IUT de Paris - Rives de Seine
Université Paris Cité

Séries chronologiques - Partie 2

Ajustement de la tendance par moyennes mobiles

BUT STID, deuxième année

Contexte et objectif

Contexte : On dispose d'une série chronologique $(y_i)_{i=1,\dots,n}$ où les composantes présentes sont :

- La **tendance**.
- La **composante saisonnière**.
- La **composante résiduelle** (les fluctuations irrégulières).

On a donc :

$$y_i = f_i + s_i + e_i \quad \text{si le modèle est additif.}$$

$$y_i = F_i \times S_i \times E_i \quad \text{si le modèle est multiplicatif.}$$

Objectif : Trouver une fonction simple du temps qui modélise *au mieux* la tendance de la série $(y_i)_{i=1,\dots,n}$.

Important : notation

Désormais, sauf mention contraire, on confondra les instants t_i et les indices temporels i :

$$t_i = i.$$

Une méthode non paramétrique : le lissage par moyennes mobiles

Les moyennes mobiles permettent de lisser directement la série sans hypothèse a priori sur la forme du modèle sous-jacent (aucune hypothèse sur l'allure de la tendance).

Il s'agit d'une méthode non paramétrique car la courbe obtenue n'est pas déterminé par un nombre fini de paramètres (comme par exemple dans le cas d'un ajustement polynomial par moindres carrés).

Cette méthode se rapproche de la courbe de régression, vue en première année (voir les illustrations sous R sur le taux de CO2).

On dispose d'une série chronologique $(y_i)_{i=1,\dots,n}$.

Idée : On propose d'estimer la tendance en un point par une moyenne des observations qui l'entourent.

La série des moyennes mobiles d'ordre k, notée $MM(k)$, est la série des moyennes de k observations consécutives qui prend ses valeurs aux dates moyennes correspondantes.

Les moyennes mobiles simples

Plus précisément, on calcule :

- Les dates « moyennes » = les moyennes de k termes consécutifs pour les dates :

$$\frac{t_1 + t_2 + \ldots + t_k}{k}, \quad \text{puis} \quad \frac{t_2 + t_3 + \ldots + t_{k+1}}{k},$$

$$\ldots \text{ jusqu'à } \quad \frac{t_{n-k+1} + t_{n-k+2} + \ldots + t_n}{k}.$$

- Les moyennes de k termes consécutifs pour la variable y :

$$\frac{y_1 + y_2 + \ldots + y_k}{k}, \quad \text{puis} \quad \frac{y_2 + y_3 + \ldots + y_{k+1}}{k},$$

$$\ldots \text{ jusqu'à } \quad \frac{y_{n-k+1} + y_{n-k+2} + \ldots + y_n}{k}.$$

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

- Si k est impair et vaut $2m+1$:
  La série des moyennes mobiles simples d'ordre k est calculée aux instants d'observation $t_{m+1}, \dots, t_{n-m}$.
  Pour $i = m+1, \dots, n-m$:

$$MM(k)_i = \frac{\overbrace{y_{i-m} + \dots + y_{i-1}}^{m \text{ termes}} + y_i + \overbrace{y_{i+1} + \dots + y_{i+m}}^{m \text{ termes}}}{2m+1}.$$

- Si k est pair et vaut $2m$:
  Les valeurs des moyennes obtenues se trouvent « entre les lignes » du tableau, et ne se rapportent donc plus aux dates d'observation.
  En fait, la série des moyennes mobiles simples d'ordre k est calculée aux instants $\left(\frac{t_i+t_{i+1}}{2}\right)_{i=m,\dots,n-m}$.

  Pour $i = m, \dots, n-m$:

$$MM(k)_i = \frac{\overbrace{y_{i-m+1}+\dots+y_{i-1}}^{m-1 \text{ termes}} + y_i + \overbrace{y_{i+1}+\dots+y_{i+m}}^{m \text{ termes}}}{2m}.$$

Le tableau suivant présente les moyennes mobiles simples d'ordre 2 d'une série de 7 observations ($n = 7$).

Calculons $MM(2)$.

t_i	$\frac{t_i+t_{i+1}}{2}$	y_i	$MM(2)_i$
1		5	
2		4	
3		6	
4		8	
5		7	
6		9	
7		8	

Exercice

Le tableau suivant présente les moyennes mobiles simples d'ordre 2 d'une série de 7 observations ($n = 7$).

t_i	$\frac{t_i+t_{i+1}}{2}$	y_i	$MM(2)_i$	$MM(3)_i$	$MM(4)_i$
1		5			
	1.5		$\frac{5+4}{2} = 4.5$		
2		4		$\frac{5+4+6}{3} = 5$	
	2.5		$\frac{4+6}{2} = 5$		$\frac{5+4+6+8}{4} = 5.75$
3		6		$\frac{4+6+8}{3} = 6$	
	3.5		$\frac{6+8}{2} = 7$		$\frac{4+6+8+7}{4} = 6.25$
4		8		$\frac{6+8+7}{3} = 7$	
	4.5		$\frac{8+7}{2} = 7.5$		$\frac{6+8+7+9}{4} = 7.50$
5		7		$\frac{8+7+9}{3} = 8$	
	5.5		$\frac{7+9}{2} = 8$		$\frac{8+7+9+8}{4} = 8.00$
6		9		$\frac{7+9+8}{3} = 8$	
	6.5		$\frac{9+8}{2} = 8.5$		
7		8			

Calculer les séries des moyennes mobiles d'ordre 2, 3 et 4 de la série 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

Complétons le tableau suivant...

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

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

Remarques

Un certain nombre de valeurs, correspondant aux extrémités de l'intervalle de variation de la série originale sont perdues.

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

Les moyennes mobiles centrées

On a vu qu'une moyenne mobile d'ordre pair se calcule à des dates qui ne coïncident pas avec les dates des observations.

Cela est gênant : si l'on veut comparer la série lissée par moyennes mobiles avec la série initiale, on a besoin d'avoir les valeurs pour les mêmes dates d'observation.

A une date d'observation donnée, on convient de lui affecter la moyenne arithmétique des deux moyennes mobiles qui l'encadrent, pour les moyennes mobiles d'ordre pair. C'est le principe des moyennes mobiles centrées.

On définit la série des moyennes mobiles centrées d'ordre k avec $k = 2m$, notée $MMC(k)$, à partir de la moyenne mobile d'ordre pair $MM(k)$, par :

$$MMC(k)_i \quad = \quad \frac{MM(k)_{i-1} \ + \ MM(k)_i}{2}$$

pour $i = m+1, \ldots, n-m$.

On a :

$$MMC(k)_i \quad = \quad \frac{\overbrace{\frac{1}{2}y_{i-m} + y_{i-m+1} + \ldots}^{m \text{ termes}} + y_i + \overbrace{\ldots + y_{i+m-1} + \frac{1}{2}y_{i+m}}^{m \text{ termes}}}{2m}.$$

Remarques

- Lorsque $k = 2m$, la série

$$\Big(MMC(k)_i\Big)_{i=m+1,\ldots,n-m}$$

prend ses valeurs aux instants $m+1, \ldots, n-m$
alors que la série

$$\Big(MM(k)_i\Big)_{i=m,\ldots,n-m}$$

prend ses valeurs aux instants $\left(\frac{t_i+t_{i+1}}{2}\right)_{i=m,\ldots,n-m}$.

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

Exercice

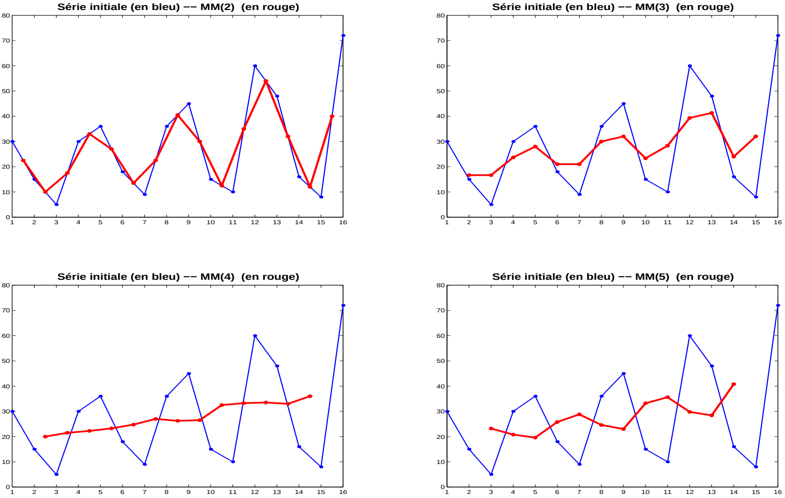
Reprendre la série de l'exercice précédent et calculer les séries des moyennes mobiles centrées d'ordre 2 et 4.

Complétons le tableau suivant...

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

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

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


Série initiale (en bleu) –– MM(2) (en rouge)
Série initiale (en bleu) –– MM(3) (en rouge)
Série initiale (en bleu) –– MM(4) (en rouge)
Série initiale (en bleu) –– MM(5) (en rouge)

Les moyennes mobiles pondérées

- Les moyennes mobiles simples sont des moyennes équipondérées (chaque observation a le même poids).
- Les moyennes mobiles centrées (k pair) accordent 2 fois moins de poids aux 2 valeurs extrêmes ($\frac{1}{2}$ aux deux observations extrêmes et 1 aux $n-1$ intermédiaires).
- Plus généralement, on peut définir des moyennes mobiles pondérées par des poids ω_i, par exemple si l'on veut accorder plus d'importance aux observations centrales.

Moyennes mobiles pondérées

La série des moyennes mobiles pondérées d'ordre k avec $k = 2m + 1$, notée $MMP(k)$, est définie par :

$$MMP(k)_i = \sum_{\ell=-m}^{m} \omega_\ell \, y_{i+\ell}$$

pour $i = m + 1, \ldots, n - m$,

où $(\omega_\ell)_{-m \leq \ell \leq m}$ est une suite de poids, c'est à dire une suite de réels positifs satisfaisant :

$$\sum_{\ell=-m}^{m} \omega_\ell = 1.$$

Notation

On adopte souvent la notation suivante pour désigner le type de moyenne mobile considéré : la valeur des pondérations apparaît entre crochet, ie :

$$[\omega_{-m}, \ldots, \omega_0, \ldots, \omega_m].$$

De plus, dans le cas de coefficients égaux, on emploie le symbole de multiplication. Ainsi :

- $MMC(2)$ s'écrit $[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$.
- $MMC(4)$ s'écrit $[\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}]$ ou bien $[\frac{1}{8}, 3 \times \frac{1}{4}, \frac{1}{8}]$.
- $MMC(12)$ s'écrit $[\frac{1}{24}, 11 \times \frac{1}{12}, \frac{1}{24}]$.

On parle de *moyenne mobile symétrique* lorsque $\omega_{-\ell} = \omega_\ell$ pour tout $\ell = 1, \ldots, m$.

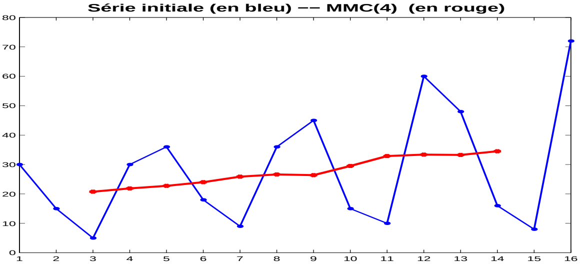
Les moyennes mobiles centrées sont symétriques.

Si la série chronologique $(y_i)_{i=1,\ldots,n}$ possède une composante saisonnière de période p, alors l'application d'une moyenne mobile d'ordre p supprime cette saisonnalité.

La série $MM(p)$ ou $MMC(p)$ ne possède plus de composante saisonnière de période p.

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

Exemple : Ventes trimestrielles d'un produit

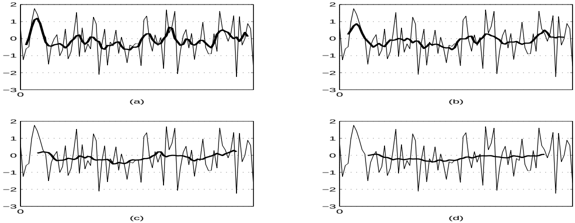


Atténuation des fluctuations irrégulières

Par construction, une moyenne mobile consiste à faire des moyennes partielles de proche en proche. On obtient donc un **lissage** de la série.

- Une moyenne mobile atténue l'amplitude des fluctuations irrégulières d'une chronique.
- Plus l'ordre de la moyenne mobile est élevé, plus cette atténuation est importante.

Exemple Moyennes mobiles sur une série de fluctuations irrégulières



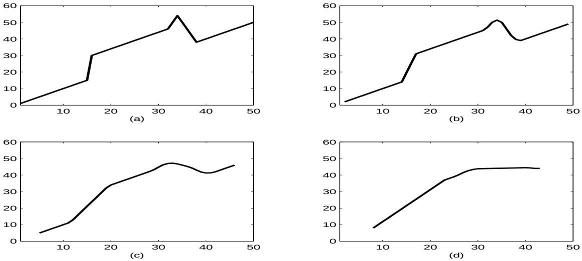
La série des fluctuations irrégulières, avec en

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

Effet d'une moyenne mobile sur la tendance

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

Exemple : Application de moyennes mobiles sur une tendance non régulière



- **(a)** : Tendance
- **(b)** : $MM(3)$
- **(c)** : $MM(8)$
- **(d)** : $MM(15)$

Résumé

Un **lissage par moyenne mobile** va permettre :

- De faire apparaître l'**allure de la tendance** (si celle-ci n'a pas une forte courbure).
- D'**éliminer la composante saisonnière** (dans la mesure où celle-ci est rigoureusement périodique).
- D'**atténuer les fluctuations irrégulières**.

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 :

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

Remarques

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 :

- On supprime la composante périodique de période p avec une moyenne mobile d'ordre p.
- On réduit l'amplitude des fluctuations irrégulières avec une moyenne mobile d'ordre élevé.
- Mais les détails de la tendance disparaîtront d'autant plus que l'ordre de la moyenne mobile sera grand.

Remarques

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

Une méthodologie consiste donc :

- A appliquer $MM(p)$ ou $MMC(p)$ pour supprimer la composante périodique de période p 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.