Cours SG042

Modélisation et prévision

Séries chronologiques - Séance 1
Décomposition d'une chronique

Frédéric Sur
École des Mines de Nancy

www.loria.fr/∼sur/enseignement/modprev/

Qu'est-ce qu'une série chronologique ?

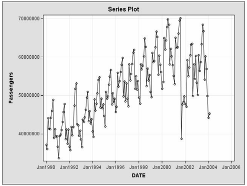
Séries temporelles / Chroniques / *Time Series*
→ suite d'observations d'une grandeur au cours du temps.

Exemples :

- économétrie (taux de chômage),
- finance (cours d'action),
- écologie (pollution),
- démographie (population),
- météorologie (relevé de températures),
- astronomie (fluctuations de la magnitude d'un astre)...

Modèle sous-jacent liant les observations.

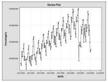
Exemple de série chronologique



Trafic aérien aux USA de 1990 à 2004.

But du cours : décrire l'évolution des chroniques à l'aide de modèles basés sur des propriétés statistiques.

Dans le modèle...



- **Perturbations ponctuelles** : variations forte amplitude. *(grève, krach boursier, attentats, erreur de mesure...)*
  → À traiter en premier. Cf "modèles d'intervention".
- **Tendance $\mathbf{T_t}$** : évolution globale.
- **Variations saisonnières $\mathbf{S_t}$** : fluctuations périodiques. *Cf "données corrigées des variations saisonnières".*
  → moyenne nulle sur une période.
- **Composante aléatoire $\mathbf{u_t}$** : fluctuations irrégulières imprévisibles de faible amplitude.
  → moyenne nulle.

Pourquoi étudier les chroniques ?

- Description / explication d'un phénomène.

chômage : variation tendancielle ou fluctuation saisonnière ?

- Prévision.

consommation d'électricité, démographie, restauration rapide...

- Étude de la dynamique.

cours d'actions.

- Impact d'un événement sur une variable.

sécurité routière : impact d'une nouvelle loi.

SG042 : techniques statistiques, aspects *temporels*.

Autres points de vue : automatique / théorie du contrôle, traitement du signal.
→ cf cours électifs.

À faire avant chaque séance...

Polycopié du cours :
chaque chapitre correspond à une séance

⟶ à lire avant le cours !
("pour en savoir plus" optionnel)

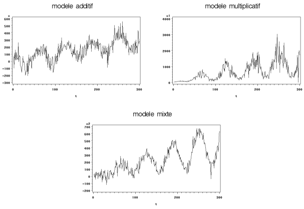
Pour approfondir : bibliographie du polycopié disponible à la bibliothèque.

Séance 1

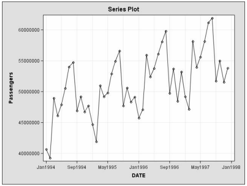
1. Introduction
2. Modèles de décomposition
   - Tendance
   - Composante saisonnière
   - Composante aléatoire
3. Estimation des paramètres et décomposition
   - Moindres carrés
   - Filtrage
4. Conclusion

Modèles de décomposition

- Modèle additif : $X_t = T_t + S_t + u_t$
- Modèle multiplicatif : $X_t = T_t \cdot S_t \cdot u_t$
- Modèle mixte : $X_t = T_t \cdot S_t + u_t$



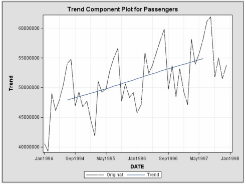
Exemple de décomposition



Trafic aérien aux USA de 1994 à 1997.

→ modèle multiplicatif.

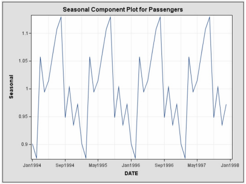
Exemple de décomposition



Tendance T_t.

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

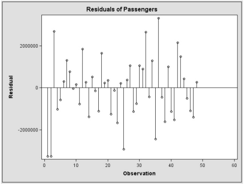
Exemple de décomposition



Composante saisonnière S_t.
(modèle multiplicatif, donc S_t centré sur 1 et pas 0)

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Exemple de décomposition



Résidus u_t.
(pas de structure apparente)

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Composante tendancielle T_t

Exemples de composante tendancielle T_t :

- linéaire : $T_t = at + b$
- quadratique : $T_t = at^2 + bt + c$
- exponentielle : $T_t = T_0 e^{at}$
- ...

avec les paramètres $a, b, c, T_0 \ldots$ à déterminer.

Composante saisonnière S_t

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

S_t périodique, de période p.

$\longrightarrow$ éventuellement plusieurs composantes périodiques superposées.

Exemple : comp. annuelle + comp. trimestrielle...

Question : comment trouver les périodes ?

Comment trouver la (les) période(s) ?

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Transformée de Fourier Discrète pour un signal *stationnaire*

Pour une chronique $X_0, X_1, \ldots, X_{T-1}$, TFD :

$$\forall k \in \{0, 1, \ldots, T-1\},\ \widehat{X}_k = \frac{1}{T} \sum_{t=0}^{T-1} X_t e^{-2i\pi kt/T}$$

(signal réel, donc $\widehat{X}_1 = \overline{\widehat{X}_{T-1}}, \widehat{X}_2 = \overline{\widehat{X}_{T-2}} \ldots$)

TFD inverse : (formule de reconstruction)

$$\forall t \in \{0, 1, \ldots, T-1\},\ X_t = \sum_{k=0}^{T-1} \widehat{X}_k e^{2i\pi kt/T}$$

Remarque : les coefficients $\widehat{X}_k$ et $\widehat{X}_{T-k}$ correspondent à la composante de fréquence k/T (ou période T/k).

Le périodogramme

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Définition : périodogramme

$$\forall\ 0 < k < T/2 + 1,\ J_{T/k} = T \left| \widehat{X}_k \right|^2$$

est l'amplitude de la composante de période $\frac{T}{k}$ (fréquence $\frac{k}{T}$).
Le graphe de J est appelé *périodogramme*.

$\longrightarrow$ les "pics" dans le périodogramme permettent d'identifier les composantes saisonnières.

Attention : on peut démontrer que les $(\widehat{X}_k)$ on tendance à décroître (donc le périodogramme à croître). Donc on ne regarde que les pics sur le "début" du périodogramme.

Composante aléatoire u_t

$\rightarrow$ pas de tendance ou de phénomène périodique superposé.
Modélisation : (u_t) réalisation d'un processus aléatoire (ε_t) *stationnaire.*

Définition **: (ε_t) stationnaire (au second ordre)**

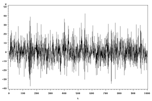
- $\forall t,\ \mathbb{E}(\varepsilon_t) = m$ (moyenne constante)
- $\forall t, s,\ \mathrm{cov}(\varepsilon_t, \varepsilon_s) = \gamma(|t-s|)$ (covariance symétrique, invariante par translation)

En particulier $\forall t$, $\mathrm{var}(\varepsilon_t) = \gamma(0)$ (variance constante).

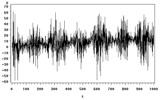
Cas particuliers de processus stationnaires :
Bruit blanc faible : $m = 0$ et $\forall h > 0,\ \gamma(h) = 0$.
Bruit blanc fort : $m = 0$ et (ε_t) i.i.d.
Bruit blanc gaussien : (ε_t) i.i.d. $\sim \mathcal{N}(0, \sigma^2)$.
(cf résidus dans la régression)

Exemples

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion



bruit blanc gaussien



processus non stationnaire

Caractérisation des processus stationnaires

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Processus stationnaires considérés : $X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$

avec $\sum_{j=0}^{+\infty} |a_j| < +\infty$ et (ε_t) **bruit blanc** (faible)
(motivation : décomposition de Wold)

Remarque : en fait (ε_t) sera même généralement un bruit blanc *gaussien.*

Par définition, (X_t) caractérisé par

- moyenne m,
- fonction de covariance γ.

m estimé par la moyenne empirique :

$$\overline{X}_T = \frac{1}{T} \sum_{t=1}^{T} X_t$$

Processus stationnaires : corrélogramme

Modélisation et prévision
F. Sur - ENSMN
Introduction
Modèles de décomposition
Tendance
Composante saisonnière
Composante aléatoire
Estimation des paramètres et décomposition
Moindres carrés
Filtrage
Conclusion

Définition **: fonction d'autocorrélation**

$$\forall h \geqslant 0,\ \rho(h) = \frac{\mathrm{Cov}(X_t, X_{t+h})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t+h})}} = \frac{\gamma(h)}{\gamma(0)}.$$

Graphe de (l'estimation de) $\rho(h)$: *corrélogramme* (**ACF**).

Proposition **: intérêt du corrélogramme**

Si $X_t = m + \sum_{j=0}^{k} a_j \varepsilon_{t-j}$, alors $\forall h > k,\ \rho(h) = 0$.

Preuve : $\mathrm{Cov}(X_t, X_{t+h}) = \sum_{i=0}^{k} \sum_{j=0}^{k} a_i a_j \mathrm{Cov}(\varepsilon_{t-i}, \varepsilon_{t+h-j})$.
Or $\mathrm{Cov}(\varepsilon_{t-i}, \varepsilon_{t+h-j}) = 0$ si $h \neq j - i$ (car (ε_t) b.b.).

Estimation du corrélogramme

Corrélogramme empirique :

$$\widehat{\rho}(h) = \frac{T}{T-h} \frac{\sum_{t=h+1}^{T}(X_t - \overline{X}_T)(X_{t-h} - \overline{X}_T)}{\sum_{t=1}^{T}(X_t - \overline{X}_T)^2}$$

Remarque : $\widehat{\rho}(h)$ calculé pour $h << T$ (il faut suffisamment d'échantillons pour le numérateur), et alors $\frac{T}{T-h} \simeq 1$.

Preuve de convergence : cf poly.

Remarque : le corrélogramme d'une chronique périodique est périodique (même période).

En effet : $\rho(h) = \dfrac{\mathrm{Cov}(X_t, X_{t+h})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t+h})}} = \dfrac{\gamma(h)}{\gamma(0)}$.

Séance 1

1. Estimation de paramètres aux moindres carrés

(X_t) : chronique mensuelle. (trimestrielle dans le poly.)

Modèle de Buys-Ballot :

$$X_t = \underbrace{a_1 + a_2 t}_{T_t} + \underbrace{\sum_{i=1}^{12} b_i \delta_i(t)}_{S_t} + u_i(t)$$

δ_i : indicatrice du mois i.

Estimation des paramètres aux moindres carrés des résidus :

$$\begin{cases} \min\limits_{a_i, b_j} \displaystyle\sum_{t=1}^{T} \left(X_t - a_1 - a_2 t - \sum_{i=1}^{12} b_i \delta_i(t) \right)^2 \\ \text{t.q. } \displaystyle\sum_{i=1}^{12} b_i = 0 \end{cases}$$

Rôle de la contrainte : lever l'indétermination sur les b_i.
(S_t est de moyenne nulle sur la période)

2. Filtrage par moyennes mobiles

***Définition* : moyenne mobile**

Soit (X_t) une chronique, et
$(Y_t) = M(X_t)$ la chronique telle que

$$\forall t,\ Y_t = \sum_{i=-m_1}^{m_2} \theta_i X_{t+i}$$

(Y_t) est obtenue de (X_t) par *filtrage par moyenne mobile*.

Remarque : M linéaire.

Premier objectif : décomposition de $X_t = T_t + S_t + u_t$.
Trouver un filtre tel que

- $M(T_t) = T_t$;
- $M(S_t) = 0$;
- $M(u_t)$ "aussi petit que possible".

Les moyennes mobiles arithmétiques

Filtre symétrique : $m_1 = m_2 \ (= m)$, $\theta_i = \theta_{-i}$
$M(X_t) = \sum_{i=-m}^{m} \theta_i X_{t+i}$.

Propriété

Une moyenne mobile symétrique telle que $\sum_{i=-m}^{m} \theta_i = 1$ conserve les chroniques affines $X_t = at + b$.

Propriété

Parmi ces moy. mob., celles qui "minimisent" $M(u_t)$ avec (u_t) réalisation d'un b.b. faible (ε_t) vérifient : $\forall i, \theta_i = \frac{1}{2m+1}$.

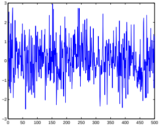
Preuve : $\mathrm{Var}(M(\varepsilon_t)) = \mathrm{Var}(\sum_{i=-m}^{m} \theta_i \varepsilon_{t+i}) = \sigma^2 \sum_{i=-m}^{m} \theta_i^2$,
à minimiser, sous contrainte $\sum_{i=-m}^{m} \theta_i = 1$.

Définition : moyenne mobile arithmétique

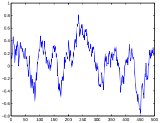
$$\forall i \in \{-m, \ldots, m\},\ \theta_i = \frac{1}{2m+1}.$$

Effet de Yule-Slutsky

Attention à l'apparition éventuelle d'une périodicité "artificielle".



bruit blanc gaussien, $\sigma = 1$



lissage arithmétique, $2m + 1 = 19$

Remarque : la variance du processus est divisée par 19.

Saisonnalité et moyennes mobiles arithmétiques

$$M_{2m+1}(X_t) = \frac{1}{2m+1} \left(X_{t-m} + X_{t-m+1} + \cdots + X_{t+m} \right).$$

Propriété

Les composantes saisonnières S_t de période $2m + 1$ et de moyenne nulle sur une période sont filtrées par M : $M(S_t) = 0$.

Généralisation : composantes saisonnières de période $2m$:

$$M_{2m}(X_t) = \frac{1}{2m} \left(\frac{1}{2} X_{t-m} + X_{t-m+1} + \cdots + X_{t+m-1} + \frac{1}{2} X_{t+m} \right).$$

Discussion détaillée du filtrage : cf traitement du signal.

Récapitulatif

Hypothèses : (X_t) chronique à décomposer sous la forme :

$$X_t = T_t + S_t + u_t,$$

avec T_t linéaire et S_t de période p, moyenne nulle sur la période.

Alors :

- $M_p(X_t) \simeq T_t$.
- $M_k(S_t + u_t) \simeq S_t$ si $k << p$.
- de plus, $(Id - M_p)(Y_t)$ a une moyenne nulle sur une période p.

Un algorithme de décomposition par filtrage

Décomposition de $X_t = T_t + S_t + u_t$.

Étape 0 : périodogramme, connaissances $\rightarrow$ période p de S_t.

1. estimation de la tendance : $T_t = M(X_t)$
2. estimation de $\Sigma_t = S_t + u_t$: $\Sigma_t = X_t - T_t$
3. estimation de la composante saisonnière : $S_t = M'(\Sigma_t)$
4. estimation de la série corrigée des variations saisonnières (cvs) : $X'_t = X_t - S_t$; et $u_t = X'_t - T_t$

où

- M conserve T_t et filtre (élimine) S_t et u_t,
- M' conserve S_t et filtre u_t.

Choix de M : tendance linéaire (par exemple) $\rightarrow M = M_p$.
Choix de M' : $M'(\Sigma_t) = M''(\Sigma_t) - M_p(M''(\Sigma_t))$
avec M'' moy. mob. de support "petit" (filtre u_t)
composée avec $\mathrm{Id} - M_p$ pour que la moyenne sur une période soit nulle.

Remarque : c'est l'idée du programme **Census X11**.

Séance 1

Conclusion

Décomposition (additive / multiplicative / mixte) d'une chronique :

- tendance,
- composante saisonnière (période ?),
- composante aléatoire (processus stationnaire).

En pratique :

- estimation aux moindres carrés d'un modèle paramétrique (*table de Buys-Ballot*)
- ou filtrage (*Census X11*).