).

- La **composante cyclique** : C_t
  Pour des séries très longues, elle correspond à des variations connues régulières. Par exemple, en économie, des fluctuations correspondant à des périodes de prospérité ou de récession.
  En général, la composante cyclique n'est pas détectable sur la période étudiée et on suppose alors que C_t n'existe pas.

– La **composante saisonnière** : S_t
Elle correspond à des variations régulières inconnues mais détectables sur des périodes courtes, généralement à l'intérieur d'une année : semaine, mois, trimestre,.... Par exemple, l'influence des congés annuels sur la production d'une entreprise.
– La **composante accidentelle** : A_t
Elle correspond à des variations accidentelles de forte amplitude dues à des phénomènes imprévisibles. Par exemple, la grève, le risque de guerre.

3.2.2 Les différents modèles de composition

Il y a plusieurs manières de composer les mouvements ci-dessus.

– Le **modèle additif** :

$$\boxed{X_t = T_t + C_t + S_t + A_t}$$

si C_t, S_t et A_t sont indépendants de T_t.

– Le **modèle multiplicatif** :

$$\boxed{X_t = T_t \times C_t \times S_t \times A_t}$$

si C_t, S_t et A_t sont dépendants de T_t.

– Le **modèle mixte** : les modèles concernés composent les mouvements avec produits et sommes.

Le choix du modèle est souvent délicat. En général, le modèle le mieux adapté est le modèle multiplicatif.

3.3 Analyse des différentes composantes

Exemple 3.3.1 On se donne les livraisons trimestrielles d'essence en super sans plomb 98 (en millions de m^3) dans un hypermarché de la région pendant quatre années :

Trimestres / Années	1	2	3	4
1997	1050	1300	1500	1300
1998	1050	1400	1750	1350
1999	1100	1550	1850	1450
2000	1150	1700	2000	1550

TABLE 3.1 – Livraisons trimestrielles de SSP98

3.3.1 Analyse de la tendance à long terme

Cette analyse peut être réalisée uniquement si la série est assez longue.

(a) Méthode graphique

On représente tout d'abord sur un graphique les livraisons trimestrielles (sur l'axe des ordonnées) en fonction des différents trimestres (sur l'axe des abscisses). On obtient la figure 3.1.

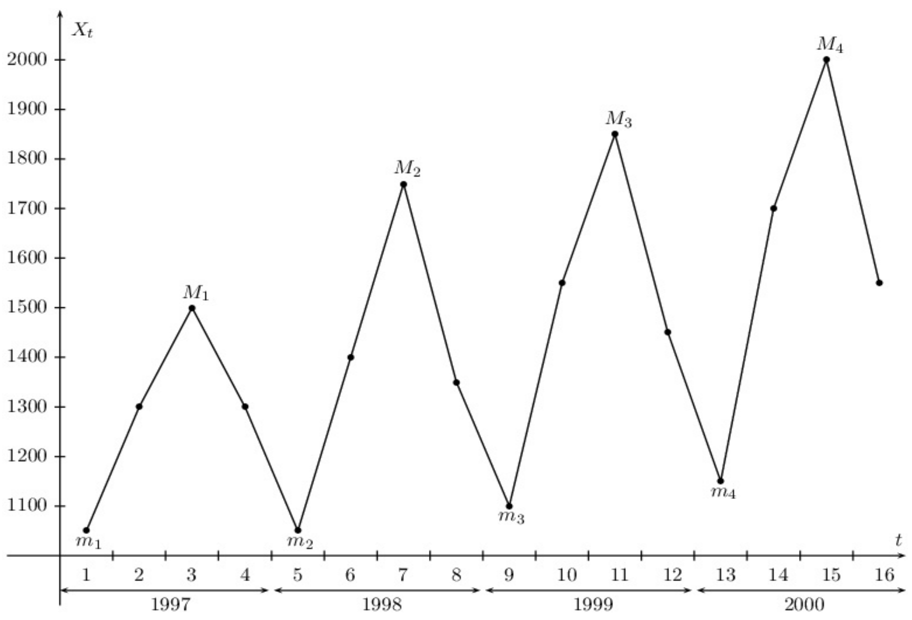

FIGURE 3.1 – Représentation graphique

L'objectif de l'analyse graphique est d'exhiber du graphe un mouvement, une tendance générale afin de préciser la manière dont les livraisons progressent au cours du temps. La période d'observation est ici de quatre années.
La procédure est ensuite la suivante :

- on joint les maxima M_i à la règle,
- on joint les minima m_i à la règle,
- on définit les projections orthogonales M_i' des maxima M_i sur les segments $[m_i, m_{i+1}]$,
- on définit les projections orthogonales m_i' des minima m_i sur les segments $[M_i, M_{i+1}]$,
- on trace les segments verticaux $[M_i, M_i']$ et $[m_i, m_i']$.

La ligne brisée qui joint les milieux des segments $[M_i, M_i']$ et $[m_i, m_i']$ constitue une tendance (voir figure 3.2.).

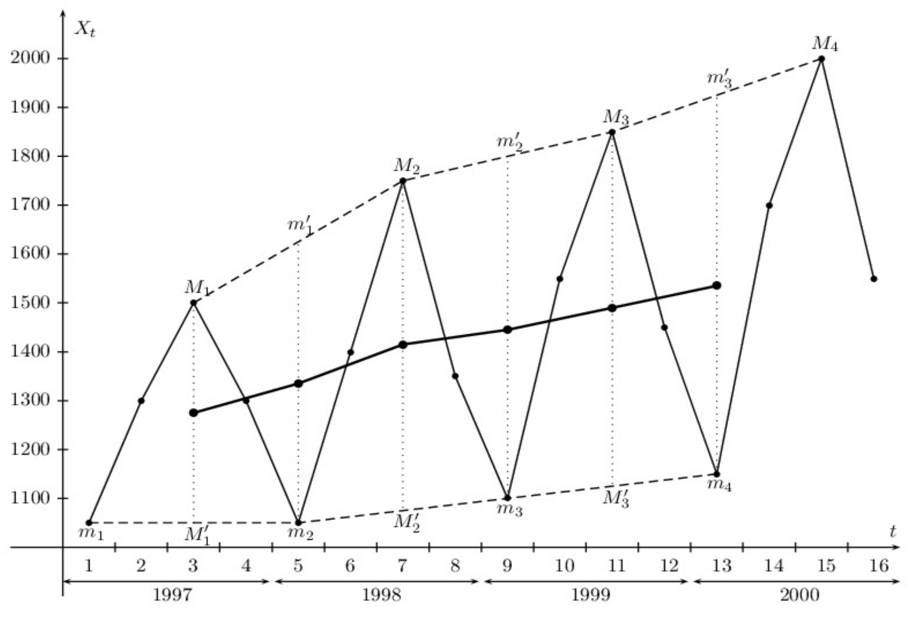

FIGURE 3.2 – Tendance - Méthode graphique

(b) Ajustement linéaire par la méthode des moindes carrés

Si le coefficient de corrélation de $(t, X(t)))$ est assez grand en valeur absolue (on suppose qu'en pratique, r doit être plus proche de 1 que de 0), la droite de régression de X_t en fonction de t donne la tendance.

Recherchons l'équation de régression de X_t dans le cadre de l'exemple 3.3.1. On utilisera donc le tableau de valeurs de la page suivante où t_i représente le rang d'un trimestre et $X_t(i)$ la livraison d'essence correspondante : On récupère les valeurs suivantes :

Rang i	Trimestre t_i	Livraisons $X(t)_i$	t_i^2	$X(t)_i^2$	$t_i.X(t)_i$
1	1	1050	1	1102500	1050
2	2	1300	4	1690000	2600
3	3	1500	9	2250000	4500
4	4	1300	16	1690000	5200
5	5	1050	25	1102500	5250
6	6	1400	36	1960000	8400
7	7	1750	49	3062500	12250
8	8	1350	64	1822500	10800
9	9	1100	81	1210000	9900
10	10	1550	100	2402500	15500
11	11	1850	121	3422500	20350
12	12	1450	144	2102500	17400
13	13	1150	169	1322500	14950
14	14	1700	196	2890000	23800
15	15	2000	225	4000000	30000
16	16	1550	256	2402500	24800
Total	136	23050	1496	34432500	206750

TABLE 3.2 – Tableau de calculs (trimestres) - Exemple 3.3.1

- $\bar{t} = \dfrac{136}{16} = 8,5$
- $\overline{X}(t) = \dfrac{23050}{16} = 1440,625$
- $V(t) = \dfrac{1496}{16} - (8,5)^2 = 21,25$
- $V(X(t)) = \dfrac{34432500}{16} - (1440,625)^2 = 76630,859$
- $\mathrm{Cov}(t, X(t)) = \dfrac{206750}{16} - (8,5 \times 1440,625) = 676,5625.$

Le coefficient directeur et l'ordonnée à l'origine de la droite de régression $D_{X(t)/t}$ sont respectivement donnés par :

$$a = \frac{\mathrm{Cov}(t, X(t))}{V(t)} \text{ et } b = \overline{X}(t) - a\bar{t}$$

soit

$$a = \frac{676,5625}{21,25} = 31,838 \text{ et } b = 1440,625 - 31,838 \times 8,5 = 1170,002.$$

Finalement, la droite admet pour équation

$$D_{X(t)/t} : X(t) = 31,838t + 1170,002$$

où t représente le trimestre, qu'on renomme D_1. On peut calculer accessoirement le coefficient de corrélation linéaire entre t et $X(t)$:

$$r = \frac{\text{Cov}(t, X(t))}{\sqrt{V(t)}\sqrt{V(X(t))}} = \frac{676,5625}{\sqrt{21,25}\sqrt{76630,859}} = 0,53.$$

Remarque 3.3.1 On peut également chercher par la méthode des moindres carrés l'équation d'une droite qui nous permettra d'évaluer, en fonction cette fois-ci du rang d'une année, les quantités livrées trimestriellement en moyenne sur une année. On utilisera dans ce cas le tableau ci-dessous où x représente le rang d'une année et m_x la moyenne trimestrielle correspondant à cette année :

Rang	x	m_x	x^2	m_x^2	xm_x
1	1	$1287,5$	1	$1657656,25$	$1287,5$
2	2	$1387,5$	4	$1925156,25$	$2775,0$
3	3	$1487,5$	9	$2212656,25$	$4462,5$
4	4	$1600,0$	16	2560000	$6400,0$
Total	10	$5762,5$	30	$8355468,75$	$14925,0$

TABLE 3.3 – Tableau de calculs (années) - Exemple 3.3.1

$$\left(\text{Par exemple, } m_2 = 1387,5 = \frac{(1050 + 1400 + 1750 + 1350)}{4}\right).$$

On trouve les valeurs

- $\overline{x} = \dfrac{10}{4} = 2,5$
- $\overline{m}_x = \dfrac{5762,5}{4} = 1440,625$
- $V(x) = \dfrac{30}{4} - (2,5)^2 = 1,25$
- $V(m_x) = \dfrac{8355468,75}{4} - (1440,625)^2 = 13466,797$
- $\text{Cov}(x, m_x) = \dfrac{14925}{4} - (2,5 \times 1440,625) = 129,687.$

Le coefficient directeur et l'ordonnée à l'origine de la droite de régression $D_{m_x/x}$ sont respectivement donnés par :

$$a = \frac{\text{Cov}(x, m_x)}{V(x)} \text{ et } b = \overline{m}_x - a\overline{x}$$

soit

$$\begin{cases} a = \dfrac{129,687}{1,25} = 103,75 \\ b = 1440,625 - 103,75 \times 2,5 = 1181,25 \end{cases}$$

On trouve ainsi que

$$D_{m_x/x} : m_x(x) = 103,75x + 1181,25$$

où X représente l'année, qu'on renomme D_2.

On trace ces deux droites de régression sur le graphe :

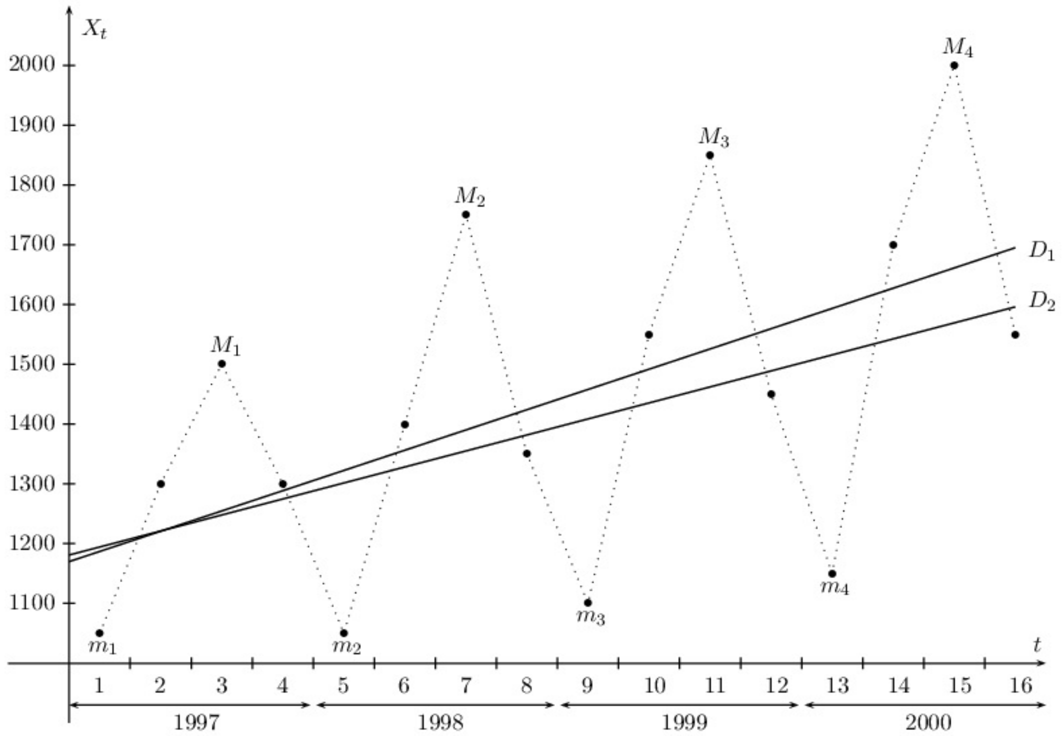

FIGURE 3.3 – Tendance - Ajustement linéaire

(c) Méthode des moyennes échelonnées

Soit $k \geq 2$ un entier. On partage l'ensemble des observations en sous-ensembles comprenant k observations successives. Pour chaque sous-ensemble, on calcule son point moyen. La tendance est alors la ligne brisée qui passe par ces points.

Remarque 3.3.2 Les inconvénients de cette méthode sont :
- le choix de k,
- la réduction du nombre de points du graphique.

Retour à l'exemple 3.3.1. Si on choisit $k = 8$, on a le tableau suivant :

t	Y_t
Sous-ensemble 1	$\dfrac{1050 + 1300 + 1500 + 1300 + 1050 + 1550 + 1850 + 1450}{8} = 1337,5$
Sous-ensemble 2	$\dfrac{1100 + 1550 + 1850 + 1450 + 1150 + 1700 + 2000 + 1550}{8} = 1543,75$

TABLE 3.4 – Moyennes échelonnées, $k = 8$

Si on choisit $k = 4$, on obtient les données :

t	Y_t
Sous-ensemble 1	$\frac{1050+1300+1500+1300}{4} = 1287,5$
Sous-ensemble 2	$\frac{1050+1400+1750+1350}{4} = 1387,5$
Sous-ensemble 3	$\frac{1100+1550+1850+1450}{4} = 1487,5$
Sous-ensemble 4	$\frac{1150+1700+2000+1550}{4} = 1600$

TABLE 3.5 – Moyennes échelonnées, $k = 4$

ainsi que le graphique associé :

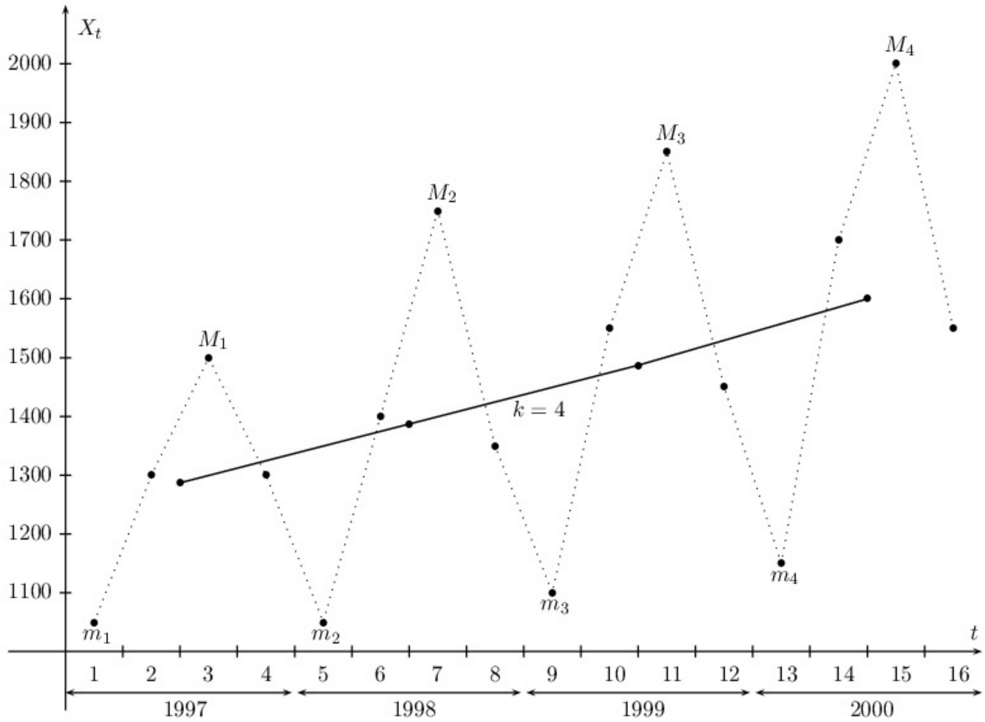

FIGURE 3.4 – Tendance - Moyennes échelonnées, $k = 4$

(d) Méthode des moyennes mobiles

Afin d'éliminer ou d'amortir les mouvements cycliques, saisonniers et accidentels, on peut également utiliser la technique des moyennes mobiles. Le principe de cette méthode est similaire à celui de la méthode des moyennes échelonnées. Il consiste à construire une nouvelle série obtenue en calculant des moyennes arithmétiques successives de longueur $k \geq 2$ fixe à partir des données originales. Chacune de ces moyennes obtenues correspondra au "milieu" de la période pour laquelle la moyenne arithmétique vient d'être calculée.

- Le premier est $(1, X_1)$, $(2, X_2)$, $\ldots$, (k, X_k), représenté par $\left(1, \frac{1}{k}(X_1 + X_2 + \ldots + X_k)\right)$,
- le deuxième est $(2, X_2)$, $(3, X_3)$, $\ldots$, $(k+1, X_{k+1})$, représenté par $\left(2, \frac{1}{k}(X_2 + X_3 + \ldots + X_{k+1})\right)$,
- et ainsi de suite...

On remarque dans ce cas que les $k-1$ dernières moyennes sont incalculables ($X_{N+1}, \ldots, X_{N+k-1}$ n'existent pas).

On peut perfectionner la méthode en considérant des moyennes mobiles centrées. On appelle **moyennes mobiles centrées** d'ordre k de la série des $\{X_t, t = 1, \ldots, N\}$ les moyennes arithmétiques calculées sur k valeurs successives et rapportées à la date du milieu de la période.

- Si k est impair ($k = 2p + 1$), on considérera les points de coordonnées

$$\boxed{\left(j, \frac{1}{2p+1}(X_{j-p} + \ldots + X_j + \ldots + X_{j+p})\right)}$$

  Il est évident que des problèmes se posent aux extrémités de la série. Par exemple, si $k = 3 = 2 \times 1 + 1$ (donc $p = 1$), les moyennes mobiles de la première ($j = 1$) et de la dernière période ($j = N$) ne peuvent pas être calculées (X_0 et X_{N+1} n'existent pas). Si $k = 5$, ce sont les moyennes des deux premières et des deux dernières périodes qui sont incalculables.
- En choisissant k pair ($k = 2p$), nous sommes confrontés au problème que les moyennes obtenues ne correspondront pas à une abscisse existante, mais chevaucheront entre deux de ces valeurs. Afin de contourner ce problème, plutôt que de considérer le point moyen de $(j-p, X_{j-p}), \ldots, (j+p-1, X_{j+p-1})$ ou $(j-p+1, X_{j-p+1}), \ldots, (j+p, X_{j+p})$, on considérera les points de coordonnées

$$\boxed{\left(j, \frac{1}{2p}(\frac{1}{2}X_{j-p} + X_{j-p+1} + \ldots + X_{j+p-1} + \frac{1}{2}X_{j+p})\right)}$$

  Les moyennes incalculables sont les mêmes que dans le cas impair.

Remarque 3.3.3

- Concernant le choix de k, pour des données trimestrielles, on prendra $k = 4$. Pour des données mensuelles, on prendra $k = 12$,...
- Cette méthode élimine les influences saisonnière et accidentelle, c'est une méthode de lissage.

Retour à l'exemple 3.3.1. Choisissons $k = 4$:

- $1^{\text{ère}}$ méthode (sans considérer le caractère pair de k) :

t	1	2	3	4	5	6	7	8
M'_t	1287,5	1287,5	1312,5	1375	1387,5	1400	1437,5	1462,5
t	9	10	11	12	13	14	15	16
M'_t	1487,5	1500	1537,5	1575	1600	−−	−−	−−

TABLE 3.6 – Moyennes mobiles centrées, $k = 4$

- $2^{\text{ème}}$ méthode (en considérant le caractère pair de k) :

t	1	2	3	4	5	6	7	8
M'_t	−−	−−	1287,5	1300	1343,75	1381,25	1393,75	1418,75
t	9	10	11	12	13	14	15	16
M'_t	1450	1475	1493,75	1518,75	1556,25	16	−−	−−

TABLE 3.7 – Moyennes mobiles centrées, $k = 4 = 2 \times 2$

On représente graphiquement la tendance à l'aide de la deuxième méthode (cf. figure 3.5).

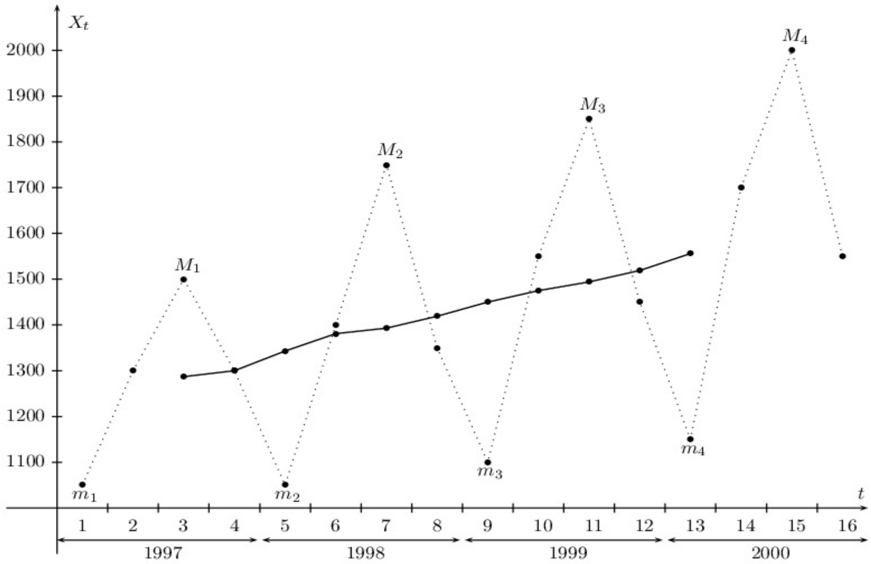

FIGURE 3.5 – Tendance - Moyennes mobiles centrées, $k = 4 = 2 \times 2$

3.3.2 Analyse de la composante saisonnière

L'analyse de la composante saisonnière a deux objectifs :

- étudier les mouvements saisonniers,
- les éliminer afin de faire apparaître les composantes cyclique et accidentelle.

Pour effectuer l'analyse des mouvements saisonniers, on essaie d'abord de déterminer si on est en présence d'une série dans laquelle pour une observation donnée :

- la variation saisonnière S_t s'ajoute simplement à la résultante des autres composantes, c'est le modèle additif ;
- la variation saisonnière S_t est proportionnelle à la résultante des autres composantes, et alors c'est le modèle multiplicatif.

(a) Modèle additif ou multiplicatif ?

- Méthode graphique de superposition.
  Pour faire cette détermination (modèle additif, multiplicatif) graphiquement, on peut par exemple superposer les saisons représentées par des droites de profil sur un même graphique. Cela donne dans le cas de l'exemple 3.3.1 la figure suivante :

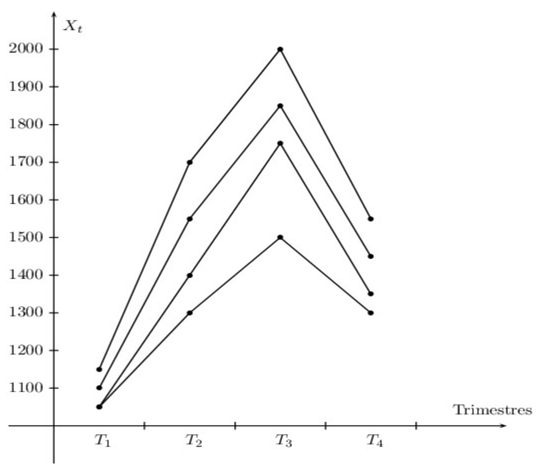

FIGURE 3.6 – Détermination du modèle - Méthode graphique de superposition

De manière générale, si les droites du graphique superposé sont parallèles, le modèle est additif, autrement le modèle est multiplicatif. Sur le graphique de notre exemple, les droites de profil ne sont pas parallèles pour toutes les saisons, le modèle est multiplicatif.

- Méthode graphique de la bande.
  La nature du modèle peut être confirmée à l'aide d'une autre méthode graphique : la méthode de la bande. On fait un graphique représentant la série chronologique, puis on trace une droite passant respectivement par les minima et par les maxima de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, c'est un modèle multiplicatif. Illustrons ceci à l'aide de l'exemple 3.3.1 de référence. On a la figure 3.7, nous constatons que les deux droites ne sont pas parallèles, nous sommes donc en présence d'un modèle multiplicatif.

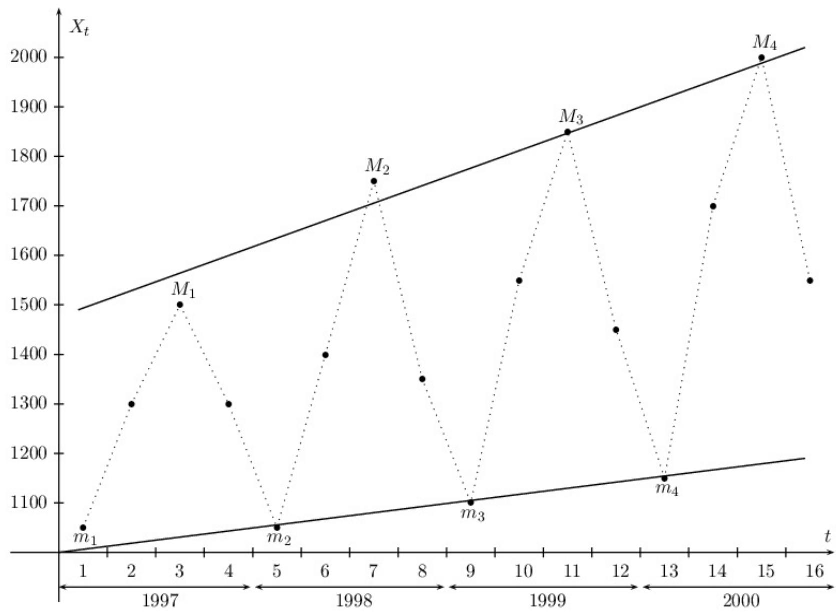

FIGURE 3.7 – Détermination du modèle - Méthode de la bande

- Méthode analytique.
  Présentons maintenant la méthode analytique. On calcule les moyennes et écarts-types pour chacune des périodes considérées et on calcule la droite des moindres carrés $\sigma = a\overline{x} + b$. Si la pente a est nulle, c'est un modèle additif, sinon, le modèle est multiplicatif. Dans le cadre de l'exemple 3.3.1, on obtient (en utilisant partiellement la table 3.3) le tableau suivant :

Rang	Moyenne $\overline{x}$	Écart-type σ	$\overline{x}^2$	σ^2	$\overline{\sigma}$
1	1287,5	159,59	1657656,25	25468,97	205472,13
2	1387,5	248,43	1925156,25	61717,46	344696,63
3	1487,5	267,80	2212656,25	71716,84	398352,5
4	1600	306,19	2560000	93752,32	489904
Total	5762,5	982,01	8355468,8	252655,59	1438425,26

TABLE 3.8 – Détermination du modèle - Méthode analytique

En déterminant l'équation de la droite des moindres carrés associée, on obtient $a = 0,44$ ($\text{Cov}(\overline{x}, \sigma) = 5932,88$ et $V(\overline{x}) = 13466,81$) et $b = -388,38$ ($\overline{\sigma} = 245,50$ et $\overline{\overline{x}} = 1440,625$), ce qui confirme encore une fois que pour cet exemple, nous sommes bien en présence d'un modèle multiplicatif.

On peut maintenant effectuer l'analyse des mouvements saisonniers. On présente deux méthodes.

(b) Analyse des mouvements saisonniers - La méthode graphique

La visualisation de la composante saisonnière peut se faire naturellement à l'aide du graphique dit "superposé". Dans le cas de l'exemple 3.3.1 de référence, on constate à l'aide de la figure 3.7 qu'il y a périodicité annuelle (les courbes sont sensiblement de même forme) même si certains phénomènes sous-jacents semblent intervenir.

(c) Analyse des mouvements saisonniers - La méthode des moyennes saisonnières

Soit j l'indice relatif à la saison.
- Si on dispose de données trimestrielles, $j = 1, \ldots, 4$
- Si on dispose de données mensuelles, $j = 1, \ldots, 12$

On calcule la moyenne des données pour chaque saison $\overline{x}_j$ puis $\overline{x}_j - \overline{x}$ ce qui permet de situer la valeur saisonnière moyenne par rapport à la valeur globale moyenne.
On place sur un graphique les points $(j, \overline{x}_j - \overline{x})$.

On reprend l'exemple de référence 3.3.1. On a la moyenne globale des livraisons qui vaut $\overline{x} = 1440,625$ ainsi que les moyennes saisonnières
- $\overline{x}_1 = 1087,5$ avec $\overline{x}_1 - \overline{x} = -353,125$,
- $\overline{x}_2 = 1487,5$ avec $\overline{x}_2 - \overline{x} = 46,875$,
- $\overline{x}_3 = 1775$ avec $\overline{x}_3 - \overline{x} = 334,375$,
- $\overline{x}_4 = 1412,5$ avec $\overline{x}_4 - \overline{x} = -28,125$

On a donc le graphe suivant

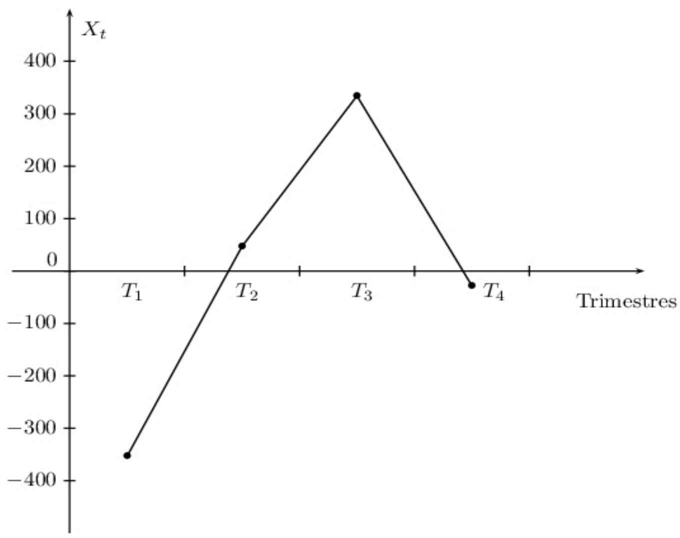

FIGURE 3.8 – Graphique des moyennes saisonnières

On peut ainsi remarquer par exemple que les livraisons d'essence moyennes du premier trimestre sont inférieures aux ventes globales moyennes d'environ 353 millions de m^3.

3.3.3 Correction des variations saisonnières

Dans beaucoup de situations, il est préférable de travailler sur des données qui ne sont pas affectées par un mouvement saisonnier. C'est pour cela que l'on transforme la série chronologique initiale en données désaisonnalisées ou corrigées des variations saisonnières.

(a) Le modèle multiplicatif

On a $X_t = T_t \times C_t \times S_t \times A_t$.

- On détermine la tendance T_t par la droite de régression ou par la méthode des moyennes mobiles,
- pour chaque t, on calcule le rapport $\dfrac{X_t}{T_t} = r_t$ appelé **rapport à la tendance**,
- on calcule les moyennes des rapports r_t correspondant à une même saison et on obtient ainsi les **coefficients saisonniers** s_j où j est l'indice relatif à la saison. Si les rapports à la tendance ne sont pas égaux, cela est dû à une influence accidentelle,
- on calcule la moyenne $\overline{s}$ des coefficients saisonniers. Si le modèle est bien choisi, la moyenne $\overline{s}$ doit être proche de 1. Dans le cas contraire,
- on calcule les corrections
$$s_j^\star = \frac{s_j}{\overline{s}}$$
qui vérifient dorénavant la propriété $\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} s_j^\star = 1$ où n est le nombre de coefficients saisonniers.
- On calcule enfin la série désaisonnalisée ou série corrigée des variations saisonnières
$$\boxed{\left(t, X_t^\star = \frac{X_t}{s_t^\star}\right)}$$
où $s_t^\star = s_j^\star$ si t appartient à la saison j. Par ce biais, on estompe l'influence accidentelle de la série.

On reprend l'exemple 3.3.1 de référence. On sait que le modèle est multiplicatif. La tendance à la saison t (soit T_t) peut être donnée à l'aide de la droite de régression dont on rappelle l'équation :

$$X(t) = 31,838t + 1170,002$$

On est donc en mesure de calculer les rapports à la tendance :

t	1	2	3	4	5	6	7	8
X_t	1050	1300	1500	1300	1050	1400	1750	1350
T_t	1201,84	1233,68	1265,52	1297,35	1329,19	1361,03	1392,87	1424,71
$r_t = X_t/T_t$	0,87	1,05	1,19	1,00	0,79	1,03	1,26	0,95
t	9	10	11	12	13	14	15	16
X_t	1100	1550	1850	1450	1150	1700	2000	1550
T_t	1456,54	1488,38	1520,22	1552,06	1583,90	1615,73	1647,57	1679,41
$r_t = X_t/T_t$	0,76	1,04	1,22	0,93	0,73	1,05	1,21	0,92

TABLE 3.9 – Rapports à la tendance

On calcule ensuite les coefficients saisonniers

- $s_1 = \dfrac{1}{4}(r_1 + r_5 + r_9 + r_{13}) = \dfrac{1}{4}(0,87 + 0,79 + 0,76 + 0,73) = 0,79$
- $s_2 = \dfrac{1}{4}(r_2 + r_6 + r_{10} + r_{14}) = \dfrac{1}{4}(1,05 + 1,03 + 1,04 + 1,05) = 1,04$
- $s_3 = \dfrac{1}{4}(r_3 + r_7 + r_{11} + r_{15}) = \dfrac{1}{4}(1,19 + 1,26 + 1,22 + 1,21) = 1,22$
- $s_4 = \dfrac{1}{4}(r_4 + r_8 + r_{12} + r_{16}) = \dfrac{1}{4}(1,00 + 0,95 + 0,93 + 0,92) = 0,95$

Comme $\overline{s} = \dfrac{1}{4}(s_1 + s_2 + s_3 + s_4) = \dfrac{1}{4}(0,79 + 1,04 + 1,22 + 0,95) = 1$, le modèle est bien choisi.
On se sert ensuite des coefficients saisonniers pour désaisonnaliser la série. On obtient le tableau suivant :

t	1	2	3	4	5	6	7	8
X_t	1050	1300	1500	1300	1050	1400	1750	1350
$X_t^\star$	1329, 11	1250	1229, 51	1368, 42	1329, 11	1346, 15	1434, 43	1421, 05
t	9	10	11	12	13	14	15	16
X_t	1100	1550	1850	1450	1150	1700	2000	1550
$X_t^\star$	1392, 41	1490, 38	1516, 39	1526, 32	1455, 70	1634, 62	1639, 34	1631, 58

TABLE 3.10 – Série désaisonnalisée

Voici pour terminer les représentations graphiques de la série originale, de la droite de régression et de la série désaisonnalisée.

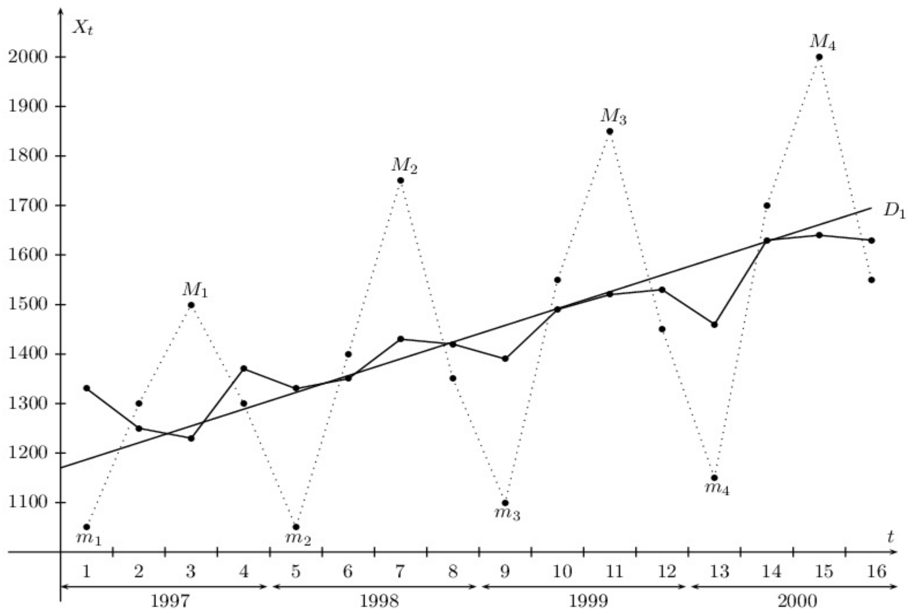

FIGURE 3.9 – Représentations graphiques de la série originale, de la droite de régression et de la série désaisonnalisée

La série désaisonnalisée fait apparaître par conséquent les composantes cycliques et accidentelles plus explicitement.

(b) Le modèle additif

On a $X_t = T_t + C_t + S_t + A_t$. La méthode est analogue à celle du modèle multiplicatif :

- on calcule la différence $d_t = X_t - T_t$ au lieu du rapport,
- on obtient les coefficients s_j en calculant les moyennes des différences d_t correspondant à une même saison,
- on calcule ensuite la moyenne $\overline{s}$ de ces coefficients s_j. Si le modèle est bien choisi, la moyenne $\overline{s}$ doit être proche de 0. Dans le cas contraire,
- on calcule les corrections

$$s_j^\star = s_j - \overline{s},$$

– on calcule enfin la série désaisonnalisée

$$\boxed{(t, X_t^\star = X_t - s_t^\star)}.$$

Exemple 3.3.2 *(cas additif)* Considérons la série statistique des chiffres d'affaires trimestriels (en milliers d'euros) d'une entreprise.

Trimestre	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Chiffre d'affaires	120	181	71	119	128	190	73	124	140	196	84	133	145	206	96	142

Table 3.11 – Chiffres d'affaire trimestriels

Le modèle de la série statistique est bien additif, pour s'en convaincre, il suffit d'employer la méthode de la bande et de constater que la droite qui passe par les maxima et celle qui passe par les minima sont parallèles.

Calculons les moyennes mobiles T_t d'ordre 4 (puisque nous sommes en présence d'une série dont la périodicité est de 4 trimestres). Ensuite nous déterminons les différences saisonnières d_t, c'est-à-dire les différences entre les valeurs de la série statistique de départ et les valeurs des moyennes mobiles correspondantes. Ces différences vont nous servir pour obtenir les coefficients saisonniers non corrigés trimestriels.

t	1	2	3	4	5	6	7	8
X_t	120	181	71	119	128	190	73	124
T_t	–	–	123,75	125,88	127,25	128,13	130,25	132,50
d_t	–	–	-52,75	-6,88	0,75	61,88	-57,25	-8,50
t	9	10	11	12	13	14	15	16
X_t	140	196	84	133	145	206	96	142
T_t	134,63	137,13	138,88	140,75	143,50	146,13	–	–
d_t	5,38	58,88	-54,88	-7,75	1,50	59,88	–	–

Table 3.12 – Différences à la tendance

Les coefficients saisonniers correspondent aux moyennes des différences saisonnières pour chacun des trimestres. On obtient :

- $s_1 = \frac{1}{3}(d_5 + d_9 + d_{13}) = \frac{1}{4}(0,75 + 5,38 + 1,50) = 2,54$
- $s_2 = \frac{1}{3}(d_6 + d_{10} + d_{14}) = \frac{1}{3}(61,88 + 58,88 + 59,88) = 60,21$
- $s_3 = \frac{1}{3}(d_3 + d_7 + d_{11}) = \frac{1}{3}(-52,75 - 57,25 - 54,88) = -54,96$
- $s_4 = \frac{1}{3}(d_4 + d_8 + d_{12}) = \frac{1}{3}(-6,88 - 8,50 - 7,75) = -7,71$

Nous supposons que la composante saisonnière est strictement périodique. L'effet net de la composante saisonnière sur une période doit être nul car il est repris dans la tendance générale de la série chronologique. Or ce n'est pas le cas dans cet exemple puisque

$$\overline{s} = \frac{1}{4}(s_1 + s_2 + s_3 + s_4) = \frac{1}{4}(2,54 + 60,21 - 54,96 - 7,71) = 0,02.$$

Ceci nous amène donc à rectifier les coefficients saisonniers non corrigés en leur retranchant la moyenne des coefficients saisonniers pour toutes les périodes. On a alors :

- $s_1^\star = s_1 - \overline{s} = 2,54 - 0,02 = 2,52$
- $s_2^\star = s_2 - \overline{s} = 60,21 - 0,02 = 60,19$
- $s_3^\star = s_3 - \overline{s} = -54,96 - 0,02 = 54,98$
- $s_4^\star = s_4 - \overline{s} = -7,71 - 0,02 = -7,73$

Disposant maintenant des coefficients saisonniers corrigés, nous pouvons désaisonnaliser la série chronologique en retranchant à chacune des valeurs initiales de la série la valeur du coefficient saisonnier correspondant.

t	1	2	3	4	5	6	7	8
X_t	120	181	71	119	128	190	73	124
d_t	117,48	120,81	125,98	126,73	125,48	129,81	127,98	131,73
t	9	10	11	12	13	14	15	16
X_t	140	196	84	133	145	206	96	142
$X_t^\star$	137,48	135,81	138,98	140,73	142,48	145,81	150,98	149,73

Table 3.13 – Série désaisonnalisée

3.4 Exercices

Exercice 33 Dans une grande concession, on dispose d'une statistique portant sur les ventes mensuelles de voitures au cours de vingt quatre mois. Le tableau suivant contient les valeurs brutes (c'est-à-dire non corrigées des variations saisonnières).

Ventes brutes x_t	janv	févr	mars	avril	mai	juin	juill	août	sept	oct	nov	déc
2002	410	435	450	430	440	455	450	430	460	470	460	440
2003	465	480	465	475	480	430	460	460	490	470	480	500

1. Calculer la série des moyennes mobiles sur quatre périodes.
2. Calculer les coefficients a et b de la droite de régression.
3. Représenter sur un même graphique la série brute (t, y_t), la série des moyennes mobiles sur quatre périodes, la droite de régression.
4. Calculer le coefficient de détermination r qui est défini par

$$r^2 = \frac{\mathrm{Cov}(x,t)^2}{V(x)V(t)}$$

La droite de régression fournit-elle une bonne estimation de la tendance d'ensemble des ventes ?

Exercice 34 La série suivante retrace les fluctuations de l'indice trimestriel pour la production industrielle au cours de six années. Le tableau suivant contient les valeurs brutes (c'est-à-dire non corrigées des variations saisonnières).

y_{ik}	printemps	été	automne	hiver
2005			105	130
2006	120	125	103	125
2007	122	125	102	126
2008	121	127	104	128
2009	124	128	109	125
2010	125	132		

1. Montrer le caractère saisonnier de cette chronique.
2. Calculer le trend par la méthode des moyennes mobiles.
3. Calculer les coefficients saisonniers.
4. Calculer la valeur désaisonnalisée du quatrième trimestre de l'année 2010 si la valeur brute est 132.
5. Calculer la valeur corrigée des variations saisonnières.

Exercice 35 Dans un magasin d'habillement, on a enregistré à la fin de chaque trimestre et pendant trois années le bilan des ventes d'un article d'habillement en milliers d'unités. Les résultats sont donnés dans le tableau suivant qui en fournit les valeurs brutes (c'est-à-dire non corrigées des variations saisonnières).

2001-2003	1	2	3	4	5	6	7	8	9	10	11	12
Ventes brutes x_t	4,16	5,27	1,70	2,94	7,04	8,33	2,80	4,27	10,56	11,73	3,70	5,53

1. Construire la représentation graphique de la série brute (t, x_t)
2. On se propose d'étudier maintenant comme modèle de série chronologique approprié un modèle multiplicatif de la forme $x_t = C_t S_t \varepsilon_t$. Donner le tableau des valeurs brutes (c'est-à-dire non corrigées des variations saisonnières) du logarithme $y_t = \ln x_t = \ln C_t + \ln S_t + \ln \varepsilon_t$ de x_t. Ici $\ln C_t$ sera la tendance linéaire, $\ln S_t$ sera la série des moyennes mobiles sur quatre périodes de y_t et enfin $\ln \varepsilon_t$ une suite chronologique de moyenne nulle.
3. Calculer les coefficients a et b de la droite de régression de y_t en fonction de t. En déduire l'expression de C_t.
4. Calculer la série $\ln S_t$ des moyennes mobiles sur quatre périodes de y_t.