

Introduction aux Séries Chronologiques

Master IDESSE 2^e année

Université de Savoie



Novembre–Décembre 2010

Première partie I

Techniques Descriptives

Techniques Descriptives

Introduction aux Séries Chronologiques

Introduction

Estimation de la tendance et de la saisonnalité

- Tendance et Saisonnalité

- Moyenne Mobile (Moving Average)

- Tendances et saisonnalité via moyennes mobiles

Filtrage de la tendance et de la saisonnalité

- Notations

- Filtrage

Série chronologique

Définition

Une série chronologique ou série temporelle, « time serie » en anglais, est une série d'observations ordonnées dans le temps.

- Les séries chronologiques se retrouvent un peu partout :
 - ★ en économie : taux de chômage, PNB
 - ★ en finance : cours d'action, taux d'intérêt, ...
 - ★ en écologie : pollution, taux de CO₂, ...
 - ★ dans les transports, en démographie
- Plusieurs objectifs : prévision, détection de tendance, ... ,

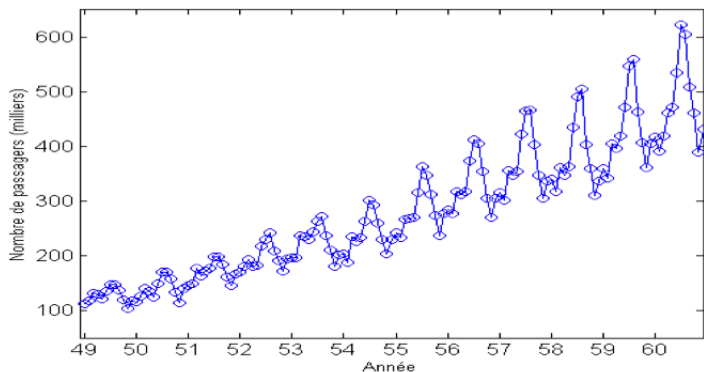
Trafic aérien international

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

○○○
○○
○○
○○○○○

○○○○○
○○○○○○○

Trafic aérien international



Constats

- Augmentation régulière du trafic ;
- Fluctuation saisonnière :
 - ★ Creux en avril ;
 - ★ Pic en juillet août ;
- Données de plus en plus dispersées !
- Les deux premières observations relèvent de la partie déterministe de la série
- La dernière de la composante aléatoire de la série

Réduction du bruit

- Nous noterons dans la suite X_t la série chronologique à étudier
 - ★ Transformation visant à réduire le bruit : en général

$$Y_t = \ln(X_t).$$

- ★ On peut aussi utiliser l'une des fonctions

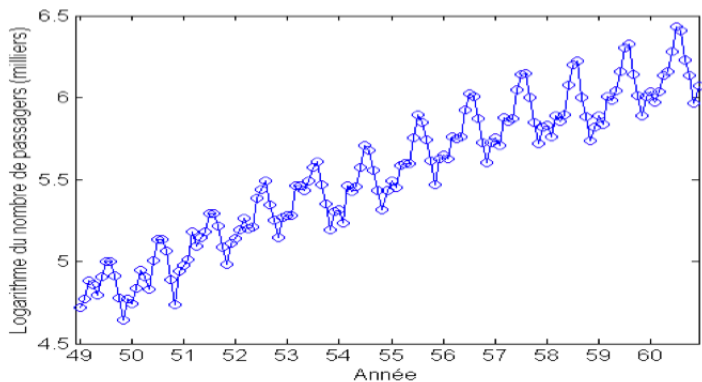
$$g^\lambda(x) = \frac{x^\lambda - 1}{\lambda}, \quad g(x) = \ln[x/(1 - x)]$$

- $Y_{t+1} - Y_t = \ln\left(1 + \frac{X_{t+1} - X_t}{X_t}\right) \simeq \frac{X_{t+1} - X_t}{X_t}$
 - ★ variation de Y = variation de X en pourcentage

○○○
○○
○
○○○○○

○○○○
○○○○○○○

Trafic aérien avec transformation logarithmique





Tendance et Saisonnalité

- Pas facile de faire la distinction entre les deux !
- Tendance = évolution au cours du temps indépendamment des fluctuations saisonnières
- Saisonnalité = variations saisonnières pures
- Exemple idéal : $Y_t = t/2 + 10 \cos(\pi t/6)$
 - ★ On peut facilement isoler les deux parties
 - ★ Tendance $T_t = t/2$
 - ★ Saisonnalité $S_t = 10 \cos(12t)$ (périodicité mensuelle)

Tendance et saisonnalité

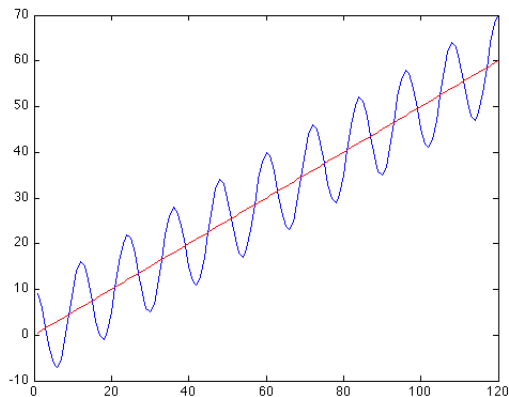


FIGURE: En rouge la tendance pour la série des logarithmes Y_t



Modèle multiplicatif

- $X_t = T_t \times S_t \times U_t$: T_t tendance, S_t saisonnalité, U_t bruit
- Y_t suit un modèle additif : $Y_t = \log(T_t) + \log(S_t) + \log(U_t)$
 - ★ Plus facile de travailler sur le modèle additif
- Si on connaît T_t , S_t et U_t s'interprètent comme des indices
 - ★ S_t est l'indice saisonnier
 - ★ U_t indice aléatoire : fluctuation inévitable
 - ★ on peut les exprimer en pourcentages



Moyenne Mobile

- Il s'agit de combinaison pondérées de valeurs de la séries :

$$M(X)_t = \sum_{i=-h}^k \theta_i X_{t+i}, \quad h \in \mathbb{N}, \quad k \in \mathbb{N}.$$

- ★ $h + k$ s'appelle le degré de la moyenne mobile
- ★ $h + k + 1$ est l'ordre de la moyenne mobile
- ★ Lorsque $h = k$, la moyenne mobile est dite centrée
- ★ Si de plus, $\theta_i = \theta_{-i}$, la moyenne mobile est dite symétrique
- Voici quelques exemples :
 - ★ $M(X)_t = (X_t + X_{t+1})/2$ n'est pas centrée (donc pas symétrique)
 - ★ $M(X)_t = (X_{t-1} + X_t + X_{t+1})/3$ est centrée et symétrique



Moyenne mobile

- Pour h impair, $h = 2p + 1$, on considère les moyennes mobiles

$$M_h(X) = \frac{1}{h} (X_{t-p} + \dots + X_t + \dots + X_{t+p})$$

- ★ Elles sont centrées et symétriques
- Pour $h = 2p$, on peut construire des moyennes centrées et symétriques d'ordre pair

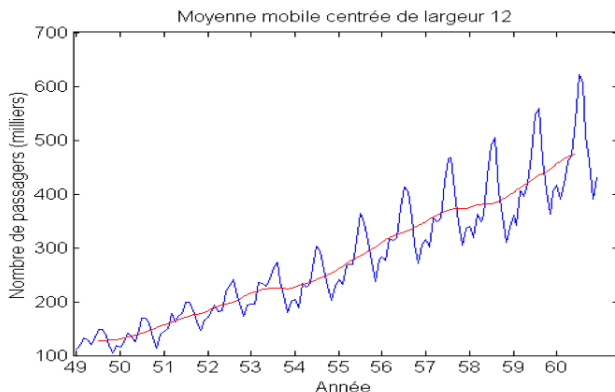
$$\begin{aligned} M_h(X) &= \frac{1}{h} (X_{t-p+1/2} + \dots + X_{t-1/2} + X_{t+1/2} + \dots + X_{t+p-1/2}) \\ &= \frac{1}{h} \left(\frac{1}{2} (X_{t-p} + X_{t-p+1}) + \dots + \frac{1}{2} (X_{t+p-1} + X_{t+p}) \right) \\ &= \frac{1}{h} \left(\frac{1}{2} X_{t-p} + X_{t-p+1} + \dots + X_t + \dots + X_{t+p-1} + \frac{1}{2} X_{t+p} \right) \end{aligned}$$

- ★ L'ordre est en fait $h + 1$!

Estimation de la tendance T_t

- On utilise la moyenne mobile d'ordre 2×6 :

$$T_t = \frac{1}{12} \left(\frac{X_{t-6}}{2} + X_{t-5} + \dots + X_{t+5} + \frac{X_{t+6}}{2} \right)$$

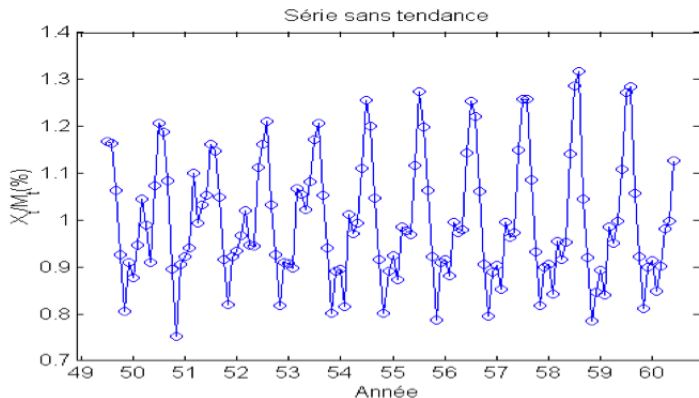


○○○
○○
○●○○○

○○○○○
○○○○○○○

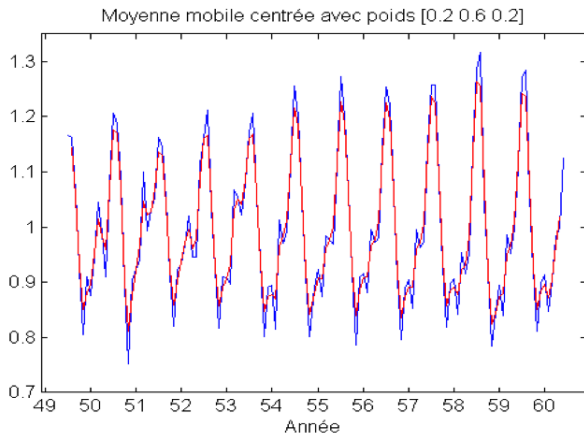
Série sans tendance

- On trace le rapport X_t/T_t



Saisonnalité

- On fait une moyenne mobile centrée symétrique sur la série obtenue avec poids 0.2, 0.6 et 0.2

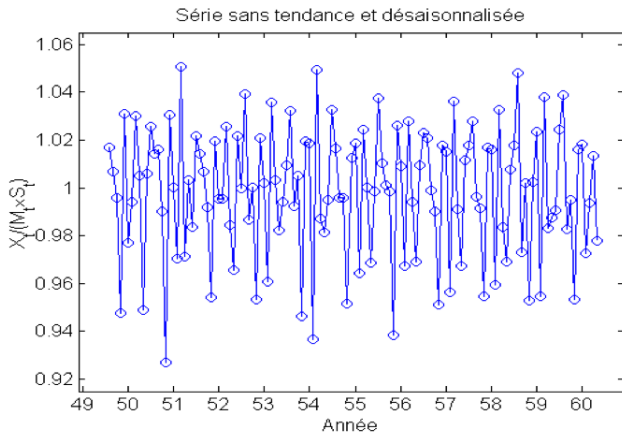


○○○
○○○
○○○●○○

○○○○○
○○○○○○○○

Saisonnalité

- On ne regarde que le bruit

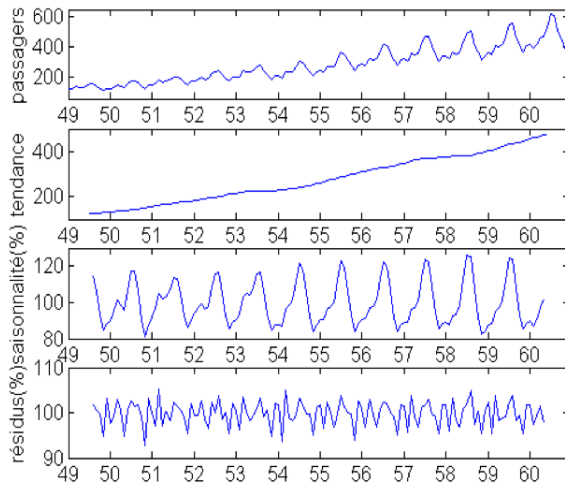


○○○
○○
○○
○○○○●○

○○○○○
○○○○○○○

Bilan

Trafic aérien international de 1949 à 1960



Techniques Descriptives

Introduction aux Séries Chronologiques

Introduction

Estimation de la tendance et de la saisonnalité

- Tendance et Saisonnalité

- Moyenne Mobile (Moving Average)

- Tendances et saisonnalité via moyennes mobiles

Filtrage de la tendance et de la saisonnalité

- Notations

- Filtrage

Motivations

- Parfois difficile de prédire la tendance et la saisonnalité !
 - ★ Une erreur sur le tendance entraîne une erreur sur la saisonnalité
 - ★ Une erreur sur la saisonnalité entraîne une erreur sur le bruit
- On cherche à éliminer la tendance et la saisonnalité directement sans les estimer
 - ★ Pour obtenir le terme aléatoire directement
- On va **filtrer** la tendance et la saisonnalité



Opérateur retard

- On considère une série chronologique $X = \{X_t\}_t$

Définition

On appelle opérateur **retard**, noté B , l'opérateur

$$B : X = \{X_t\}_t \longmapsto BX = \{X_{t-1}\}_t, \quad BX_t = X_{t-1}.$$

On appelle opérateur **avance**, noté F , l'opérateur

$$F : X = \{X_t\}_t \longmapsto FX = \{X_{t+1}\}_t, \quad FX_t = X_{t+1}.$$

- ★ La notation B vient de « backward » ; F de « forward »
- ★ On utilise aussi la notation L pour « lag » à la place de B
- On utilisera seulement l'opérateur B car F est l'opérateur inverse

$$F = B^{-1}, \quad \text{comme } x \mapsto x^2 \text{ et } x \mapsto \sqrt{x}.$$

Opérateur retard itéré

- On note Id l'opérateur identité : $\text{Id } X_t = X_t$

★ $B \circ F = F \circ B = \text{Id}$

- On peut itérer l'opérateur B

$$B^2 = B \circ B, \quad B^2 X_t = B(BX)_t = BX_{t-1} = X_{t-2}$$

- Plus généralement, si $k \in \mathbb{N}$,

$$B^k = \underbrace{B \circ B \circ \dots \circ B}_{k \text{ fois}},$$

$$B^k X_t = X_{t-k},$$

$$B^{-k} = \underbrace{B^{-1} \circ B^{-1} \circ \dots \circ B^{-1}}_{k \text{ fois}},$$

$$B^{-k} X_t = X_{t+k}$$

- ★ Convention habituelle : $B^0 = \text{Id}$

Polynômes d'opérateurs B

- Soit P le polynôme

$$P(z) = a_0 + a_1 z^1 + \dots + a_n z^n$$

- On peut considérer l'opérateur $P(B)$

$$P(B) = a_0 B^0 + a_1 B^1 + \dots + a_n B^n,$$

$$P(B)X_t = a_0 X_t + a_1 X_{t-1} + \dots + a_n X_{t-n}$$

- ★ Les opérations usuelles sur les polynômes sont préservées
- On peut aussi définir des « séries » de l'opérateur B
- On peut faire pareil avec B^{-1}

Retour sur les moyennes mobiles

- Moyenne mobile générale :

$$M(X)_t = \sum_{i=-h}^k \theta_i X_{t+i}, \quad h \in \mathbb{N}, \quad k \in \mathbb{N}.$$

- ★ $h + k$ s'appelle le degré de la moyenne mobile
- ★ $h + k + 1$ est l'ordre de la moyenne mobile
- On peut réécrire cette moyenne mobile sous la forme

$$M(X)_t = \sum_{i=-h}^k \theta_i B^{-i} X_t = \left(\sum_{i=-h}^k \theta_i B^{-i} \right) X_t$$

- De façon plus synthétique

$$M = \sum_{i=-h}^k \theta_i B^{-i} = B^h \circ \sum_{i=-h}^k \theta_i B^{-i-h} = B^h \sum_{j=0}^{h+k} \theta_{j-h} B^{-j}$$

Retour sur les moyennes mobiles (2)

- On peut aussi écrire une moyenne mobile avec l'opérateur F

$$M = B^h \sum_{j=0}^{h+k} \theta_{j-h} B^{-j} = F^{-h} \sum_{j=0}^{h+k} \theta_{j-h} F^j = B^h \sum_{j=0}^{h+k} \theta_{j-h} F^j$$

- Par exemple, $M(X)_t = (X_t + X_{t+1})/2$ peut se représenter par

$$M = (\text{Id} + F) / 2 = (\text{Id} + B^{-1}) / 2$$

- $M(X)_t = (X_{t-1} + X_t + X_{t+1}) / 3$ par

$$M = (B^1 + \text{Id} + B^{-1}) / 3 = B (\text{Id} + B^{-1} + B^{-2}) / 3.$$

Filtrage de la tendance

- Le filtrage est efficace sur un modèle additif

$$Y_t = T_t^a + S_t^a + U_t^a$$

- Si on part d'un modèle multiplicatif : $Y_t = \log X_t$
- Pour éliminer la tendance, on fait une « dérivation »

Définition

L'opérateur de **différentiation** est l'opérateur noté ∇ ou Δ défini par

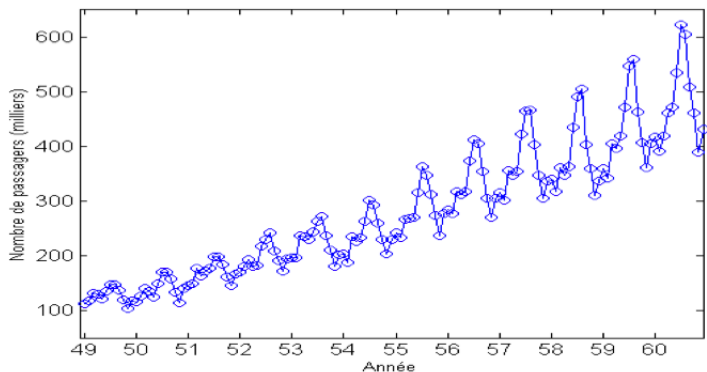
$$\nabla Y_t = \Delta Y_t = Y_t - Y_{t-1} = (\text{Id} - B) Y_t$$

- On filtre la tendance en considérant la série différenciée $Z_t = Y_t - Y_{t-1}$
 - ★ Très efficace si la tendance est linéaire

○○○
○○
○○
○○○○○

○○○○○
●○○○○○○

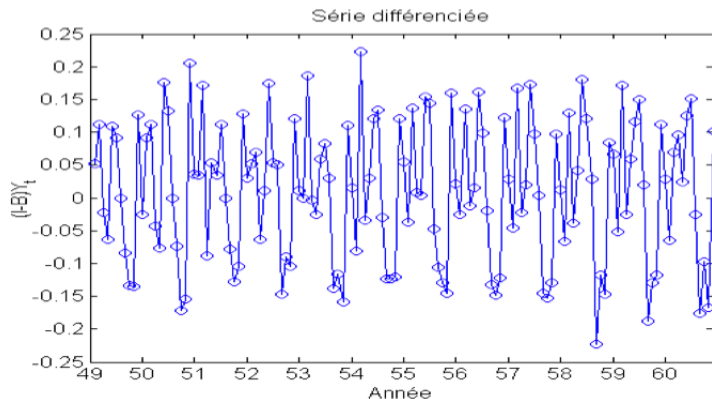
Trafic aérien (le retour)



○○○
○○
○○○○○○

○○○○○
○○●○○○○○

Filtrage de la tendance : trafic aérien



Cas linéaire

- Modèle sans bruit avec tendance linéaire
- $Y_t = a \times t + b + c \cos(\pi t/6)$
 - ★ tendance linéaire : $a \times t + b$
 - ★ terme saisonnier : $c \cos(\pi t/6)$
- $Z_t = Y_t - Y_{t-1} = a + c \cos(\pi t/6) - c \cos(\pi(t-1)/6)$
 - ★ plus de terme en t !
 - ★ seulement la constante a et un terme saisonnier (12-périodique)

Que faire si on a toujours une tendance ?

- L'idée est d'itérer l'opérateur de différentiation
- Par exemple $\nabla^2 = \nabla \circ \nabla$:

$$\begin{aligned}\nabla^2 Y_t &= \nabla (\nabla Y)_t = \nabla (Y_t - Y_{t-1}) = \nabla Y_t - \nabla Y_{t-1} \\ &= Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-1-1}) = Y_t - 2Y_{t-1} + Y_{t-2} \\ &= (\text{Id} - B)^2 Y_t\end{aligned}$$

- ★ Calculer $\nabla^2 t^2$ et $\nabla^2 t$
- Pour k entier, on peut appliquer le filtre $\nabla^k = (\text{Id} - B)^k$
 - ★ il transforme les polynômes de degré inférieur à k en constante
 - ★ élimine les tendances polynomiales

Filtrage de la saisonnalité

- Cas d'une série périodique $Y_t = \cos(\pi t/6)$
 - ★ $Y_{t-12} = Y_t$ pour tout t
 - ★ L'opérateur $\nabla_{12} : \nabla_{12} Y_t = Y_t - Y_{t-12}$ élimine une saisonnalité de période 12

Définition

On appelle opérateur de différentiation saisonnière de période s l'opérateur noté ∇_s défini par

$$\nabla_s Y_t = (\text{Id} - B^s) Y_t = Y_t - Y_{t-s}$$

- Les opérateurs ∇ et ∇_s commutent

$$\nabla \nabla_s Y_t = \nabla_s \nabla Y_t$$

Filtrage de la saisonnalité

- En pratique on choisit s en fonction des données
 - ★ $s = 12$ pour des données mensuelles
 - ★ $s = 7$ pour des données journalières
- Comme les opérateurs ∇ et ∇_s commutent, on peut éliminer tendance et saisonnalité dans un sens ou dans l'autre
- Après application de ces deux filtres, on doit obtenir une série complètement aléatoire

○○○
○○
○○
○○○○○○○○○○○
○○○○○○○●

Trafic aérien (le retour du retour)

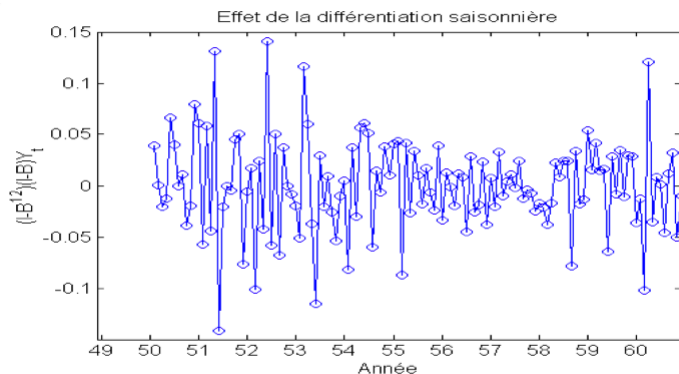


FIGURE: Différenciation de période 12 de la série différenciée