

# Cours 1 – Analyse descriptive des Séries Chronologiques

A. Godichon-Baggioni

INSA – GM4 – Cours de Statistiques

2017–2018

# Bibliographie

- 1 *Méthodes de prévision à court terme*, Guy Mélard, 1991, Ellipses
- 2 *Séries temporelles et modèles dynamiques*, Christian Gourieroux, Alain Monfort, 1995, Economica

# 1. Introduction

## 1.1 Définition et objectifs

### Définition (*Série chronologique*)

On appelle **série chronologique**, ou bien encore **chronique** ou **série temporelle**, une suite finie de données quantitatives indexée par le temps. L'indice temps peut être selon les cas, la seconde, la minute, l'heure, le jour, le mois, l'année, ...

**Nombreux domaines d'application** : on rencontre des séries chronologiques en

- économie, finance ;
- démographie ;
- biologie, médecine ;
- météorologie, pollution ;
- ...

**Objet du cours** : mise en place de techniques mathématiques pour l'étude des séries chronologiques, dans le but de :

- comprendre le passé : analyser et expliquer les valeurs observées ;
- prédire le futur : bâtir des prévisions pour les valeurs non encore observées ;
- étudier le lien avec d'autres séries chronologiques.

## Notation

Considérons une série chronologique de longueur  $n$ .

Si

- $t_1, t_2, \dots, t_n$  sont les  $n$  instants successifs d'observation
- et  $y_{t_j}$  est la valeur mesurée à l'instant  $t_j$ ,

on notera la série chronologique  $\{y_t\}_{t \in T}$  où  $T$  est l'ensemble ordonné des instants d'observation, ie.  $T = \{t_1, t_2, \dots, t_n\}$ .

## Remarque

La série chronologique  $\{y_t\}_{t \in T}$  avec  $T = \{t_1, \dots, t_n\}$  n'est rien d'autre que la série statistique double  $(t_j, y_{t_j})_{1 \leq j \leq n}$ , où

- la première composante est le temps  $t$ ,
- la deuxième composante est une variable numérique  $y$  prenant ses valeurs aux instants  $t$ .

## Convention

On supposera dans toute la suite, que les dates d'observations sont équidistantes et donc nous adopterons la notation simplifiée

$$(y_j)_{j=1,\dots,n}$$

pour désigner la série chronologique

$$\{y_t\}_{t \in T} \text{ avec } T = \{t_1, t_2, \dots, t_n\}.$$



## 1.2 Représentation graphique

### Représentation graphique de la série

On représente graphiquement la série chronologique  $\{y_t\}_{t \in \mathcal{T}}$

- 1 en dessinant le nuage formé par les points  $(t_j, y_j)_{1 \leq j \leq n}$
- 2 en reliant les points entre eux par des segments de droite, pour indiquer la chronologie

**Intérêt de la représentation graphique** : essayer de repérer les caractéristiques de la chronique, comme

- une tendance
- un cycle
- un phénomène périodique
- des variations accidentelles
- des fluctuations irrégulières
- ...

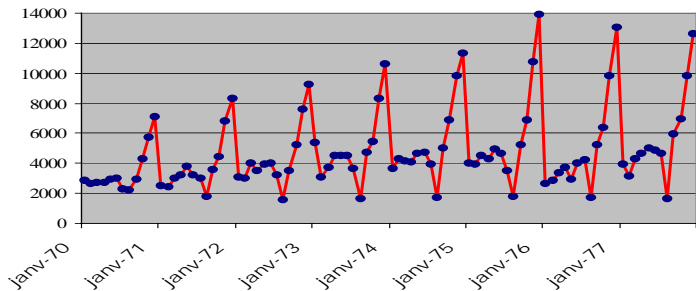
## 1.3 Exemples

**Exemple 1.** On considère la série des ventes mensuelles de bouteilles de champagne en France entre Janvier 1970 et Décembre 1977.

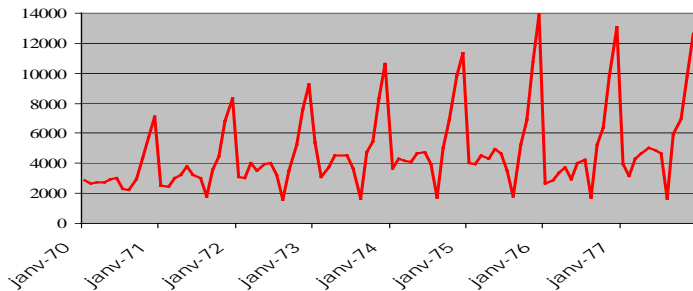
	J	F	M	A	M	J	J	A	S	O	N	D
1970	2851	2672	2755	2721	2946	3036	2282	2212	2922	4301	5764	7132
1971	2541	2475	3031	3266	3776	3230	3028	1759	3595	4474	6838	8357
1972	3113	3006	4047	3523	3937	3986	3260	1573	3528	5211	7614	9254
1973	5375	3088	3718	4514	4520	4539	3663	1643	4739	5428	8314	10651
1974	3633	4292	4154	4121	4647	4753	3965	1723	5048	6922	9858	11331
1975	4016	3957	4510	4276	4968	4677	3523	1821	5222	6873	10803	13916
1976	2639	2899	3370	3740	2927	3986	4217	1738	5221	6424	9842	13076
1977	3934	3162	4286	4676	5010	4874	4633	1659	5951	6981	9851	12670

(en milliers de bouteilles)

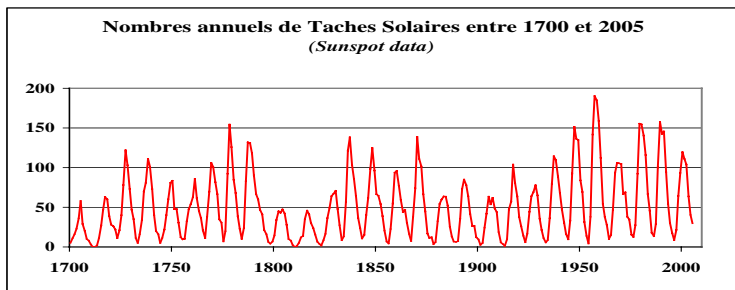
### Ventes mensuelles de Champagne entre Janvier 1970 et Décembre 1977 (en milliers de bouteilles)



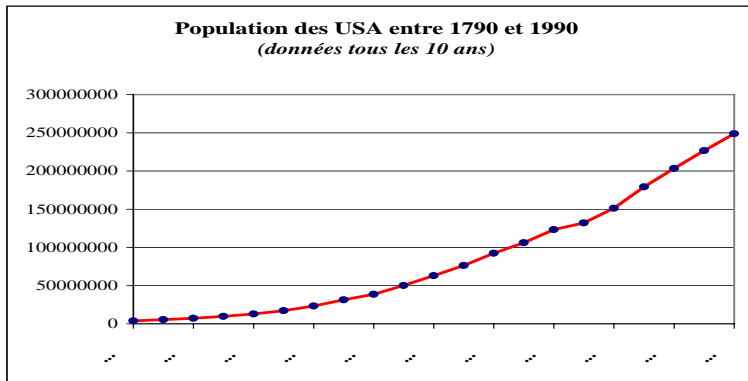
### Ventes mensuelles de Champagne entre Janvier 1970 et Décembre 1977 (en milliers de bouteilles)



**Exemple 2.** On considère la série du nombre annuel de tâches solaires entre 1700 et 2005.



**Exemple 3.** On s'intéresse à l'évolution de la population des USA. On dispose de données recueillies tous les 10 ans entre 1790 et 1990.



## 1.4 Les modèles de décomposition déterministe étudiés

On étudiera deux modèles de décomposition déterministes :

- le modèle additif
- le modèle multiplicatif

combinant chacun :

- une tendance ( $f_i$ ) ;
- une composante saisonnière ( $s_i$ ) ;
- une composante résiduelle ( $e_i$ ).



## 1.4.1 Le modèle additif

### Définition (*Modèle additif*)

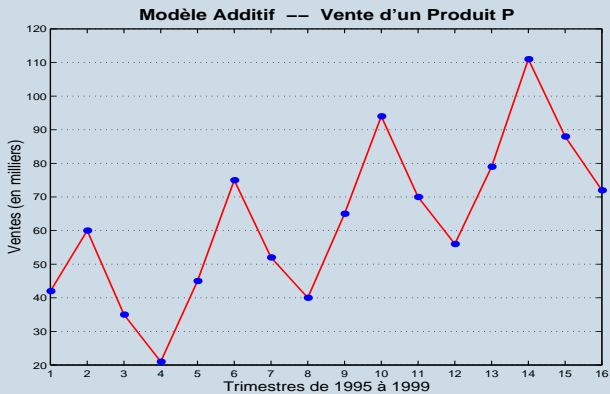
Le modèle additif combine une tendance, une saisonnalité de période  $p$  et une composante résiduelle de la manière suivante :

$$y_i = f_i + s_i + e_i \text{ pour } i = 1, \dots, n \quad (1)$$

avec

$$\sum_{j=1}^p s_j = 0 \text{ et } \sum_{i=1}^n e_i = 0.$$

# Exemple



## 1.4.2 Le modèle multiplicatif

### Définition (*Modèle multiplicatif*)

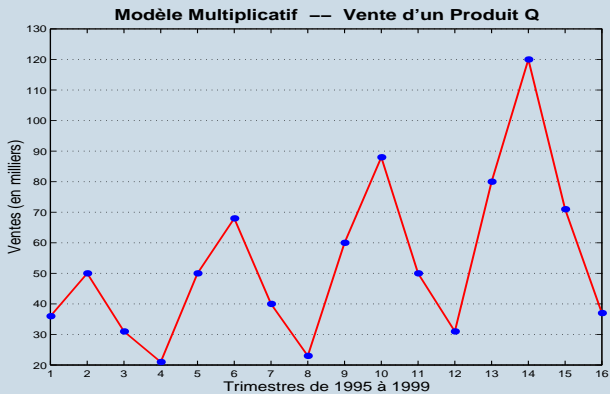
Le modèle multiplicatif combine une tendance, une saisonnalité et une composante résiduelle de la manière suivante :

$$y_i = f_i (1 + s_i) (1 + e_i) \quad \text{pour } i = 1, \dots, n \quad (2)$$

avec

$$\sum_{j=1}^p s_j = 0 \quad \text{et} \quad \sum_{i=1}^n e_i = 0.$$

# Exemple



## 2. Ajustement de la tendance

### 2.1 Introduction

**Le contexte.** On dispose d'une série chronologique  $(y_i)_{i=1,\dots,n}$  où les seules composantes présentes sont

- la **tendance**
- et la **composante résiduelle** (les fluctuations irrégulières).

**Le problème.** Peut-on trouver une fonction simple du temps qui modélise au mieux la tendance de la série  $(y_i)_{i=1,\dots,n}$  ?

On distingue deux types de méthodes :

- **les méthodes dites non-paramétriques**

On suppose que la tendance  $(f_i)_{i=1,\dots,n}$  est de la forme  $(f(t_i))_{i=1,\dots,n}$  où  $f$  est un paramètre fonctionnel, donc de dimension infinie.

Parmi ces méthodes non-paramétriques, on trouve

- ▶ le **lissage par moyennes mobiles**
- ▶ les ondelettes, particulièrement adaptées pour le débruitage
- ▶ les méthodes d'estimation fonctionnelle par noyau
- ▶ ...

## ● les méthodes dites paramétriques

On suppose que la tendance  $(f_i)_{i=1,\dots,n}$  est de la forme  $(f_\theta(t_i))_{i=1,\dots,n}$  où  $\theta$  est un paramètre inconnu de dimension finie, qu'on estimera à l'aide des observations.

Il nous faut alors :

- ▶ choisir une famille de fonctions  $\{f_\theta\}$  dans une collection donnée de fonctions paramétriques, et en général, c'est l'**analyse graphique** de la série chronologique qui détermine la famille de fonctions paramétriques à considérer.
- ▶ une fois la famille choisie, déterminer la valeur de  $\theta$  qui conduit au **meilleur ajustement** de la série  $(y_i)_{i=1,\dots,n}$  et en général, on choisit  $\theta$  à l'aide du critère des moindres carrés, ie. on cherche la valeur de  $\theta$  qui rend minimale

$$\sum_{i=1}^n (y_i - f_\theta(t_i))^2$$

## 2.2 Ajustement linéaire

### Objectif

On veut ajuster les données par une droite et donc modéliser la tendance par une fonction de la forme :

$$f_{\theta}(t) = at + b \text{ avec } \theta = (a, b)'$$

Deux méthodes :

- la méthode des moindres carrés
- la méthode des 2 points



## 2.2.1 La méthode des moindres carrés

- Cette méthode consiste à choisir les coefficients  $a$  et  $b$  de sorte que la quantité

$$\sum_{i=1}^n \left( y_i - (a t_i + b) \right)^2 = g(a, b)$$

soit minimale.

- On cherche donc  $a$  et  $b$  solutions du système :

$$(S) \quad \begin{cases} \frac{\partial g}{\partial a}(a, b) = 0 \\ \frac{\partial g}{\partial b}(a, b) = 0 \end{cases}$$

On démontre que le couple solution est  $(\hat{a}, \hat{b})$  avec :

$$\bullet \hat{a} = \frac{\sum_{i=1}^n (y_i - \bar{y})(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2} = \frac{\sum_{i=1}^n y_i t_i - n\bar{y}\bar{t}}{\sum_{i=1}^n t_i^2 - n\bar{t}^2}$$

$$\bullet \hat{b} = \bar{y} - \hat{a}\bar{t}$$

$$\text{où } \bar{t} = \frac{1}{n} \sum_{i=1}^n t_i \text{ et } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

## Propriétés

- 1 La droite d'équation  $y = \hat{a}t + \hat{b}$  s'appelle **la droite des moindres carrés** (DMC).
- 2 Le point moyen de coordonnées  $(\bar{t}, \bar{y})$  appartient à la DMC.
- 3 On peut apprécier la qualité de l'ajustement linéaire à l'aide du **coefficient de corrélation linéaire** noté  $r$  et défini par :

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^n (t_i - \bar{t})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

## Propriétés (Suite)

- 1 La variance de la série  $(y_i)_{i=1,\dots,n}$  se décompose en

$$\underbrace{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}_{\text{Variance totale de } y} = \underbrace{\frac{1}{n} \sum_{i=1}^n (\hat{a} t_i + \hat{b} - \bar{y})^2}_{\text{Variance expliquée par la DMC}} + \underbrace{\frac{1}{n} \sum_{i=1}^n (y_i - (\hat{a} t_i + \hat{b}))^2}_{\text{Variance résiduelle}}$$

La proportion de variance expliquée par le modèle linéaire est donnée par :

$$\frac{\sum_{i=1}^n (\hat{a} t_i + \hat{b} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = r^2$$

## 2.2.2 La méthode des 2 points

- Cette méthode consiste à choisir arbitrairement deux points et à faire passer la droite d'ajustement par ces deux points.

- Si ces points ont pour coordonnées

$$(t_\ell; y_\ell) \text{ et } (t_{\ell\ell}; y_{\ell\ell})$$

et si la droite a pour équation

$$y = a t + b$$

alors on obtient :

$$\begin{cases} a t_\ell + b = y_\ell \\ a t_{\ell\ell} + b = y_{\ell\ell} \end{cases} \implies \begin{cases} a = \frac{y_{\ell\ell} - y_\ell}{t_{\ell\ell} - t_\ell} \\ b = y_{\ell\ell} - a t_{\ell\ell} \end{cases}$$

- Pour le choix des 2 points, on constitue deux sous-séries d'observations de tailles égales (à 1 près), puis, on prend **les points médians** ou **les points moyens** de chaque série.

## Remarques

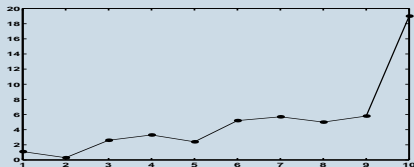
- 1 La méthode des deux points est empirique. Elle n'est basée sur aucun critère d'erreur à minimiser.
- 2 Elle peut cependant s'avérer efficace en présence de valeurs aberrantes, ce qui n'est pas le cas de la méthode des moindres carrés.

## 2.2.3 Illustration

### Exemple

Considérons la série chronologique suivante :

$t_i$	1	2	3	4	5
$y_i$	1.1	0.3	2.6	3.3	2.4
$t_i$	6	7	8	9	10
$y_i$	5.2	5.7	5.0	5.8	19.0



# Calcul des 2 ajustements linéaires

- ① La méthode des moindres carrés donne :

$$\hat{a} = 1.3430 \text{ et } \hat{b} = -2.3467$$

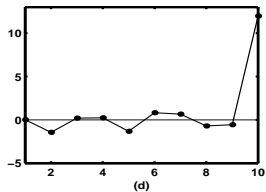
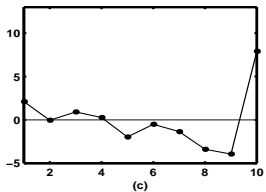
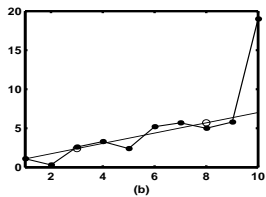
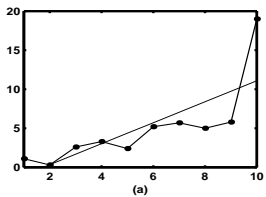
- ② En prenant les points médians des deux sous-séries

$$(3; 2.4) \text{ et } (8; 5.7)$$

la méthode des 2 points donne :

$$a = 0.66 \text{ et } b = 0.42$$





## 2.3 Ajustement polynômial

### Objectif

On veut ajuster les données par un **polynôme de degré  $d$**  et donc modéliser la tendance par une fonction de la forme :

$$f_{\theta}(t) = a_0 + a_1 t + \dots + a_{d-1} t^{d-1} + a_d t^d$$

avec  $\theta = (a_0, a_1, \dots, a_d)'$

Le cas  $d = 0$  correspond à la tendance constante et le cas  $d = 1$  à la tendance linéaire.

## 2.3.1 Le polynôme des moindres carrés

### Objectif

Trouver la valeur de  $\theta = (a_0, a_1, \dots, a_d)'$  qui minimise

$$\sum_{i=1}^n \left( y_i - (a_0 + a_1 t_i + \dots + a_d t_i^d) \right)^2$$

autrement dit, calculer l'estimation des moindres carrés  $\theta^{MC}$  définie par

$$\theta^{MC} = \arg \min_{\theta \in \mathbb{R}^{d+1}} \sum_{i=1}^n \left( y_i - (a_0 + a_1 t_i + \dots + a_d t_i^d) \right)^2$$

On peut montrer en introduisant les notations

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{et} \quad T = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^d \\ 1 & t_2 & t_2^2 & \dots & t_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^d \end{pmatrix}$$

qu'on a encore

$$\theta^{MC} = \arg \min_{\theta \in \mathbb{R}^{d+1}} \|Y - T\theta\|^2$$

La solution de ce problème de minimisation est bien connue :

$$\theta^{MC} = (T'T)^{-1}T'Y$$

où  $T'$  désigne la matrice transposée de  $T$  et

$$T'T = \begin{pmatrix} n & \sum t_i & \sum t_i^2 & \dots & \sum t_i^d \\ \sum t_i & \sum t_i^2 & \sum t_i^3 & \dots & \sum t_i^{d+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum t_i^d & \sum t_i^{d+1} & \sum t_i^{d+2} & \dots & \sum t_i^{2d} \end{pmatrix} \quad T'Y = \begin{pmatrix} \sum y_i \\ \sum t_i y_i \\ \vdots \\ \sum t_i^d y_i \end{pmatrix}$$

En particulier, on peut vérifier que :

- lorsque  $d = 0$ , alors  $T'T = n$  et  $T'Y = \sum_{i=1}^n y_i$ , et donc

$$\hat{a}_0 = \bar{y}.$$

- lorsque  $d = 1$ , alors

$$T'T = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix} \quad \text{et} \quad T'Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n t_i y_i \end{pmatrix}$$

et donc

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \begin{pmatrix} \frac{\sum t_i^2 \sum y_i - \sum t_i \sum t_i y_i}{n \sum t_i^2 - (\sum t_i)^2} \\ \frac{n \sum t_i y_i - \sum t_i \sum y_i}{n \sum t_i^2 - (\sum t_i)^2} \end{pmatrix} = \begin{pmatrix} \frac{\bar{y} \sum t_i^2 - \bar{t} \sum t_i y_i}{\sum t_i^2 - n \bar{t}^2} \\ \frac{\sum t_i y_i - n \bar{t} \bar{y}}{\sum t_i^2 - n \bar{t}^2} \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{a}_1 \bar{t} \\ \frac{\sum (t_i - \bar{t})(y_i - \bar{y})}{\sum (t_i - \bar{t})^2} \end{pmatrix}$$

## 2.3.2 Le choix du degré du polynôme

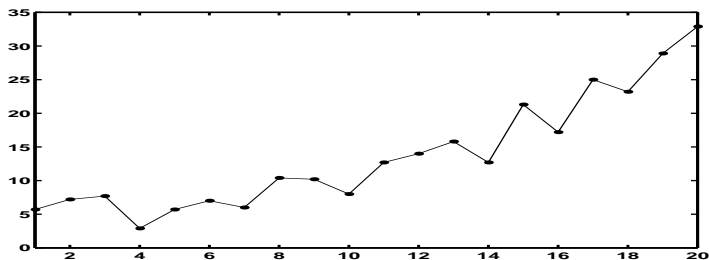
En général, le choix du degré du polynôme est basé sur l'examen

- du graphe de la série  $(y_i)_{1 \leq i \leq n}$ , qui permet de se faire une première idée du degré ;
- des résidus obtenus après l'ajustement. En particulier, on voudra :
  - ▶ obtenir des résidus qui fluctuent autour de 0 et de faible amplitude.
  - ▶ utiliser un polynôme de degré le plus faible possible.

## 2.3.3 Illustration sur un exemple

Considérons la chronique suivante :

$t_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	5.7	7.2	7.7	2.9	5.7	7.0	6.0	10.4	10.2	8.0
$t_i$	11	12	13	14	15	16	17	18	19	20
$y_i$	12.7	14.0	15.8	12.7	21.3	17.2	25.0	23.2	28.9	32.9





On veut ajuster la chronique à l'aide d'un polynôme de degré  $d$  avec  $d = 0, 1, 2$  et  $3$ .

Déterminons les 4 ajustements.

- **Ajustement par un polynôme de degré 0.** On a

$$\hat{a}_0 = \bar{y} = 13.725,$$

et donc

$$f(t) = 13.725$$

- **Ajustement par un polynôme de degré 1.** On a

$$T'T = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix} = \begin{pmatrix} 20 & 210 \\ 210 & 2870 \end{pmatrix}$$

et

$$T'Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n t_i y_i \end{pmatrix} = \begin{pmatrix} 274.5 \\ 3757.6 \end{pmatrix}$$

Donc

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \begin{pmatrix} 0.21579 & -0.015789 \\ -0.015789 & 0.0015038 \end{pmatrix} \begin{pmatrix} 274.5 \\ 3757.6 \end{pmatrix} = \begin{pmatrix} -0.0963 \\ 1.316 \end{pmatrix}$$

et

$$f(t) = -0.0963 + 1.316 t$$

- **Ajustement par un polynôme de degré 2.** On a

$$\begin{aligned}
 T'T &= \begin{pmatrix} n & \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 & \sum_{i=1}^n t_i^3 \\ \sum_{i=1}^n t_i^2 & \sum_{i=1}^n t_i^3 & \sum_{i=1}^n t_i^4 \end{pmatrix} \\
 &= \begin{pmatrix} 20 & 210 & 2870 \\ 210 & 2870 & 44100 \\ 2870 & 44100 & 722666 \end{pmatrix}
 \end{aligned}$$

et

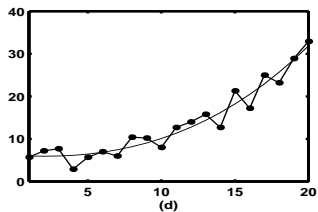
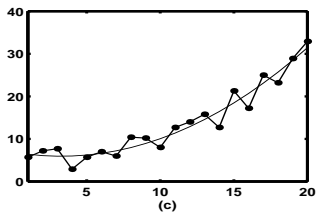
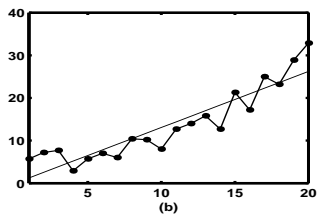
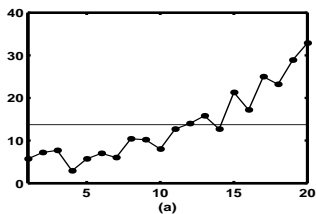
$$T'Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n t_i y_i \\ \sum_{i=1}^n t_i^2 y_i \end{pmatrix} = \begin{pmatrix} 274.5 \\ 3\,757.6 \\ 59\,373 \end{pmatrix}$$

Donc

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \begin{pmatrix} 0.55351 & -0.10789 & 0.00439 \\ -0.10789 & 0.0266 & -0.001196 \\ 0.00439 & -0.001196 & 0.00005696 \end{pmatrix} \begin{pmatrix} 274.5 \\ 3\,757.6 \\ 59\,373 \end{pmatrix}$$
$$= \begin{pmatrix} 6.921 \\ -0.597 \\ 0.091 \end{pmatrix}$$

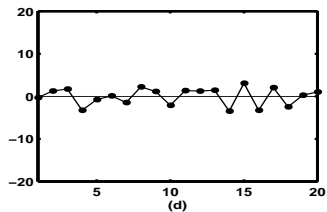
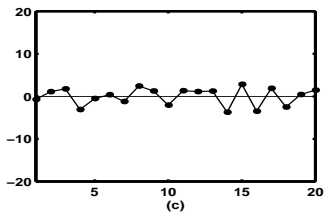
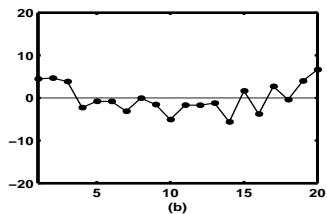
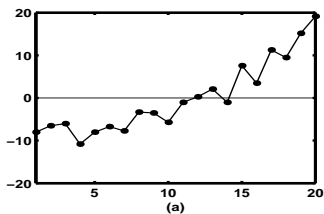
et

$$f(t) = 6.92 - 0.60 t + 0.09 t^2$$



La série  $(y_i)_{1 \leq i \leq 20}$  et son ajustement polynômial  $(f(t_i))_{1 \leq i \leq 20}$  avec en

- (a),  $f(t) = 13.73$
- (b),  $f(t) = -0.10 + 1.32t$
- (c),  $f(t) = 6.92 - 0.60t + 0.09t^2$
- (d),  $f(t) = 6.15 - 0.20t + 0.05t^2 + 0.0015t^3$



La série des résidus  $(y_i - f(t_i))_{1 \leq i \leq 20}$  avec en

- (a),  $f(t) = 13.73$
- (b),  $f(t) = -0.10 + 1.32t$
- (c),  $f(t) = 6.92 - 0.60t + 0.09t^2$
- (d),  $f(t) = 6.15 - 0.20t + 0.05t^2 + 0.0015t^3$

# Critères de choix numériques

- Somme des carrés des résidus

$d$	0	1	2	3
$\sum_{i=1}^{20} (y_i - f(t_i))^2$	1376.82	224.58	78.78	77.84

- Part de variance expliquée

$d$	0	1	2	3
$\frac{\sum_{i=1}^{20} (f(t_i) - \bar{y})^2}{\sum_{i=1}^{20} (y_i - \bar{y})^2}$	0%	83.69%	94.28%	94.35%

## 2.4 Ajustement non linéaire

### Objectif

On souhaite ajuster à la série  $(y_i)_{1 \leq i \leq n}$  une tendance non linéaire paramétrique  $(f_\theta(t_i))_{1 \leq i \leq n}$ .

Plusieurs approches :

- le changement de variable
- l'approche "directe"



## 2.4.1 Le changement de variables

**Principe.** Se ramener à des tendances linéaires ou polynomiales grâce à un simple changement de variables.

Cette approche est possible pour des tendances de la forme :

- $f(t) = 1/(a + bt)$ ,
- $f(t) = b \exp(at)$ ,
- $f(t) = \exp(a + bt)$ ,
- $f(t) = \ln(a + bt + ct^2)$ ,
- ...

## Exemple 1

On souhaite ajuster à une série  $(y_i)_{1 \leq i \leq n}$  une tendance de la forme

$$f(t) = \frac{1}{at + b}.$$

Cela revient à ajuster une tendance de la forme

$$g(t) = 1/f(t) = at + b$$

à la série  $(z_i)_{1 \leq i \leq n} = (1/y_i)_{1 \leq i \leq n}$ .

Il suffit donc de calculer la nouvelle série  $(z_i)$ , puis de calculer les valeurs de  $a$  et  $b$  au moyen de la méthode des moindres carrés ou des 2 points.

## Exemple 2

On souhaite ajuster à une série  $(y_i)_{1 \leq i \leq n}$  une tendance de la forme

$$f(t) = b \exp(at).$$

Cela revient à ajuster une tendance linéaire de la forme

$$g(t) = \log(f(t)) = at + \log(b)$$

à la série  $(z_i)_{1 \leq i \leq n} = (\log(y_i))_{1 \leq i \leq n}$ .

Autrement dit, il suffit de calculer la nouvelle série  $(z_i)$ , puis de calculer les valeurs de  $a$  et  $\log(b)$  au moyen de la méthode des moindres carrés ou des 2 points.

## Remarque

- Bien évidemment, la méthode du changement de variables n'est pas toujours possible.
- C'est la représentation graphique de la série chronologique qui va nous guider pour le choix d'un changement de variables.
- Entre plusieurs changements de variables possibles, c'est la somme des carrés des résidus ou la part de variance expliquée qui déterminera le meilleur choix.

## 2.4.2 L'ajustement "direct" (sur l'exemple de la tendance exponentielle)

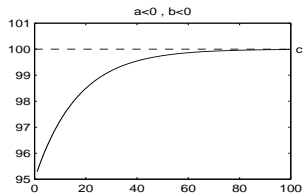
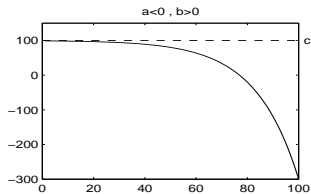
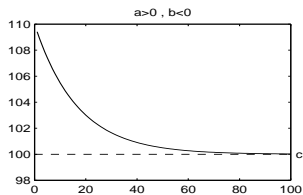
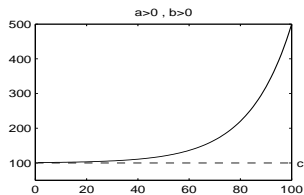
### Objectif

On souhaite ajuster à la série  $(y_i)_{1 \leq i \leq n}$  une tendance de type exponentielle.

La courbe de type exponentiel (ou courbe exponentielle modifiée) est définie par l'équation :

$$f_{\theta}(t) = a e^{bt} + c$$

où  $\theta = (a, b, c)^t$ .



## Calcul des coefficients ( $a, b, c$ ). Plusieurs méthodes :

- Changement de variables lorsque  $c$  est connu
- méthode des moindres carrés
- méthode des 3 points

- **Lorsque  $c$  est connu**, on peut se ramener à une tendance linéaire en posant

$$\begin{aligned}g(t) &= \ln(f(t) - c) \\ &= \ln(a) + bt\end{aligned}$$

On fait alors le changement de variable  $z_i = \ln(y_i - c)$  et on ajuste une droite d'équation  $g(t) = \ln(a) + bt$  à la nouvelle série  $(z_i)_{1 \leq i \leq n}$ .



- **La méthode des moindres carrés** : elle consiste à déterminer  $(a, b, c)$  en minimisant

$$h(a, b, c) = \sum_{i=1}^n (y_i - a \exp(b t_i) - c)^2 .$$

*(mais on ne connaît pas de solutions explicites)*

- **La méthode des 3 points** : on choisit 3 points représentatifs

$$(t_\ell, y_\ell), \quad (t_{\ell\ell}, y_{\ell\ell}) \quad \text{et} \quad (t_{\ell\ell\ell}, y_{\ell\ell\ell})$$

tels que les abscisses  $t_\ell$ ,  $t_{\ell\ell}$  et  $t_{\ell\ell\ell}$  soient équidistantes :

$$\Delta = t_{\ell\ell\ell} - t_{\ell\ell} = t_{\ell\ell} - t_\ell.$$

La fonction exponentielle étant monotone, il faut donc que

- ▶  $y_\ell < y_{\ell\ell} < y_{\ell\ell\ell}$  si la tendance est croissante ;
- ▶ ou  $y_\ell > y_{\ell\ell} > y_{\ell\ell\ell}$  si la tendance est décroissante.

Une fois les trois points choisis, on fait passer la tendance  $y = f(t)$  par ces points. On a donc :

$$\begin{cases} y_\ell & = ae^{bt_\ell} + c \\ y_{\ell\ell} & = ae^{bt_{\ell\ell}} + c \\ y_{\ell\ell\ell} & = ae^{bt_{\ell\ell\ell}} + c \end{cases}$$

On obtient alors

$$\begin{cases} y_{ll} - y_l &= a (e^{bt_{ll}} - e^{bt_l}) \\ y_{lll} - y_{ll} &= a (e^{bt_{lll}} - e^{bt_{ll}}) \end{cases}$$

et donc, en posant  $\Delta = t_{lll} - t_{ll} = t_{ll} - t_l$ ,

$$\frac{y_{lll} - y_{ll}}{y_{ll} - y_l} = \frac{e^{bt_{lll}} - e^{bt_{ll}}}{e^{bt_{ll}} - e^{bt_l}} = \frac{e^{bt_{ll}}(e^{b\Delta} - 1)}{e^{bt_l}(e^{b\Delta} - 1)} = e^{b\Delta}.$$

Donc,

$$\begin{aligned} b &= \frac{1}{\Delta} \ln \left( \frac{y_{lll} - y_{ll}}{y_{ll} - y_l} \right) \\ a &= \frac{y_{lll} - y_{ll}}{e^{bt_{lll}} - e^{bt_{ll}}} \\ c &= y_l - ae^{bt_l} \end{aligned}$$

## Remarques

Le choix des trois points n'est pas unique.

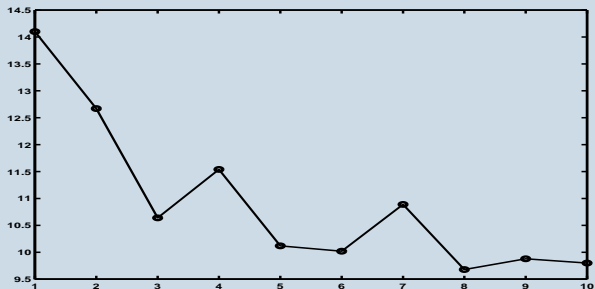
Voici quelques alternatives :

- On répartit les observations en trois sous-séries  $S_1$ ,  $S_2$  et  $S_3$ , tels que les sous ensembles  $S_1$  et  $S_3$  soient exactement de même taille (environ  $n/3$ ). On définit alors les trois points comme les médianes des  $(t_i)$  et des  $(y_i)$  pour ces trois sous-séries.
- Une variante moins robuste que la précédente consiste à utiliser les moyennes, plutôt que les médianes des trois sous-séries.
- On choisit les trois quartiles des séries  $(t_i)_{1 \leq i \leq n}$  et  $(y_i)_{1 \leq i \leq n}$ .
- On choisit les trois points "à la main".

## Exemple

Considérons la chronique suivante :

$t_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	14.10	12.67	10.64	11.54	10.12	10.02	10.89	9.68	9.88	9.80



## 1. Choix des 3 points médians :

- On répartit les données en trois sous-séries de taille 3, 4 et 3.

$t_i$	1	2	3
$y_i$	14.10	12.67	10.64

$t_i$	4	5	6	7
$y_i$	11.54	10.12	10.02	10.89

$t_i$	8	9	10
$y_i$	9.68	9.88	9.80

- Les trois points définis comme les médianes de ces trois sous-séries sont :

$t_\ell$	$y_\ell$
2	12.67

$t_{ll}$	$y_{ll}$
5.5	10.50

$t_{lll}$	$y_{lll}$
9	9.80

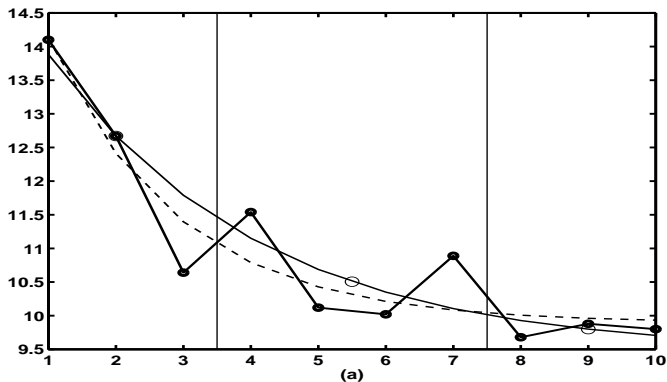
## 2. Calcul des coefficients ( $a, b, c$ ) :

- Les formules précédentes donnent alors :

$$a = 6.10, \quad b = -0.32 \quad \text{et} \quad c = 9.46.$$

- A titre de comparaison, les coefficients obtenus par la méthode des moindres carrés sont :

$$a = 7.08, \quad b = -0.52 \quad \text{et} \quad c = 9.89.$$



Ajustement de la série  $(y_i)_{1 \leq i \leq 10}$  par une courbe exponentielle modifiée obtenue par :

- la méthode des 3 points médians (*courbe en trait continu*)
- la méthode des moindres carrés (*courbe en trait pointillé*)



### 3. La méthode des quartiles pour le choix des trois points.

- Comme la tendance est clairement décroissante, il faut ordonner correctement ces quartiles, c'est-à-dire dans l'ordre croissant pour  $t$  et dans l'ordre décroissant pour  $y$ .

$t_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	14.10	12.67	11.54	10.89	10.64	10.12	10.02	9.88	9.80	9.68

- On trouve alors les points :

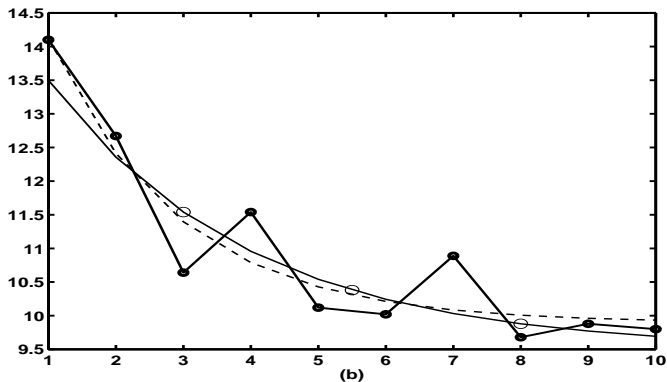
$t_\ell$	$y_\ell$
3	11.54

$t_{\ell\ell}$	$y_{\ell\ell}$
5.5	10.38

$t_{\ell\ell\ell}$	$y_{\ell\ell\ell}$
8	9.88

- et les coefficients :

$$a = 5.60, \quad b = -0.34 \quad \text{et} \quad c = 9.50.$$



Ajustement de la série  $(y_i)_{1 \leq i \leq 10}$  par une courbe exponentielle modifiée obtenue par :

- la méthode des 3 points "quartiles" (*en trait continu*)
- la méthode des moindres carrés (*en trait pointillé*)

## 4. Qualité des ajustements.

- Somme des carrés des résidus

Courbe des 3 pts médians	Courbe des Moindres Carrés	Courbe des 3 pts "quartiles"
0.2495	0.2777	0.5176

- Part de variance expliquée

Courbe des 3 pts médians	Courbe des Moindres Carrés	Courbe des 3 pts "quartiles"
98.68%	98.54%	97.27%

sachant que  $\sum_{i=1}^{10} (y_i - \bar{y})^2 = 18.9602$

### 3. Lissage par moyennes mobiles

#### Objectif

L'objectif d'un lissage par moyenne mobile est de faire apparaître l'allure de la tendance

- en éliminant la composante saisonnière ;
- en atténuant les fluctuations irrégulières.

## 3.1 Définitions des moyennes mobiles

### 3.1.1 Les moyennes mobiles simples

#### Définition (*Moyennes mobiles simples*)

On appelle **série des moyennes mobiles d'ordre  $k$**  de la série  $(t_i, y_i)_{1 \leq i \leq n}$ , la série des moyennes de  $k$  observations consécutives prenant ses valeurs aux dates moyennes correspondantes.

Plus précisément, cette série est définie par :

- la série des instants moyens  $(\bar{t}_i)_{1 \leq i \leq n-k+1}$  avec

$$\bar{t}_i = \frac{t_i + t_{i+1} + \cdots + t_{i+k-1}}{k} \quad \text{pour } i = 1, \dots, n - k + 1$$

- la série  $(\bar{y}_i)_{1 \leq i \leq n-k+1}$  des moyennes empiriques de  $k$  observations consécutives avec

$$\bar{y}_i = \frac{y_i + y_{i+1} + \cdots + y_{i+k-1}}{k} \quad \text{pour } i = 1, \dots, n - k + 1$$

## Remarque 1

On perd  $(k - 1)$  observations lorsqu'on construit la série des moyennes mobiles d'ordre  $k$ .

## Remarque 2

On limitera notre étude au cas où les instants d'observations sont équidistants. Par conséquent,

- si  $k$  est **impair** et vaut  $2m + 1$ , alors la série des moyennes mobiles d'ordre  $k$  est calculée aux instants d'observations

$$t_{m+1}, \dots, t_{n-m}, \text{ ie. } (t_j)_{j=m+1, \dots, n-m}$$

- si  $k$  est **pair** et vaut  $2m$ , alors la série des moyennes mobiles d'ordre  $k$  est calculée aux instants d'observations

$$\frac{t_m + t_{m+1}}{2}, \dots, \frac{t_{n-m} + t_{n-m+1}}{2}, \text{ ie. } \left( \frac{t_j + t_{j+1}}{2} \right)_{j=m, \dots, n-m}$$

## Définition (*Moyennes mobiles simples (bis)*)

On note  $MM(k)$  la série des moyennes mobiles d'ordre  $k$  de la série  $(y_j)_{j=1, \dots, n}$ , et on a

- **lorsque  $k$  est pair et vaut  $2m$  :**

$$MM(k)_j = \frac{\overbrace{y_{j-m+1} + \dots + y_j}^{m \text{ termes}} + \overbrace{y_{j+1} + \dots + y_{j+m}}^{m \text{ termes}}}{2m}$$

pour  $j = m, \dots, n - m$ .

- **lorsque  $k$  est impair et vaut  $2m + 1$  :**

$$MM(k)_j = \frac{y_{j-m} + \dots + y_j + \dots + y_{j+m}}{2m + 1}$$

pour  $j = m + 1, \dots, n - m$ .

### Remarque 3

Il est possible de calculer les moyennes mobiles simples d'ordre  $k$  de manière récursive. En effet,

- **lorsque  $k$  est pair et vaut  $2m$** , on calcule

- ▶  $MM(k)_m = \frac{y_1 + \dots + y_{2m}}{2m}$

- ▶ puis, pour  $j = m, \dots, n - m - 1$ ,

$$MM(k)_{j+1} = MM(k)_j + \frac{y_{j+m+1} - y_{j-m+1}}{2m}$$

- **lorsque  $k$  est impair et vaut  $2m + 1$** , on calcule

- ▶  $MM(k)_{m+1} = \frac{y_1 + \dots + y_{2m+1}}{2m + 1}$

- ▶ puis, pour  $j = m + 1, \dots, n - m - 1$ ,

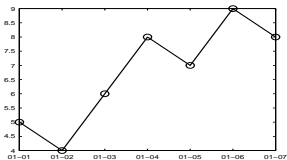
$$MM(k)_{j+1} = MM(k)_j + \frac{y_{j+m+1} - y_{j-m}}{2m + 1}$$



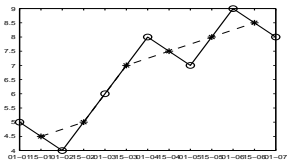
## Exemple

Le tableau suivant présente les moyennes mobiles d'ordre 2, 3 et 4 d'une même série.

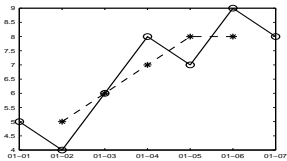
date $t_i$	date $\frac{t_i+t_{i+1}}{2}$	observ. $y_i$	$MM(2)$	$MM(3)$	$MM(4)$
01-01		5			
	15-01		$\frac{5+4}{2} = 4.5$		
01-02		4		$\frac{5+4+6}{3} = 5$	
	15-02		$\frac{4+6}{2} = 5$		$\frac{5+4+6+8}{4} = 5.75$
01-03		6		$\frac{4+6+8}{3} = 6$	
	15-03		$\frac{6+8}{2} = 7$		$\frac{4+6+8+7}{4} = 6.25$
01-04		8		$\frac{6+8+7}{3} = 7$	
	15-04		$\frac{8+7}{2} = 7.5$		$\frac{6+8+7+9}{4} = 7.50$
01-05		7		$\frac{8+7+9}{3} = 8$	
	15-05		$\frac{7+9}{2} = 8$		$\frac{8+7+9+8}{4} = 8.00$
01-06		9		$\frac{7+9+8}{3} = 8$	
	15-06		$\frac{9+8}{2} = 8.5$		
01-07		8			



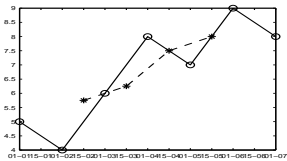
(a)



(b)



(c)



(d)

(a) la série observée ( $y_i, 1 \leq i \leq 7$ ), (b)  $MM(2)$ , (c)  $MM(3)$ , (d)  $MM(4)$

## Exercice (E.1)

Calculer les séries des moyennes-mobiles d'ordre 2, 3 et 4 de la série initiale ( $y_i$ ) suivante.

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

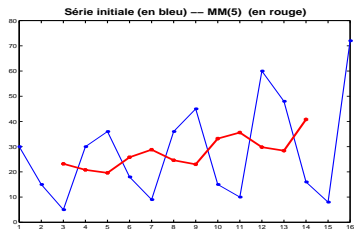
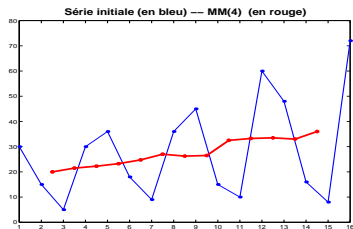
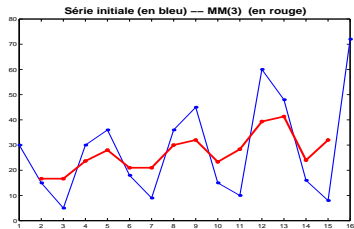
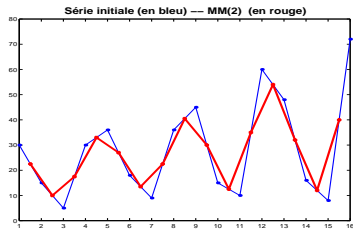
On complétera le tableau suivant.

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	$MM(2)$	$MM(3)$	$MM(4)$
1	30	1.5			
2	15	2.5			
3	5				
4	30				
5	36				
6	18				
7	9				
8	36				
9	45				
10	15				
11	10				
12	60				
13	48				
14	16				
15	8				
16	72				

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MM(3)	MM(4)
1	30	1.5	22.5		
2	15	2.5	10		
3	5	3.5	17.5		
4	30	4.5	33		
5	36	5.5	27		
6	18	6.5	13.5		
7	9	7.5	22.5		
8	36	8.5	40.5		
9	45	9.5	30		
10	15	10.5	12.5		
11	10	11.5	35		
12	60	12.5	54		
13	48	13.5	32		
14	16	14.5	12		
15	8	15.5	40		
16	72				

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MM(3)	MM(4)
1	30	1.5	22.5		
2	15	2.5	10	<b>16.67</b>	
3	5	3.5	17.5	<b>16.67</b>	
4	30	4.5	33	<b>23.67</b>	
5	36	5.5	27	<b>28</b>	
6	18	6.5	13.5	<b>21</b>	
7	9	7.5	22.5	<b>21</b>	
8	36	8.5	40.5	<b>30</b>	
9	45	9.5	30	<b>32</b>	
10	15	10.5	12.5	<b>23.33</b>	
11	10	11.5	35	<b>28.33</b>	
12	60	12.5	54	<b>39.33</b>	
13	48	13.5	32	<b>41.33</b>	
14	16	14.5	12	<b>24</b>	
15	8	15.5	40	<b>32</b>	
16	72				

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MM(3)	MM(4)
1	30	1.5	22.5		
2	15	2.5	10	16.67	20
3	5	3.5	17.5	16.67	21.5
4	30	4.5	33	23.67	22.25
5	36	5.5	27	28	23.25
6	18	6.5	13.5	21	24.75
7	9	7.5	22.5	21	27
8	36	8.5	40.5	30	26.25
9	45	9.5	30	32	26.5
10	15	10.5	12.5	23.33	32.5
11	10	11.5	35	28.33	33.25
12	60	12.5	54	39.33	33.5
13	48	13.5	32	41.33	33
14	16	14.5	12	24	36
15	8	15.5	40	32	
16	72				





### 3.1.2 Les moyennes mobiles centrées

Définition (*Moyennes mobiles centrées*)

On appelle **série des moyennes mobiles centrées d'ordre  $k$**  avec  $k = 2m$ , de la série  $(t_i, y_i)_{1 \leq i \leq n}$ , la série notée  $MMC(k)$  et définie par :

$$\begin{aligned} MMC(k)_j &= \frac{MM(k)_{j-1} + MM(k)_j}{2} \\ &= \frac{\frac{1}{2}y_{j-m} + y_{j-m+1} + \cdots + y_j + \cdots + y_{j+m-1} + \frac{1}{2}y_{j+m}}{2m} \end{aligned}$$

pour  $j = m + 1, \dots, n - m$ .

## Remarques

- ① Lorsque  $k = 2m$ , la série

$$\left( MMC(k)_j \right)_{j=m+1, \dots, n-m}$$

prend ses valeurs aux instants

$$(t_j)_{j=m+1, \dots, n-m}$$

alors que la série

$$\left( MM(k)_j \right)_{j=m, \dots, n-m}$$

prend ses valeurs aux instants

$$\left( \frac{t_j + t_{j+1}}{2} \right)_{j=m, \dots, n-m}$$

- ② La série  $MMC(k)$  comporte un terme de moins que la série  $MM(k)$ .

## Exercice (E.2)

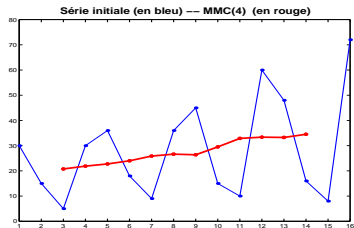
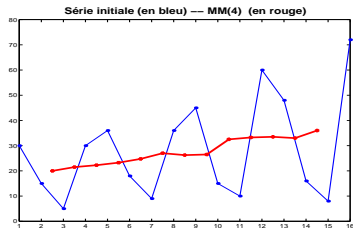
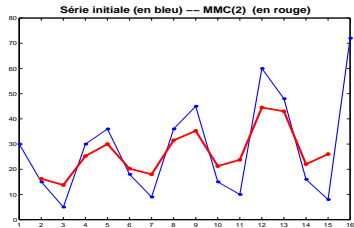
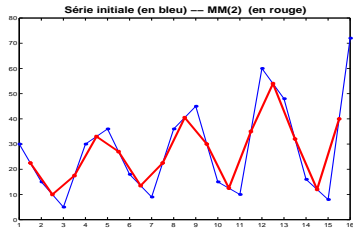
Reprendre la série de l'exercice (E.1) et calculer les séries des moyennes-mobiles centrées d'ordre 2 et 4.

On complétera le tableau suivant.

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MMC(2)	MM(3)	MM(4)	MMC(4)
1	30	1.5	22.5				
2	15	2.5	10		16.67		
3	5	3.5	17.5		16.67	20	
4	30	4.5	33		23.67	21.5	
5	36	5.5	27		28	22.25	
6	18	6.5	13.5		21	23.25	
7	9	7.5	22.5		21	24.75	
8	36	8.5	40.5		30	27	
9	45	9.5	30		32	26.25	
10	15	10.5	12.5		23.33	26.5	
11	10	11.5	35		28.33	32.5	
12	60	12.5	54		39.33	33.25	
13	48	13.5	32		41.33	33.5	
14	16	14.5	12		24	33	
15	8	15.5	40		32	36	
16	72						

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MMC(2)	MM(3)	MM(4)	MMC(4)
1	30	1.5	22.5				
2	15	2.5	10	16.25	16.67		
3	5	3.5	17.5	13.75	16.67	20	
4	30	4.5	33	25.25	23.67	21.5	
5	36	5.5	27	30	28	22.25	
6	18	6.5	13.5	20.25	21	23.25	
7	9	7.5	22.5	18	21	24.75	
8	36	8.5	40.5	31.5	30	27	
9	45	9.5	30	35.25	32	26.25	
10	15	10.5	12.5	21.25	23.33	26.5	
11	10	11.5	35	23.75	28.33	32.5	
12	60	12.5	54	44.5	39.33	33.25	
13	48	13.5	32	43	41.33	33.5	
14	16	14.5	12	22	24	33	
15	8	15.5	40	26	32	36	
16	72						

$(t_i)$	$(y_i)$	$\left(\frac{t_i+t_{i+1}}{2}\right)$	MM(2)	MMC(2)	MM(3)	MM(4)	MMC(4)
1	30	1.5	22.5				
2	15	2.5	10	16.25	16.67	20	
3	5	3.5	17.5	13.75	16.67	21.5	<b>20.75</b>
4	30	4.5	33	25.25	23.67	22.25	<b>21.875</b>
5	36	5.5	27	30	28	23.25	<b>22.75</b>
6	18	6.5	13.5	20.25	21	24.75	<b>24</b>
7	9	7.5	22.5	18	21	27	<b>25.875</b>
8	36	8.5	40.5	31.5	30	26.25	<b>26.625</b>
9	45	9.5	30	35.25	32	26.5	<b>26.375</b>
10	15	10.5	12.5	21.25	23.33	32.5	<b>29.5</b>
11	10	11.5	35	23.75	28.33	33.25	<b>32.875</b>
12	60	12.5	54	44.5	39.33	33.5	<b>33.375</b>
13	48	13.5	32	43	41.33	33	<b>33.25</b>
14	16	14.5	12	22	24	36	<b>34.5</b>
15	8	15.5	40	26	32		
16	72						



## 3.2 Propriétés des moyennes mobiles

### 3.2.1 Elimination de la composante saisonnière

#### Propriété 1.

- 1 Si la série chronologique  $(y_i)_{i=1,\dots,n}$  possède une composante saisonnière de période  $p$ , alors l'application d'une moyenne mobile d'ordre  $p$  supprime cette saisonnalité.
- 2 La série  $MM(p)$  ou  $MMC(p)$  ne possède plus de composante saisonnière de période  $p$ .

#### Remarque

On se servira donc d'une moyenne mobile d'ordre  $p$  pour éliminer une composante saisonnière de période  $p$ .



## Vérification

Considérons une série purement périodique  $(s_i, 1 \leq i \leq n)$  de période  $p$ .

On a donc

- $s_1 + s_2 + \cdots + s_p = 0$ .
- pour tout  $j$ ,  $s_j = s_{j+p}$ .

Calculons maintenant la moyenne mobile d'ordre  $p$ .

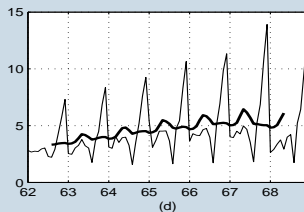
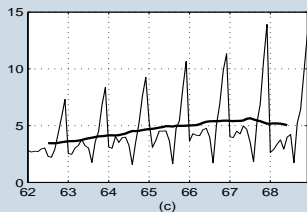
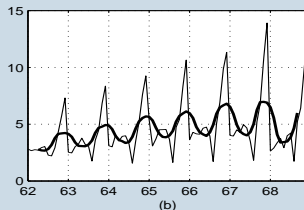
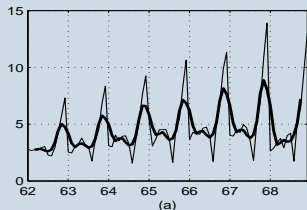
On a pour  $j = 1, \dots, n - p$  :

$$MM(p)_j = \frac{s_j + \cdots + s_{j+p-1}}{p}$$

et puisque la série est périodique de période  $p$ , alors (à une permutation circulaire près des  $p$  termes)

$$MM(p)_j = \frac{s_1 + \cdots + s_p}{p} = 0.$$

## Exemple (*Ventes mensuelles de champagne de 1962 à 1968*)



La série des Ventes mensuelles de Champagne de 1962 à 1968, avec en

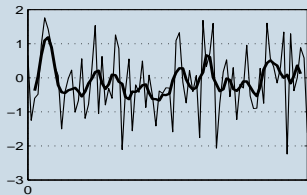
- (a), MM(4)
- (b), MM(6)
- (c), MM(12)
- (d), MM(15)

## 3.2.2 Atténuation des fluctuations irrégulières

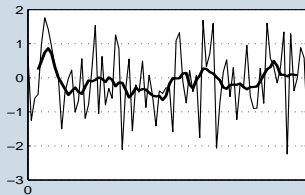
### Propriété 2.

- 1 Une moyenne mobile atténue l'amplitude des fluctuations irrégulières d'une chronique.
- 2 Plus l'ordre de la moyenne mobile est élevé, et plus cette atténuation est importante.
- 3 Donc en appliquant une moyenne mobile sur une série chronologique on obtient **un lissage** de la série.

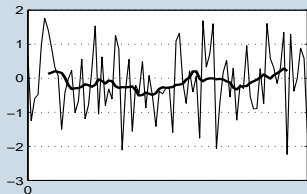
## Exemple (*Moyennes mobiles sur une série de fluctuations irrégulières*)



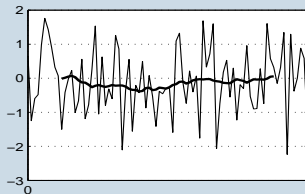
(a)



(b)



(c)



(d)

La série des fluctuations irrégulières, avec en

- (a),  $MM(4)$
- (b),  $MM(6)$
- (c),  $MM(12)$
- (d),  $MM(20)$

## Interprétation (*Probabiliste*)

- Soit  $(\varepsilon_i)_{i=1,\dots,n}$  une suite de variables aléatoires indépendantes de moyenne nulle et de variance  $\sigma^2$ .

Appliquons une  $MM(k)$  à cette suite : pour  $i = 1, \dots, n - k$ ,

$$MM(k)_i = \frac{1}{k} \sum_{j=1}^k \varepsilon_{i+j}$$

et donc 
$$\text{Var}(MM(k)_i) = \frac{1}{k^2} \sum_{j=1}^k \text{Var}(\varepsilon_{i+j}) = \frac{\sigma^2}{k}.$$

La variance est donc divisée par un facteur  $k$ , c'est-à-dire que l'écart-type est divisé par un facteur  $\sqrt{k}$ .

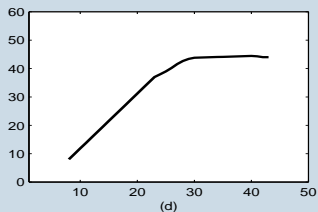
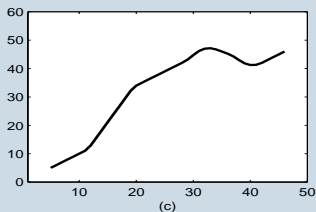
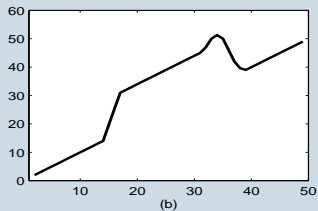
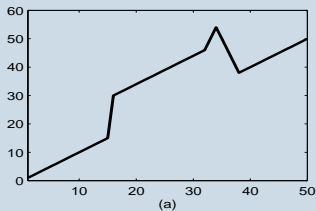
- Donc, si l'on considère une suite de fluctuations irrégulières  $(e_i)$  comme la réalisation d'une suite de variables aléatoires  $(\varepsilon_i)$  indépendantes, centrées et de même variance,  $MM(k)$  réduit l'amplitude de ces fluctuations irrégulières d'un facteur  $\sqrt{k}$ .

## 3.2.3 Effet d'une moyenne mobile sur la tendance

### Propriété 3.

- 1 Une moyenne mobile (d'ordre quelconque) ne modifie pas une tendance constante.
- 2 Une moyenne mobile simple (d'ordre impair quelconque) ou centrée (d'ordre quelconque) ne modifie pas une tendance linéaire.
- 3 Lorsque la tendance n'est pas aussi régulière, les détails disparaîtront d'autant plus que l'ordre de la moyenne mobile sera grand.

## Exemple (Application de moyennes mobiles sur une tendance non régulière)



- (a), La tendance
- (b),  $MM(3)$
- (c),  $MM(8)$
- (d),  $MM(15)$

### 3.3 Choix pratique de l'ordre d'une moyenne mobile

**Rappel.** Le but d'un lissage par moyenne mobile est de faire apparaître l'allure de la tendance.

Lorsque des fluctuations périodiques et/ou irrégulières sont présentes, on aimerait par conséquent réaliser les trois objectifs suivants :

- 1 Supprimer la composante périodique ;
- 2 Réduire le plus possible l'amplitude des fluctuations irrégulières ;
- 3 Et bien sûr, ne pas trop modifier la tendance !



## Remarque

Il est bien rare que ces trois objectifs puissent être parfaitement atteints simultanément. En effet, chacun de ces objectifs conduit à des choix différents de moyenne mobile :

- 1 On supprime la composante périodique de période  $p$  avec une moyenne mobile d'ordre  $p$  ;
- 2 On réduit l'amplitude des fluctuations irrégulières avec une moyenne mobile d'ordre élevé ;
- 3 On préserve les détails de la tendance avec une moyenne mobile d'ordre faible.

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

La solution consiste donc :

- à appliquer  $MM(k)$  ou  $MMC(k)$  pour supprimer la composante périodique dans un premier temps,
- puis à composer éventuellement différentes moyennes mobiles d'ordre peu élevé, jusqu'à l'obtention d'une série suffisamment lissée, mais telle que les détails intéressants de la tendance ne soient pas effacés.

## 4. Décomposition d'une Série Chronologique

On considèrera deux types de décomposition :

- la décomposition à partir d'un modèle additif
- la décomposition à partir d'un modèle multiplicatif

La démarche consistera à :

- identifier les coefficients saisonniers
- désaisonnaliser la série initiale, pour pouvoir ensuite ajuster une courbe de tendance
- construire la série lissée des prédictions en vue de faire de la prévision

## 4.1. Décomposition additive

### Objectif

Soit  $(y_i)_{i=1, \dots, n}$  une série chronologique.

On veut ajuster à cette série un modèle de la forme

$$y_i = f_i + s_i + e_i, \quad i = 1, \dots, n.$$

constitué

- de deux composantes déterministes :
  - ▶ la tendance ( $f_i$ )
  - ▶ la composante saisonnière ( $s_i$ ) périodique de période  $p$  telle que

$$\sum_{j=1}^p s_j = 0$$

- et d'une composante accidentelle ou bruit ( $e_i$ ) correspondant à un aléa.

## 4.1.1 Les étapes de la décomposition additive

On procèdera en suivant les étapes suivantes :

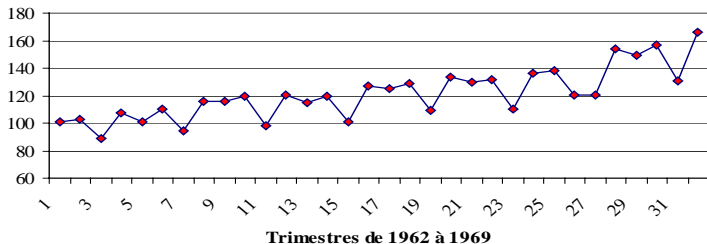
- 1 La désaisonnalisation (5 étapes)
  - 1 Lissage par moyennes mobiles
  - 2 Construction de la série des différences
  - 3 Calcul des coefficients saisonniers non centrés
  - 4 Centrage des coefficients saisonniers
  - 5 Construction de la série corrigée des variations saisonnières
- 2 La série lissée des prévisions (2 étapes)
  - 1 Ajustement d'une tendance
  - 2 Construction de la série lissée des prévisions

## Exemple (*Fil rouge*)

On veut étudier la série chronologique de l'**Indice trimestriel de la production industrielle base 100 en 1962**. Source INSEE.

Années	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
1962	101.3	102.9	88.4	107.3
1963	101	109.8	94.1	116.1
1964	115.6	119.2	97.7	120.3
1965	115.1	119.5	101.1	127.4
1966	124.8	129	109.3	133.6
1967	129.4	131.8	110.2	136.4
1968	138.5	120.1	120.8	154.4
1969	149.5	157.1	130.8	166.5

**Indice trimestriel de la production industrielle base 100  
en 1962. Source INSEE.**



## 4.1.2 La désaisonnalisation : les différentes étapes.

### Etape 1. (*Lissage par Moyennes Mobiles*)

On effectue **un lissage** par la méthode des Moyennes Mobiles afin d'obtenir une première évaluation de la tendance de la série.

On notera  $(\tilde{f}_j)$  la série obtenue.

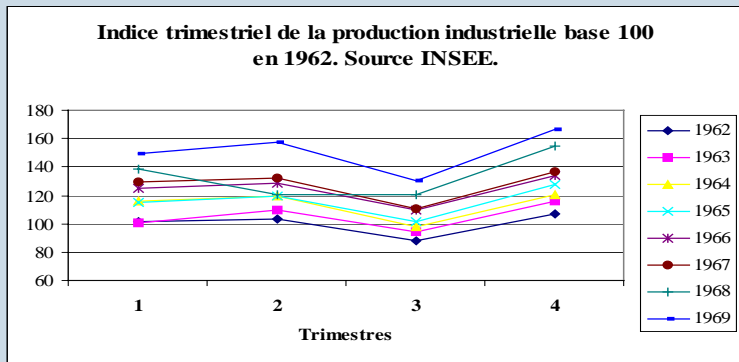
Si  $p$  désigne la période de la composante saisonnière, on aura

$$\tilde{f}_j = \begin{cases} MMC(p)_j & \text{si } p \text{ est pair} \\ MM(p)_j & \text{si } p \text{ est impair} \end{cases}$$

**Remarque.** Dans de nombreuses applications, la période  $p$  est paire et vaut  $2m$  (série mensuelle, trimestrielle, bi-mensuelle, semestrielle, etc). Avec le lissage, on perd  $m$  observations au début de la série et  $m$  observations à la fin, soit au total  $p = 2m$  observations.

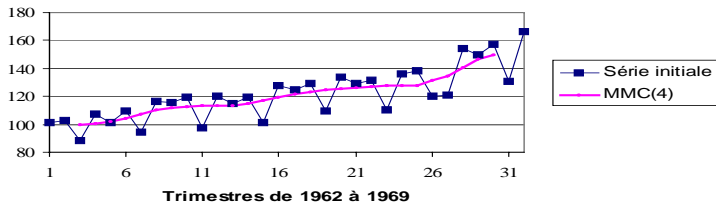


# Application sur l'exemple



Date	Y_j	MM(4)	MMC(4)
1	101.3		
2	102.9		
3	88.4	99.975	99.9375
4	107.3	99.9	100.7625
5	101	101.625	102.3375
6	109.8	103.05	104.15
7	94.1	105.25	107.075
8	116.1	108.9	110.075
9	115.6	111.25	111.7
10	119.2	112.15	112.675
11	97.7	113.2	113.1375
12	120.3	113.075	113.1125
13	115.1	113.15	113.575
14	119.5	114	114.8875
15	101.1	115.775	116.9875
16	127.4	118.2	119.3875
17	124.8	120.575	121.6
18	129	122.625	123.4
19	109.3	124.175	124.75
20	133.6	125.325	125.675
21	129.4	126.025	126.1375
22	131.8	126.25	126.6
23	110.2	126.95	128.0875
24	136.4	129.225	127.7625
25	138.5	126.3	127.625
26	120.1	128.95	131.2
27	120.8	133.45	134.825
28	154.4	136.2	140.825
29	149.5	145.45	146.7
30	157.1	147.95	149.4625
31	130.8	150.975	
32	166.5		

**Indice trimestriel de la production industrielle  
base 100 en 1962. Source INSEE.**



## Etape 2. (*Série des différences*)

On calcule **la série des différences**, soit

$$D_i = y_i - \tilde{f}_i$$

pour les valeurs de  $i$  disponibles, soit

$$i = m + 1, \dots, n - m.$$

Date	Y_j	MMC(4)	D_j
1	101.3		
2	102.9		
3	88.4	99.9375	<b>-11.5375</b>
4	107.3	100.7625	<b>6.5375</b>
5	101	102.3375	<b>-1.3375</b>
6	109.8	104.15	<b>5.65</b>
7	94.1	107.075	<b>-12.975</b>
8	116.1	110.075	<b>6.025</b>
9	115.6	111.7	<b>3.9</b>
10	119.2	112.675	<b>6.525</b>
11	97.7	113.1375	<b>-15.4375</b>
12	120.3	113.1125	<b>7.1875</b>
13	115.1	113.575	<b>1.525</b>
14	119.5	114.8875	<b>4.6125</b>
15	101.1	116.9875	<b>-15.8875</b>
16	127.4	119.3875	<b>8.0125</b>
17	124.8	121.6	<b>3.2</b>
18	129	123.4	<b>5.6</b>
19	109.3	124.75	<b>-15.45</b>
20	133.6	125.675	<b>7.925</b>
21	129.4	126.1375	<b>3.2625</b>
22	131.8	126.6	<b>5.2</b>
23	110.2	128.0875	<b>-17.8875</b>
24	136.4	127.7625	<b>8.6375</b>
25	138.5	127.625	<b>10.875</b>
26	120.1	131.2	<b>-11.1</b>
27	120.8	134.825	<b>-14.025</b>
28	154.4	140.825	<b>13.575</b>
29	149.5	146.7	<b>2.8</b>
30	157.1	149.4625	<b>7.6375</b>
31	130.8		
32	166.5		

### Etape 3. (Calcul des coefficients saisonniers non centrés)

On calcule les  $p$  coefficients saisonniers non centrés ( $\tilde{s}_j$ ).

Pour cela, on effectue des moyennes avec les valeurs de la série  $(D_j)_{m+1 \leq j \leq n-m}$ .

Pour simplifier l'écriture, supposons que  $n$  soit multiple de la période  $p$  de la saisonnalité de telle sorte que l'on dispose de  $K_0$  périodes complètes d'observations, i.e.  $n = K_0 * p$ .

Par exemple si la saisonnalité est annuelle  $p = 12$  et que la série est observée sur  $n = 36$  mois, on a  $K_0 = 3$  périodes complètes.

- Pour  $j \geq m$ ,  $\tilde{s}_j = \frac{1}{K_0 - 1} \sum_{k=1}^{K_0-1} D_{j+p(k-1)}$
- Pour  $j \leq m$ ,  $\tilde{s}_j = \frac{1}{K_0 - 1} \sum_{k=1}^{K_0-1} D_{j+pk}$

## Application sur l'exemple

Ici,  $n = 32$  et  $p = 4$ , donc  $K_0 = 8$  et  $m = 2$ .

Par conséquent,

$$\tilde{s}_3 = \frac{1}{7} \sum_{k=1}^7 D_{3+4(k-1)} = \frac{1}{7} (D_3 + D_7 + D_{11} + D_{15} + D_{19} + D_{23} + D_{27})$$

$$\tilde{s}_4 = \frac{1}{7} \sum_{k=1}^7 D_{4+4(k-1)} = \frac{1}{7} (D_4 + D_8 + D_{12} + D_{16} + D_{20} + D_{24} + D_{28})$$

et

$$\tilde{s}_1 = \frac{1}{7} \sum_{k=1}^7 D_{1+4k} = \frac{1}{7} (D_5 + D_9 + D_{13} + D_{17} + D_{21} + D_{25} + D_{29})$$

$$\tilde{s}_2 = \frac{1}{7} \sum_{k=1}^7 D_{2+4k} = \frac{1}{7} (D_6 + D_{10} + D_{14} + D_{18} + D_{22} + D_{26} + D_{30})$$

## Application sur l'exemple (*suite*)

On trouve :

- $\tilde{s}_3 = -14.7428571$

- $\tilde{s}_4 = 8.27142857$

et

- $\tilde{s}_1 = 3.46071429$

- $\tilde{s}_2 = 3.44642857$



Date	Y_j	MMC(4)	D_j	Stild_j
1	101.3			3.46071429
2	102.9			3.44642857
3	88.4	99.9375	-11.5375	-14.7428571
4	107.3	100.7625	6.5375	8.27142857
5	101	102.3375	-1.3375	
6	109.8	104.15	5.65	
7	94.1	107.075	-12.975	
8	116.1	110.075	6.025	
9	115.6	111.7	3.9	
10	119.2	112.675	6.525	
11	97.7	113.1375	-15.4375	
12	120.3	113.1125	7.1875	
13	115.1	113.575	1.525	
14	119.5	114.8875	4.6125	
15	101.1	116.9875	-15.8875	
16	127.4	119.3875	8.0125	
17	124.8	121.6	3.2	
18	129	123.4	5.6	
19	109.3	124.75	-15.45	
20	133.6	125.675	7.925	
21	129.4	126.1375	3.2625	
22	131.8	126.6	5.2	
23	110.2	128.0875	-17.8875	
24	136.4	127.7625	8.6375	
25	138.5	127.625	10.875	
26	120.1	131.2	-11.1	
27	120.8	134.825	-14.025	
28	154.4	140.825	13.575	
29	149.5	146.7	2.8	
30	157.1	149.4625	7.6375	
31	130.8			
32	166.5			

#### Etape 4. (*Centrage des coefficients saisonniers.*)

On calcule la moyenne arithmétique des  $p$  coefficients saisonniers obtenus à l'étape 3 :

$$\bar{s} = \frac{1}{p} \sum_{j=1}^p \tilde{s}_j.$$

Puis on centre les coefficients saisonniers  $(\tilde{s}_j)_{j=1,\dots,p}$  en retranchant la moyenne  $\bar{s}$  et on obtient :

$$\hat{s}_j = \tilde{s}_j - \bar{s}, \quad \text{pour } j = 1, \dots, p.$$

## Application sur l'exemple

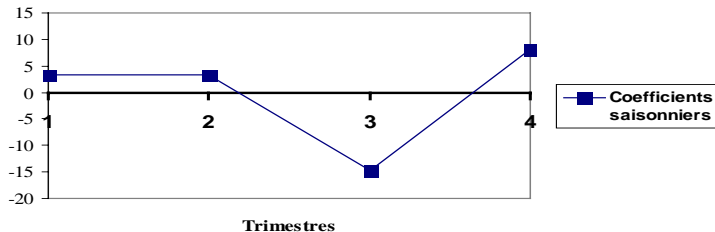
Sur l'exemple, on a

$$\begin{aligned}\bar{s} &= \frac{1}{4} \sum_{j=1}^4 \tilde{s}_j = \frac{1}{4} (\tilde{s}_1 + \tilde{s}_2 + \tilde{s}_3 + \tilde{s}_4) \\ &= 0.10892857\end{aligned}$$

et on trouve :

- $\hat{s}_1 = 3.35178571$
- $\hat{s}_2 = 3.3375$
- $\hat{s}_3 = -14.8517857$
- $\hat{s}_4 = 8.1625$

**Indice trimestriel de la production industrielle base 100  
en 1962. Source INSEE.**



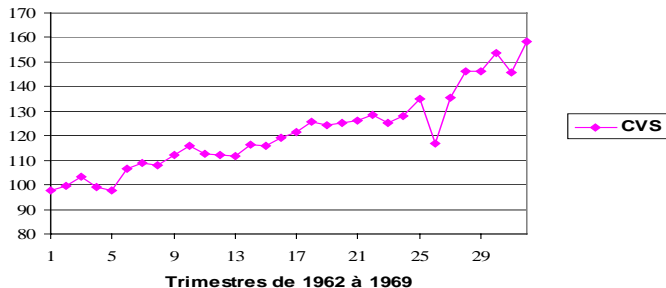
## Etape 5. (*La série corrigée des variations saisonnières*)

La série corrigée des variations saisonnières ou la série désaisonnalisée notée  $(CVS_i)_{1 \leq i \leq n}$  s'obtient en retranchant à la série initiale  $(y_i)_{1 \leq i \leq n}$ , la suite des coefficients saisonniers centrés obtenus à l'étape 4 :

$$CVS_j = y_j - \hat{s}_j, \quad \text{pour } j = 1, \dots, n.$$

Date	Y	Schap_j	CVS
1	101.3	3.35178571	97.9482143
2	102.9	3.3375	99.5625
3	88.4	-14.8517857	103.251786
4	107.3	8.1625	99.1375
5	101	3.35178571	97.6482143
6	109.8	3.3375	106.4625
7	94.1	-14.8517857	108.951786
8	116.1	8.1625	107.9375
9	115.6	3.35178571	112.248214
10	119.2	3.3375	115.8625
11	97.7	-14.8517857	112.551786
12	120.3	8.1625	112.1375
13	115.1	3.35178571	111.748214
14	119.5	3.3375	116.1625
15	101.1	-14.8517857	115.951786
16	127.4	8.1625	119.2375
17	124.8	3.35178571	121.448214
18	129	3.3375	125.6625
19	109.3	-14.8517857	124.151786
20	133.6	8.1625	125.4375
21	129.4	3.35178571	126.048214
22	131.8	3.3375	128.4625
23	110.2	-14.8517857	125.051786
24	136.4	8.1625	128.2375
25	138.5	3.35178571	135.148214
26	120.1	3.3375	116.7625
27	120.8	-14.8517857	135.651786
28	154.4	8.1625	146.2375
29	149.5	3.35178571	146.148214
30	157.1	3.3375	153.7625
31	130.8	-14.8517857	145.651786
32	166.5	8.1625	158.3375

### Indice trimestriel de la production industrielle base 100 en 1962. Source INSEE.



## 4.1.3 La série lissée des prédictions

Etape 1. (*Ajustement d'une courbe de tendance*)

Maintenant que la série est corrigée des variations saisonnières, on peut ajuster une courbe de tendance de la forme

$$y = f_{\theta}(t).$$

Par exemple, si on décide d'ajuster une **tendance linéaire** de la forme :

$$f_{\theta}(t) = at + b$$

alors on aura :

$$\hat{a} = \frac{\text{Cov}(\text{CVS}; (t_i)_{1 \leq i \leq n})}{\text{Var}((t_i)_{1 \leq i \leq n})} \quad \text{et} \quad \hat{b} = \overline{\text{CVS}} - \hat{a} \bar{t}$$



## Application sur l'exemple

Sur l'exemple, on va ajuster un polynôme de degré 2.

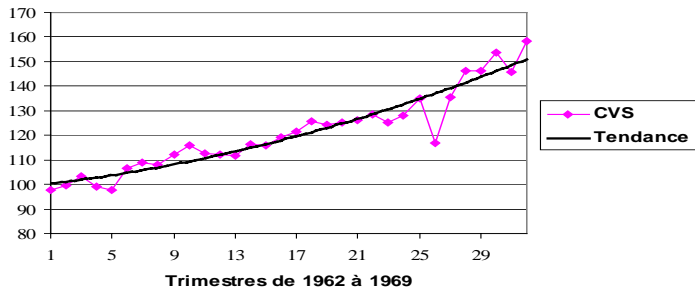
On trouve

$$\hat{f}(t) = 0.0287 t^2 + 0.6873 t + 99.621$$

et donc pour  $i = 1, \dots, 32$ ,

$$\hat{f}_i = 0.0287 t_i^2 + 0.6873 t_i + 99.621$$

### Indice trimestriel de la production industrielle base 100 en 1962. Source INSEE.



## Etape 2. (*Construction de la série lissée des prédictions*)

La série lissée des prédictions  $(\hat{y}_i)_{1 \leq i \leq n}$ , utile pour la prévision, est obtenue en additionnant la tendance  $(\hat{f}_i)_{1 \leq i \leq n}$  et la composante saisonnière  $(\hat{s}_i)_{1 \leq i \leq n}$ , ie.

$$\hat{y}_i = \hat{f}_i + \hat{s}_i \quad \text{pour } i = 1, \dots, n$$

On peut alors définir l'erreur de prévision associée

$$\hat{e}_i = y_i - \hat{y}_i \quad \text{pour } i = 1, \dots, n$$

On pourra étudier la qualité de la modélisation de la chronique

- à l'aide de représentations graphiques
- à l'aide de critères numériques, comme l'erreur quadratique moyenne

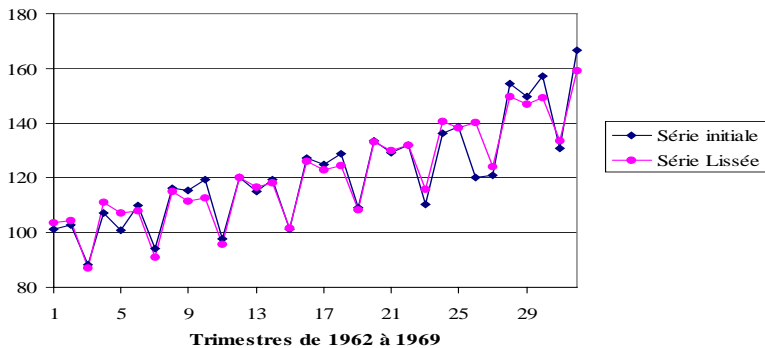
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

ou bien encore l'erreur absolue moyenne (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

qui sont des **critères d'erreurs** permettant de comparer différentes modélisations.

**Indice trimestriel de la production industrielle base 100 en 1962.**  
**Source INSEE.**





## 4.2. Décomposition multiplicative

### Objectif

Soit  $(y_i)_{i=1, \dots, n}$  une série chronologique.

On veut ajuster à cette série un modèle de la forme

$$y_i = f_i(1 + s_i)(1 + e_i), \quad i = 1, \dots, n.$$

constitué

- de deux composantes déterministes :
  - ▶ la tendance ( $f_i$ )
  - ▶ la composante saisonnière ( $s_i$ ) périodique de période  $p$  telle que

$$\sum_{j=1}^p s_j = 0$$

- et d'une composante accidentelle ou bruit ( $e_i$ ) correspondant à un aléa.

## 4.2.1 Les étapes de la décomposition multiplicative

On procèdera en suivant les étapes suivantes :

- 1 La désaisonnalisation (5 étapes)
  - 1 Lissage par moyennes mobiles
  - 2 Construction de la série des quotients
  - 3 Calcul des coefficients saisonniers non centrés
  - 4 Centrage des coefficients saisonniers
  - 5 Construction de la série corrigée des variations saisonnières
  
- 2 La série lissée des prévisions (2 étapes)
  - 1 Ajustement d'une tendance
  - 2 Construction de la série lissée des prévisions



## 4.2.2 La désaisonnalisation : les différentes étapes.

### Etape 1. (*Lissage par Moyennes Mobiles*)

On effectue **un lissage** par la méthode des Moyennes Mobiles afin d'obtenir une première évaluation de la tendance de la série.

On notera  $(\tilde{f}_j)$  la série obtenue.

Si  $p$  désigne la période de la composante saisonnière, on aura

$$\tilde{f}_j = \begin{cases} MMC(p)_j & \text{si } p \text{ est pair} \\ MM(p)_j & \text{si } p \text{ est impair} \end{cases}$$

**Remarque.** Dans de nombreuses applications, la période  $p$  est paire et vaut  $2m$  (série mensuelle, trimestrielle, bi-mensuelle, semestrielle, etc). Avec le lissage, on perd  $m$  observations au début de la série et  $m$  observations à la fin, soit au total  $p = 2m$  observations.

## Etape 2. (Série des Quotients)

On calcule **la série des quotients**, soit

$$Q_i = \frac{y_i}{\widetilde{f}_i}$$

pour les valeurs de  $i$  disponibles, soit

$$i = m + 1, \dots, n - m.$$

### Etape 3. (Calcul des coefficients saisonniers non centrés)

On calcule les  $p$  coefficients saisonniers non centrés ( $\tilde{s}_j$ ).

Pour cela, on effectue des moyennes avec les valeurs de la série  $(Q_j)_{m+1 \leq j \leq n-m}$ .

Pour simplifier l'écriture, supposons que  $n$  soit multiple de la période  $p$  de la saisonnalité de telle sorte que l'on dispose de  $K_0$  périodes complètes d'observations, i.e.  $n = K_0 * p$ .

Par exemple si la saisonnalité est annuelle  $p = 12$  et que la série est observée sur  $n = 36$  mois, on a  $K_0 = 3$  périodes complètes.

- Pour  $j \geq m$ ,  $\tilde{s}_j = \frac{1}{K_0 - 1} \sum_{k=1}^{K_0-1} Q_{j+p(k-1)}$
- Pour  $j \leq m$ ,  $\tilde{s}_j = \frac{1}{K_0 - 1} \sum_{k=1}^{K_0-1} Q_{j+pk}$

#### Etape 4. (*Centrage des coefficients saisonniers.*)

On calcule la moyenne arithmétique des  $p$  coefficients saisonniers obtenus à l'étape 3 :

$$\bar{s} = \frac{1}{p} \sum_{j=1}^p \tilde{s}_j.$$

Puis on centre les coefficients saisonniers  $(\tilde{s}_j)_{j=1,\dots,p}$  en retranchant la moyenne  $\bar{s}$  et on obtient :

$$\hat{s}_j = \tilde{s}_j - \bar{s}, \quad \text{pour } j = 1, \dots, p.$$

## Etape 5. (*La série corrigée des variations saisonnières*)

La série corrigée des variations saisonnières ou la série désaisonnalisée notée  $(CVS_i)_{1 \leq i \leq n}$  s'obtient en divisant

- les valeurs de la série initiale  $(y_i)_{1 \leq i \leq n}$
- par les valeurs de la suite des coefficients saisonniers centrés (obtenus à l'étape 4) auxquels on ajoute 1 :

$$CVS_j = \frac{y_j}{1 + \hat{s}_j}, \quad \text{pour } j = 1, \dots, n.$$

## 4.2.3 La série lissée des prédictions

Etape 1. (*Ajustement d'une courbe de tendance*)

Maintenant que la série est corrigée des variations saisonnières, on peut ajuster une courbe de tendance de la forme

$$y = f_{\theta}(t).$$

Par exemple, si on décide d'ajuster une **tendance linéaire** de la forme :

$$f_{\theta}(t) = at + b$$

alors on aura :

$$\hat{a} = \frac{\text{Cov}(\text{CVS}; (t_i)_{1 \leq i \leq n})}{\text{Var}((t_i)_{1 \leq i \leq n})} \quad \text{et} \quad \hat{b} = \overline{\text{CVS}} - \hat{a} \bar{t}$$

## Etape 2. (*Construction de la série lissée des prédictions*)

La série lissée des prédictions  $(\hat{y}_i)_{1 \leq i \leq n}$ , utile pour la prévision, est obtenue en multipliant la tendance  $(\hat{f}_i)_{1 \leq i \leq n}$  par la composante saisonnière  $(1 + \hat{s}_i)_{1 \leq i \leq n}$ , ie.

$$\hat{y}_i = \hat{f}_i (1 + \hat{s}_i) \quad \text{pour } i = 1, \dots, n$$