Séries Temporelles - cours 1

RCP217

Clément Rambour

Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Déroulé des deux séances

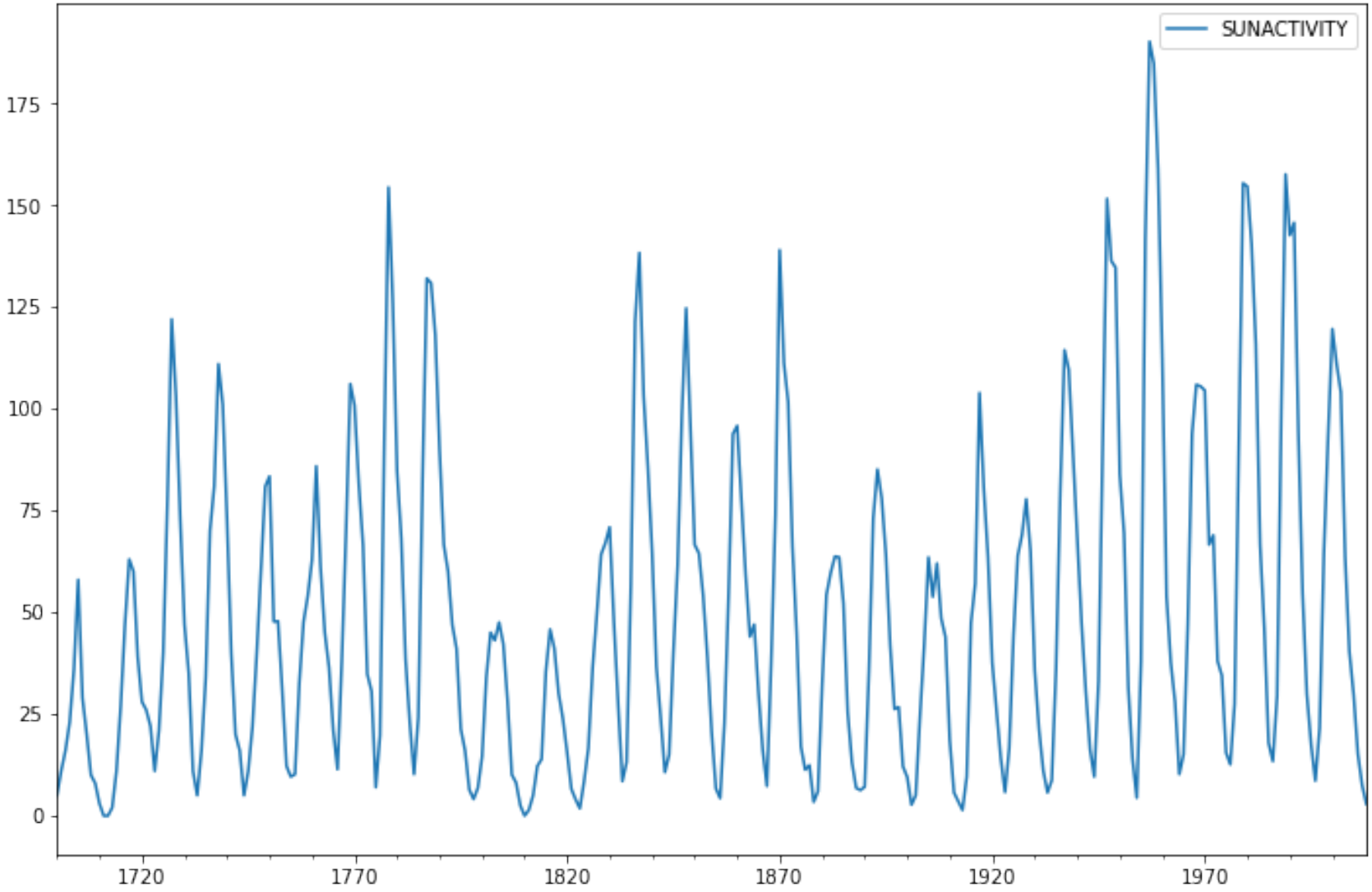
Cours 1 : Introduction aux Séries Temporelles

- Définitions de base
- Modélisation statistique
- Introduction à la prédiction *forecasting*

Cours 2 : Prédiction bayésienne et apprentissage

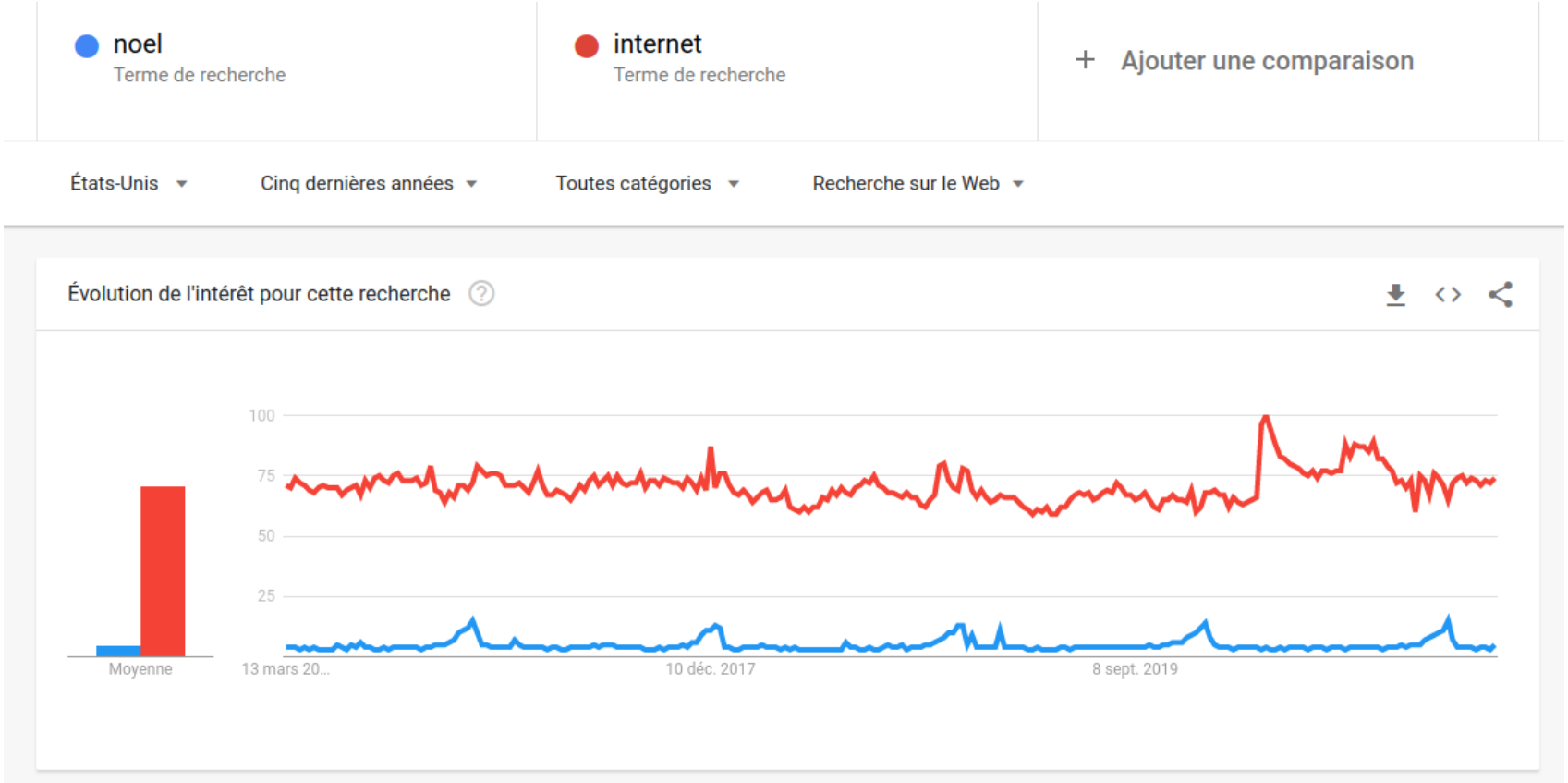
- Filtres de Kalman
- Deep Learning pour les séries temporelles
- Loss pour les séries temporelles
- prédiction statistique

Exemples : activité solaire

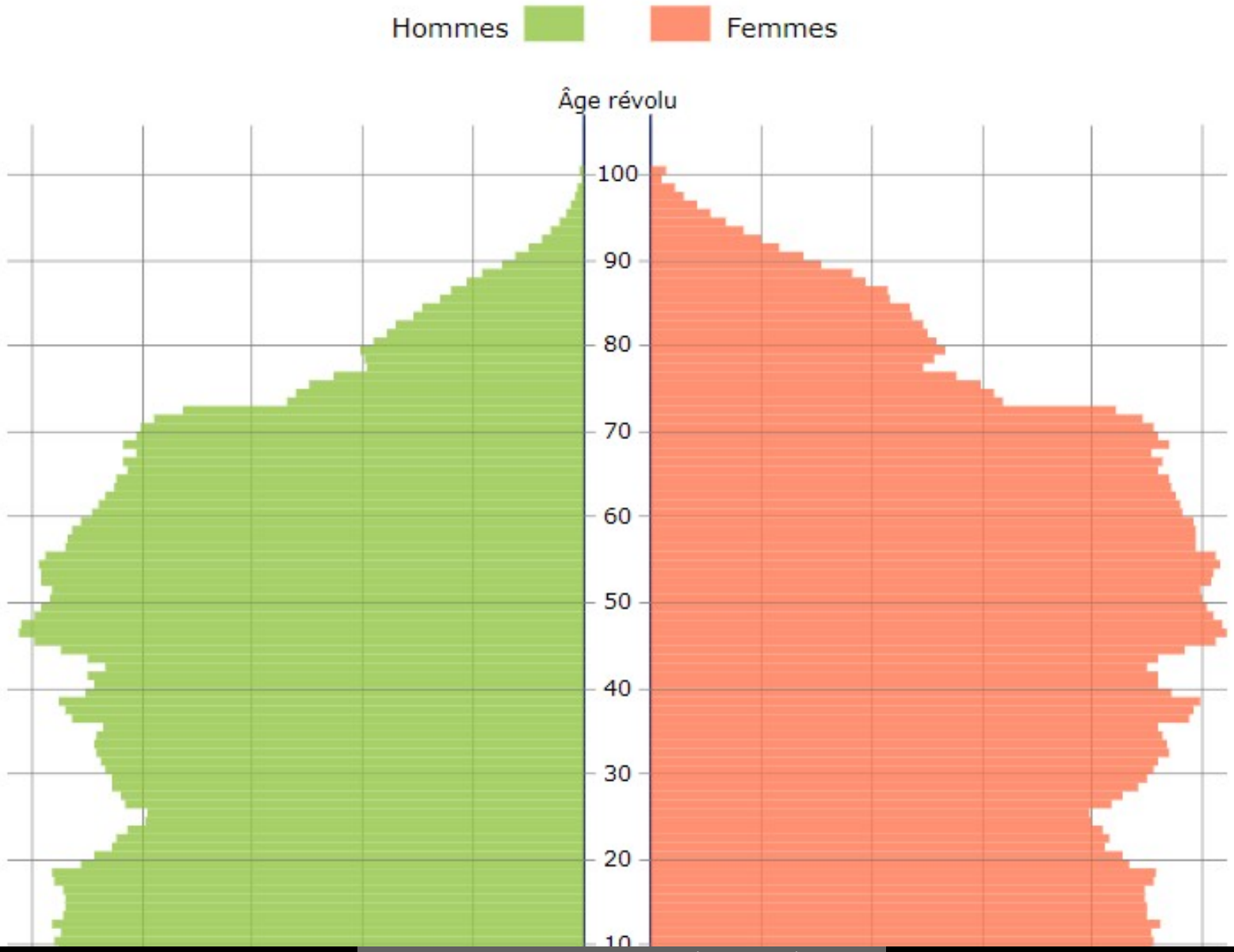


Exemples : cours en bourse

Exemples : recherches sur Google



Exemples : pyramide des âges



Série Temporelle

Définition

Une série temporelle est une ensemble d'observations x_t associées à un ensemble d'indices ordonné $\mathcal{T}$

- Généralement l'ensemble des indices est à valeur discrète : $\mathcal{T} \subseteq \mathbb{N}$ ou $\mathbb{Z}$
- x_t peut être scalaire ou multidimensionnel

Analyse de séries temporelles

Objectifs

- Visualiser et décomposer les composantes de la série :
  - Comprendre le comportement temporel des données
  - Identifier la distribution et la nature du processus générant les données
- Estimer les éléments x_t d'une séries
  - Filtrage
  - Prédiction
- Classifier différentes séries
- Détecter des anomalies

Exemple

Une série temporelle simple peut être de la forme :

$$x_t = s_t + n_t$$

où s_t est déterministe = signal et n_t est aléatoire

- Peut être estimé par moindres carrées si les caractéristiques du bruit sont constantes :

$$\hat{s}_t = \underset{s_t}{\operatorname{argmin}} ||x_t - s_t||_2^2$$

Exemple

Une série temporelle simple peut être de la forme :

$$x_t = s_t + n_t$$

où s_t est déterministe = signal et n_t est aléatoire

- Peut être estimé par moindres carrées si les caractéristiques du bruit sont constantes :

$$\hat{s}_t = \underset{s_t}{\operatorname{argmin}} ||x_t - s_t||_2^2$$

- Si s_t varie "lentement" $\rightarrow$ **tendance**
- Si s_t périodique $\rightarrow$ **composante saisonnière**
- $n_t \rightarrow$ **bruit**

Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Analyse de séries temporelles vs. régression

Particularités

- Les même données peuvent être présentées différemment selon le lieu et l'époque
- Les données sont (très) corrélées
  - Le passé permet de prédire le future
  - Les données ne sont **pas iid**
- Les structures peuvent être visibles sur des axes temporelles très différents : hautes/basses fréquences, changement de régime...

Stationnarité

Comment estimer un signal qui varie dans le temps ?

- Nécessité d'une hypothèse sur la distribution
- La distribution des x_t est la même pour tout $\forall t \rightarrow$ **stationnarité forte**
- La moyenne et la variance sont invariants dans le temps $\rightarrow$ **stationnarité au sens large**

Stationnarité

Définition

$\{x_t\}_{t\in\mathbb{Z}}$ est fortement stationnaire si $\forall h \in \mathbb{Z}, n \in \mathbb{N}$, $(x_{t_1}, \cdots, x_{t_n})$ et $(x_{t_1+h}, \cdots, x_{t_n+h})$ sont distribués selon la même loi.

Cette contrainte est très (trop !) forte, on lui préfère souvent la stationnarité au sens large :

Définition

$\{x_t\}_{t\in\mathbb{Z}}$ est stationnaire (au sens large) si sa moyenne et sa fonction d'autocovariance sont invariante dans le temps.

Moyenne et covariance

Moyenne

La moyenne de $\{x_t\}_{t\in\mathbb{N}}$ en $t \in \mathbb{Z}$ est donnée par

$$\mu_x(t) = \mathbb{E}(x_t)$$

Covariance et Autocovariance

La covariance de $\{x_t\}_{t\in\mathbb{N}}$ et $\{y_t\}_{t\in\mathbb{N}}$ est donnée par

$$\gamma_{xy}(s,t) = \mathbb{E}((x_s - \mathbb{E}(x_s))(y_t - \mathbb{E}(y_t)))$$

L'autocovariance d'un processus stationnaire $\{x_t\}_{t\in\mathbb{N}}$ est donnée par

$$\gamma_{xx}(h) = \mathbb{E}((x_t - \mathbb{E}(x_t))(x_{t+h} - \mathbb{E}(x_{t+h})))$$

Moyenne et covariance

Estimation

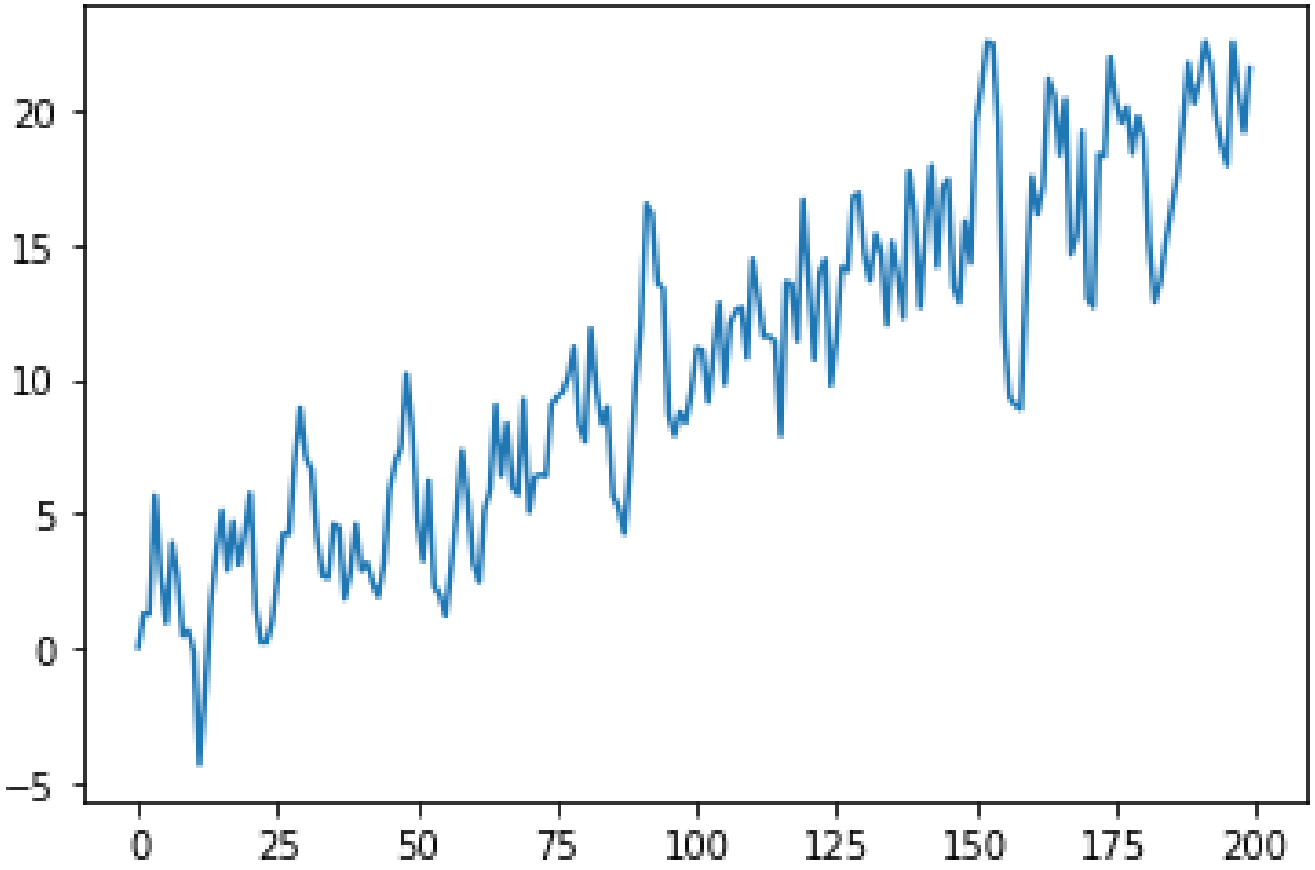
La moyenne et la covariance sont généralement estimées sur un ensemble de N échantillons. On note :

- $\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$
- $\overline{y} = \frac{1}{N}\sum_{n=1}^{N} y_n$

La covariance empirique est calculée comme :

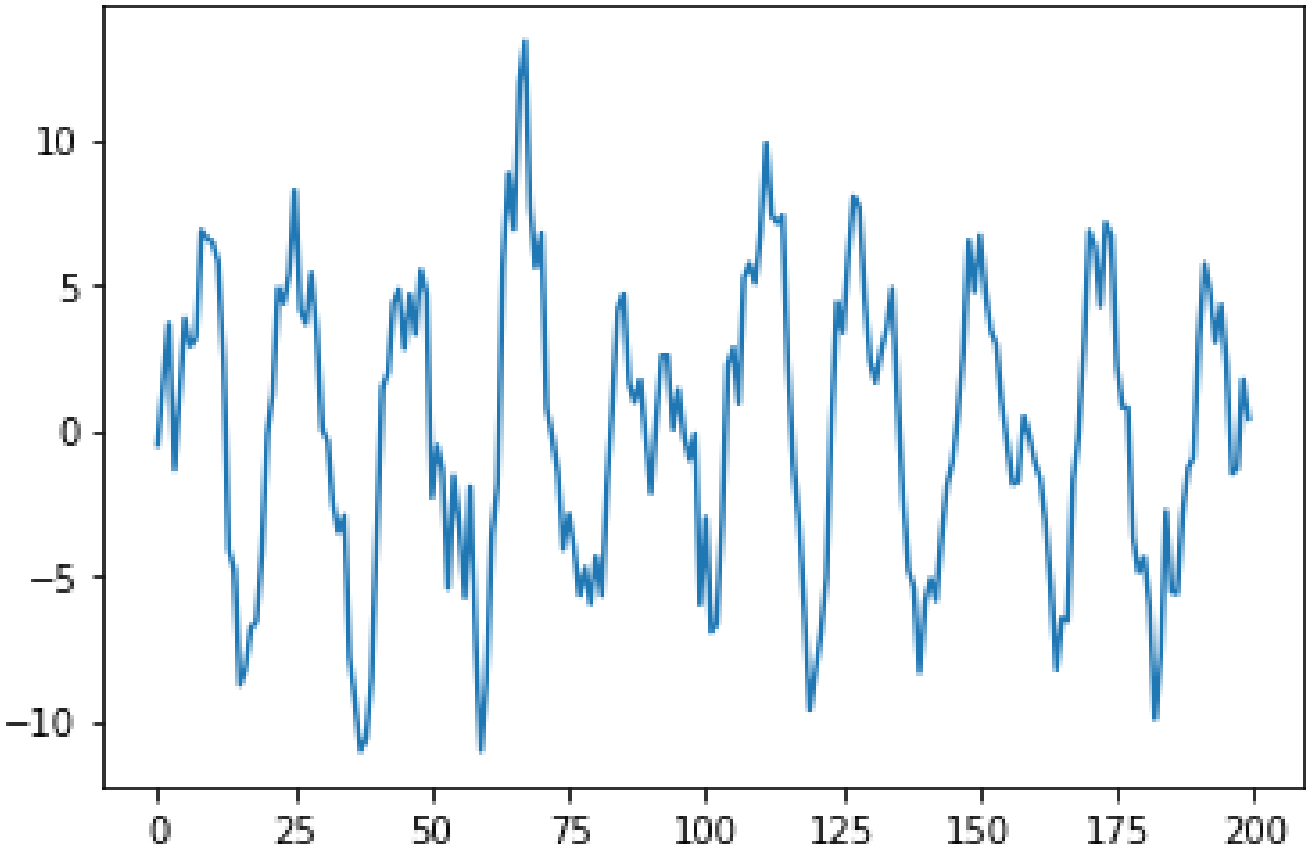
$$\hat{\gamma}_{xy}(h) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})(y_{n+h} - \overline{y}).$$

Exemples



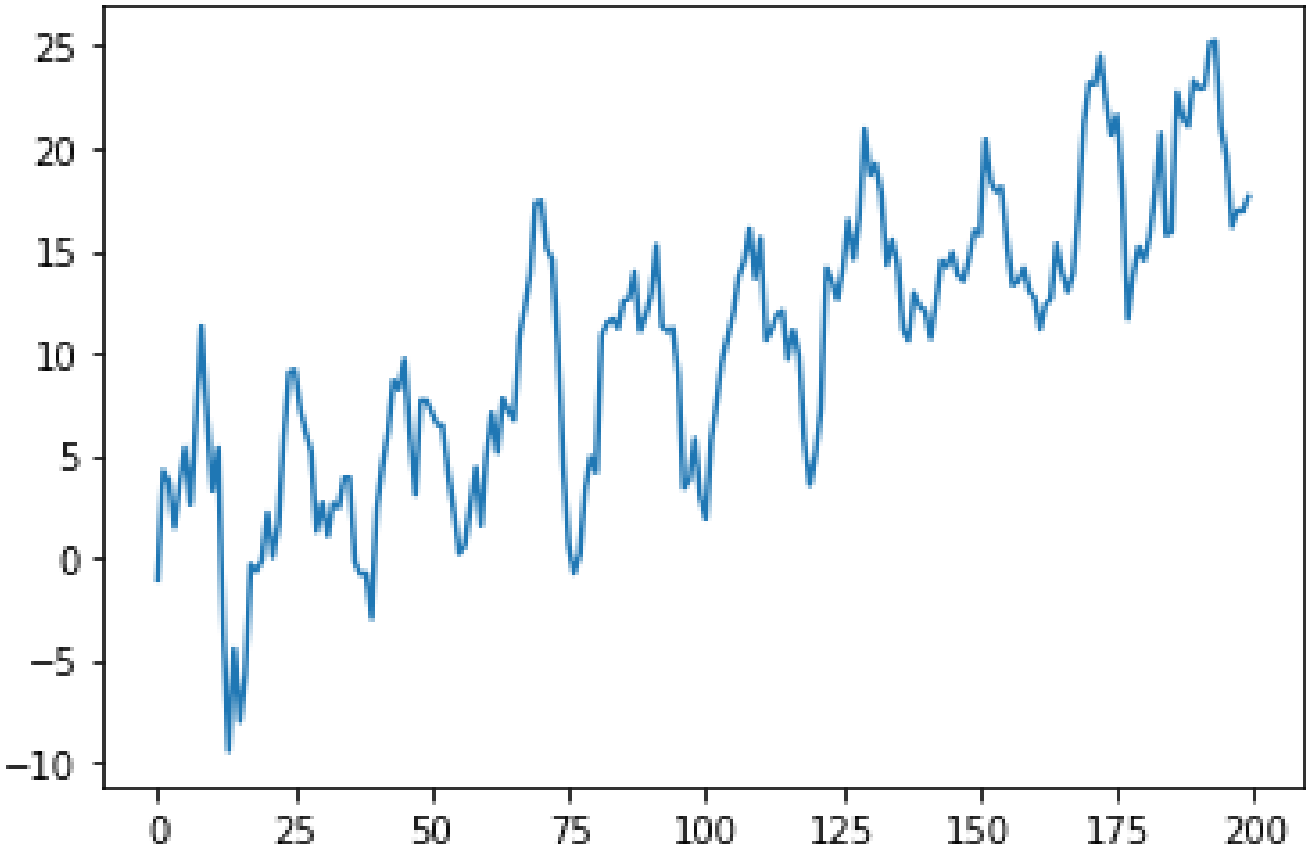
$\Rightarrow$ Non stationnaire car tendance croissante

Exemples



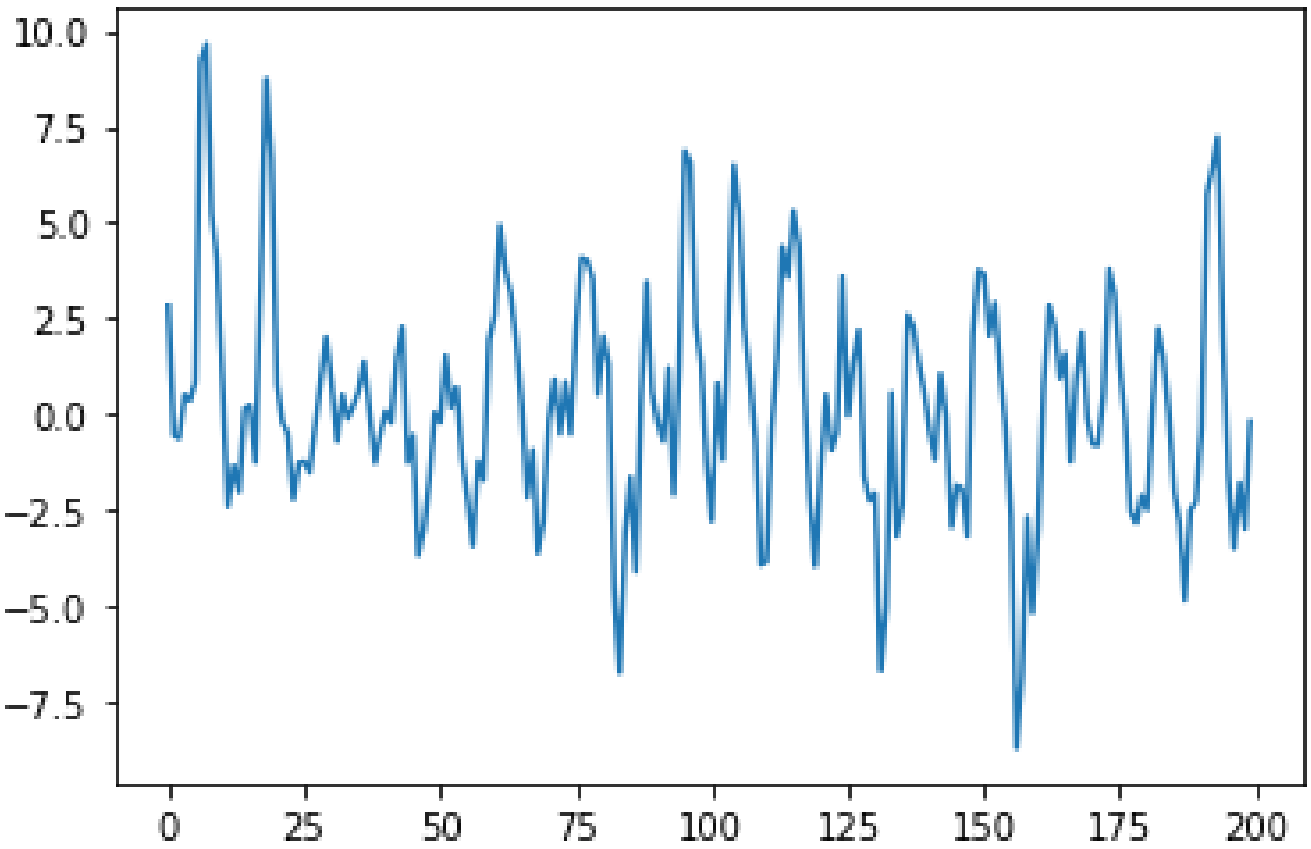
⇒Non stationnaire car périodicité

Exemples



$\Rightarrow$ Non stationnaire car tendance croissante et périodicité

Exemples



⇒Stationnaire

Corrélation

Cross-Corrélation

Soit la cross-corrélation entre deux times series $\{x_t\}_{0\leq t\leq T}$ et $\{y_t\}_{0\leq t\leq T}$ de moyenne respectives μ_x et μ_y :

$$c_{xy}(h) = \frac{\mathbb{E}[(x_t-\mu_x)(y_{t+h}-\mu_y)]}{\sqrt{\mathbb{E}[(x_t-\mu_x)^2]\mathbb{E}[(y_t-\mu_y)^2]}} = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}}.$$

$c_{xy}(h)$ donne une information de dépendance linéaire entre $\{x_t\}_{0\leq t\leq T}$ et $\{y_t\}_{0\leq t\leq T}$:

- Si $c_{xy}(h) = 1$, alors $\exists\, \alpha, \beta \in \mathbb{R}$ tels que $x_t = \alpha y_{t+h} + \beta$
- Si $c_{xy}(h) = -1$, alors $\exists\, \alpha, \beta \in \mathbb{R}$ tels que $x_t = \alpha(\mu_y - y_{t+h}) + \beta$
- Si $c_{xy}(h) = 0$, les deux séries sont indépendantes pour un lag de h

Corrélation

Estimation

La cross-corrélation est généralement estimée par la cross-corrélation empirique sur un ensemble de N échantillons :

$$\hat{c}_{xy}(h) = \frac{\sum_{n=1}^{N}(x_n - \overline{x})(y_{n+h} - \overline{y})]}{\sqrt{\sum_{n=1}^{N}(x_n - \overline{x})^2 \sum_{n=1}^{N}(y_{n+h} - \overline{y})^2}}.$$

La cross-corrélation est généralement estimée par la cross-corrélation empirique $\{\hat{c}_{xy}(h)\}$ sur un ensemble de N échantillons compris dans des fenêtres glissantes décallée d'un lag de h.

Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Prédicteur linéaire optimal

Idée : régression locale ?

- Estimations des paramètres de la série localement par régression
- Utiliser le passé pour prédire le futur

Prédicteur linéaire optimal

Idée : régression locale ?

- Estimations des paramètres de la série localement par régression
- Utiliser le passé pour prédire le futur

Prédiction progressive

$$\hat{x}_t = \boldsymbol{\varphi}^T \boldsymbol{x} \quad \text{avec} \quad \boldsymbol{\varphi} = \underset{\boldsymbol{\varphi}}{\operatorname{argmin}}\, \mathbb{E}\left[(x_t - \boldsymbol{\varphi}^T \boldsymbol{x})^2\right]$$

où $\boldsymbol{x} = [x_{t-1}, \cdots, x_{t-p}]$

Prédicteur linéaire optimal

Idée : régression locale ?

- Estimations des paramètres de la série localement par régression
- Utiliser le passé pour prédire le futur

Prédiction progressive

$$\hat{x}_t = \boldsymbol{\varphi}^T \boldsymbol{x} \quad \text{avec} \quad \boldsymbol{\varphi} = \underset{\boldsymbol{\varphi}}{\operatorname{argmin}}\, \mathbb{E}\left[(x_t - \boldsymbol{\varphi}^T \boldsymbol{x})^2\right]$$

où $\boldsymbol{x} = [x_{t-1}, \cdots, x_{t-p}]$

$\hat{x}$ = projection orthogonale de x sur $\text{vec}\{x_{t-1}, \cdots, x_{t-p}\}$

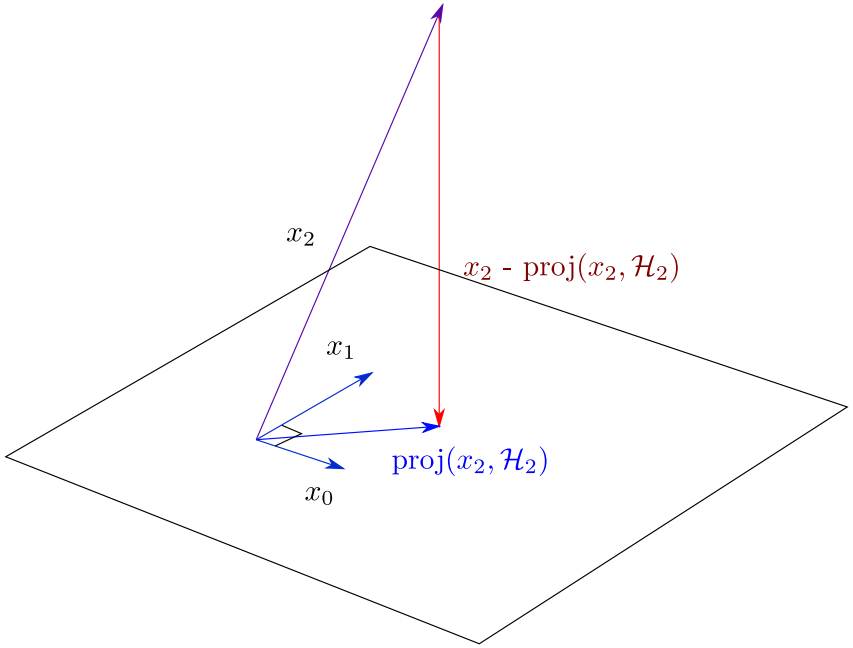
Ici on suppose $\{x_t\}_{t\in\mathbb{N}}$ est de moyenne nulle (ne pas oublier de la soustraire sinon)

Predicteur optimale : intuition géométrique

- On considère pour le produit scalaire :

$$< x, y >= \mathbb{E}(xy)$$

  attention à soustraire la norme, si $\mathbb{E}(x)$ et $\mathbb{E}(y) \neq 0$
- Le prédicteur linéaire optimal au sens de l'erreur quadratique moyenne revient à trouver la projection de x_t sur le sous espace dénifi par $\{x_{t-1}, ..., x_{t-p}\}$



Prédicteur linéaire optimal

Équation de Yule-Walker

Pour $\boldsymbol{x} = [x_{t-1}, \cdots, x_{t-p}]$,

$$\boldsymbol{\Gamma}_p \boldsymbol{\varphi}_p = \boldsymbol{\gamma}_p$$

avec

- L'indice ligne i, colonne j donné par $\boldsymbol{\Gamma}_{p_{i,j}} = \gamma_{xx}(i-j)$
- $\boldsymbol{\gamma}_p = [\gamma_{xx}(1), \cdots, \gamma_{xx}(p)]^T$

Mean Square Error

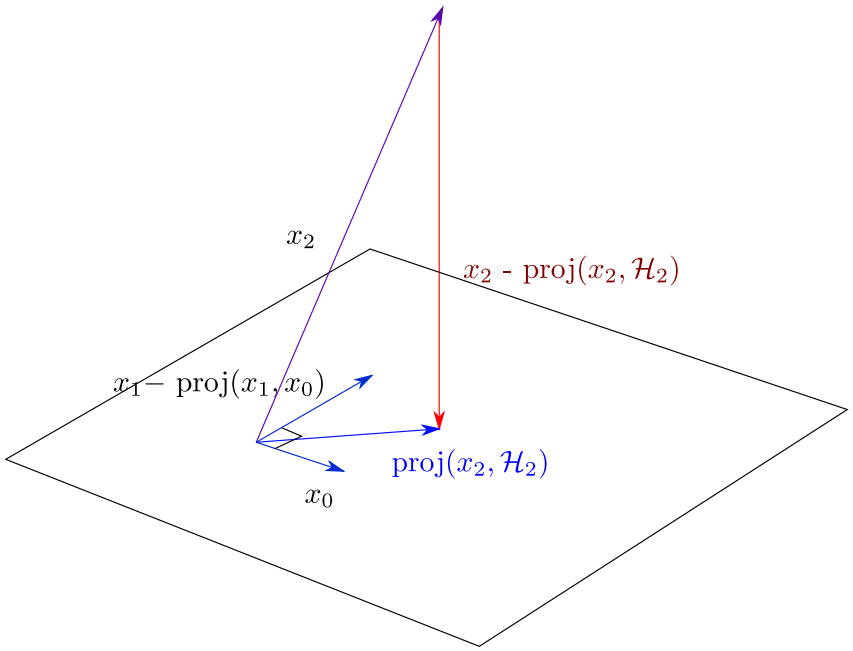
L'erreur d'approximation est donnée par

$$\text{MSE} = \mathbb{E}||x_t - \boldsymbol{\varphi}^T \boldsymbol{x}||_2^2 = \gamma(0) - \boldsymbol{\varphi}_p^T \boldsymbol{\Gamma}_p \boldsymbol{\varphi}_p$$

Estimation récursive

Algorithme de Gram-Schmidt ou des innovations

Intuition : si on connaît une base orthonormale décrivant l'ensemble des observations rencontrées, une nouvelle observation peut se décomposer simplement sur cette base



Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Théorème de représentativité de Wold

Théorème

Une série $\{x_t\}_{t\in\mathbb{N}}$, **stationnaire au sens large**, peut être parfaitement représentée par la somme d'un processus déterministe et d'un processus stochastique :

$$x_t = v_t + s_t$$

$$s_t = \sum_i^{\infty} \theta_i z_{t-i} \quad \text{avec } z_t \text{ bruit blanc}$$

$$v_t = \text{ processus déterministe}$$

Opérateur de retard

Définition

On définit l'opérateur de retard B comme un décallage un instant précédent :

$$B(x_t) = x_{t-1} \quad \text{et} \quad B^n(x_t) = x_{t-n}$$

Inversibilité

On peut définir une série temporelle auto-regressive *ie.* dont le futur dépend exclusivement des réalisations précédentes par un *polynôme de retards* :

$$x_t = \sum_{i=1}^{\infty} \varphi_i x_{t-i} \quad \text{alors} \quad x_t = \Psi(B) x_{t-1} \quad \text{avec} \quad \Psi(B) = \sum_{i=0}^{\infty} \varphi_i B^i$$

Si $\Psi(B)$ est inversible *ie.* il existe $\Psi^{-1}(B)$ tel que $\Psi^{-1}(B)\Psi(B) = B^0$ alors $\{x_t\}_{t\in\mathbb{N}}$ est dite inversible.

ARMA / ARIMA

Auto Regressif Moving Average Processus (ARMA)

Un processus est ARMA d'ordre $(p, q) \in \mathbb{N}^2$ s'il peut s'écrire de la façon suivante :

$$x_t = z_t + \sum_{k=1}^{p} \varphi_k x_{t-k} + \sum_{k=1}^{q} \theta_k z_{t-k}$$

$$\Phi(B) x_t = \Theta(B) z_t$$

où $\forall t, Z_t \sim \mathcal{N}(0, \sigma^2)$ et $\forall i,\ \varphi_i$ et $\theta_j \in \mathbb{R}$. Les polynômes Φ et Θ représentent respectivement la partie autorégressive et Moving average du processus.
$\{x_t\}_{t \in \mathbb{N}}$ est stationnaire au sens large et inversible si les racines de $\Phi(B)$ et $\Theta(B)$ sont hors du cercle unité.

Exemple : AR(1)

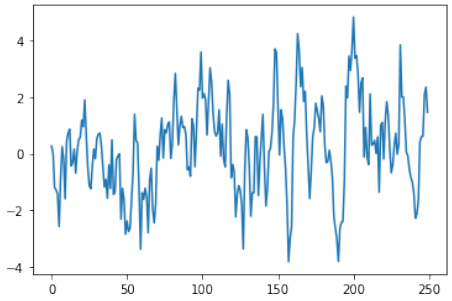


FIGURE – $|\varphi| < 1$

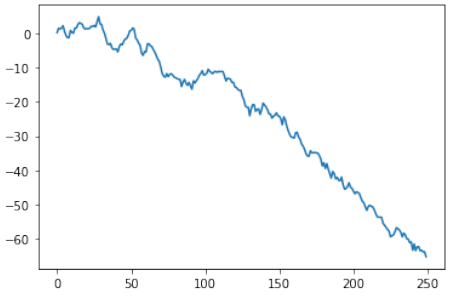


FIGURE – $|\varphi| > 1$

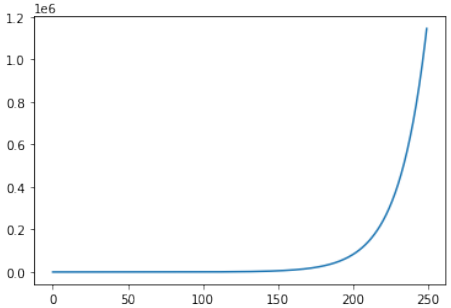


FIGURE – $|\varphi| > 1$

Opérateur de différence temporel

Définition

On note Δ l'opérateur de différence temporel tel que :

$$\Delta = 1 - B$$
$$\Delta^2 = (1-B)^2 = 1 - 2B + B^2$$
$$\Delta^k = (1-B)^k = \sum_{j=0}^{k} \binom{k}{j} (-B)^j$$

Extraction de tendance

- Si $\{x_t\}_{t\in\mathbb{N}}$ a une tendance linéaire $\{\Delta x_t\}_{t\in\mathbb{N}}$ n'a pas de tendance
- Si $\{x_t\}_{t\in\mathbb{N}}$ a une tendance quadratique $\{\Delta^2 x_t\}_{t\in\mathbb{N}}$ n'a pas de tendance
- Etc

L'opérateur de différence est donc un moyen pour **rendre une série stationnaire**

Processus ARIMA

En pratique la plupart des séries temporelles ne sont pas stationnaires. Pour, faire correspondre un modèle stationnaire *eg.* ARMA, il faut d'abord retirer les tendances composant le signal.

Integrated ARMA or ARIMA

$\{x_t\}_{t\in\mathbb{N}}$ est ARIMA d'ordre (p, d, q) si $\{\Delta^d x_t\}_{t\in\mathbb{N}}$ est ARMA d'ordre (p, q). On a donc la forme suivante pour $\{x_t\}_{t\in\mathbb{N}}$:

$$\Phi(B)\Delta^d x_t = \Theta(B) z_t$$

$\Rightarrow$ On trouve le modèle ARMA expliquant $\{w_t\}_{t\in\mathbb{N}} = \{\Delta^d x_t\}_{t\in\mathbb{N}}$

Processus SARIMA

En plus des tendances, des phénomènes périodique *ie.* saisonniers doivent souvent être gérés

Seasonal ARIMA

$\{x_t\}_{t\in\mathbb{N}}$ est un processus SARIMA d'ordre $(p,d,q)\times(P,D,Q)_s$ si $\{x_t\}_{t\in\mathbb{N}}$ peut être décrit par la récurrence suivante :

$$\Phi(B)\Phi'\left(B^s\right)\Delta^d\Delta_s^D x_t = \theta(B)\Theta\left(B^s\right)z_t$$

avec $\Delta_s^D = (1-B^s)^D$.

Estimation des paramètres

- Estimation de l'ordre des tendances d et des périodicités D
- Une fois retirées le processus doit être stationaire
- Les valeurs d'ordre p, P, q et Q doivent être déterminées par exemple par essaie erreur, en regardant les fonctions d'autocorrélation et autocorrélation réduites, par critère BIC/AIC
- Maximisation de la vraissemblance du model *eg.* par moindre de carré si hypothèse gaussienne. En pratique cette étape est faite par un software ou une bibliothèque spécialisée en statistiques

Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Simple Exponential Smoothing

SES

- $\{x_t\}_{t\in\mathbb{N}}$ une série sans tendance ni périodicité
- Prédiction des valeurs x_t en se basant sur une combinaison linéaire des valeurs passées :

$$\hat{x}_{N+1} = \sum_{M=0}^{\infty} c_n x_{N-M}$$

- Poids géométriques souvent privilégiés :

$$c_i = \alpha(1-\alpha)^i \quad \text{avec } \alpha \in [0,1]$$

- Bien adapté pour représenter des MA(∞) :

$$x_t = \mu + \alpha \sum_{j<t} z_j + z_t$$

Holt procédure

Holt forecasting

- On note L_t la moyenne locale ou le **niveau** de la série
- Sans tendance/périodicité, le niveau L_t est approximativement constant. Une estimation itérative de L_t peut donc avoir la forme :

$$L_t \simeq \underbrace{\alpha x_t}_{correction} + \underbrace{(1-\alpha)L_{t-1}}_{approximation}$$

- Avec α et $\gamma \in [0,1]$

Holt procédure

Holt forecasting

- Une tendance T_t peut être vu comme une variation locale de $L_t = L_{t-1} + T_{t-1}$
- Estimation du niveau et de la tendance par :

$$L_t = \underbrace{\alpha x_t}_{correction} + \underbrace{(1-\alpha)(L_{t-1} + T_{t-1})}_{approximation}$$

$$T_t = \underbrace{\gamma(L_t - L_{t-1})}_{correction} + \underbrace{(1-\gamma)T_{t-1}}_{approximation \simeq T_t \ constant}$$

- Prédiction donnée par

$$\hat{x}_{t+h} = L_t + hT_t$$

- Avec α et $\gamma \in [0, 1]$

Holt-Winters procédure

Périodicité

On note S_t la périodicité de la série :

- $S_{t+\tau} = S_t$. Ici τ est la période de S_t
- Multiplicative : $x_t \simeq (L_t + T_{t-1})\, S_t$

Holt-Winters : cas multiplicatif

- Approximation $x_t \simeq L_t S_t$
- Estimation des composantes par :

$$L_t = \alpha \frac{x_t}{S_{t-\tau}} + (1-\alpha)\,(L_{t-1} + T_{t-1})$$
$$T_t = \gamma\,(L_t - L_{t-1}) + (1-\gamma) T_{t-1}$$
$$S_t = \delta \frac{x_t}{L_t} + (1-\delta) S_{t-\tau}$$

- Prédiction donnée par $\hat{x}_{t+h} = (L_t + hT_t)\, S_t$

Holt-Winters procédure

Périodicité

On note S_t la périodicité de la série :

- $S_{t+\tau} = S_t$. Ici τ est la période de S_t
- Additive : $x_t \simeq L_t + T_{t-1} + S_t$

Holt-Winters : cas additif

- Approximation $x_t \simeq L_t + S_t$
- Estimation des composantes par :

$$
\begin{aligned}
L_t &= \alpha(x_t - S_{t-\tau}) + (1-\alpha)\left(L_{t-1} + T_{t-1}\right) \\
T_t &= \gamma\left(L_t - L_{t-1}\right) + (1-\gamma)T_{t-1} \\
S_t &= \delta(x_t - L_t) + (1-\delta)S_{t-\tau}
\end{aligned}
$$

- Prédiction donnée par $\hat{x}_{t+h} = L_t + hT_t + S_t$

SARIMA ou Box-Jenkins procédure

- Identification du modèle SARIMA *ie.* trouver d et D
- Estimation des paramètres
- Prédiction :
  - Grâce à l'équation
  - Autoregressive après inversion de Φ :

$$\begin{aligned} \Phi(B)_{x_t} &= \Theta(B)z_t \\ x_t &= \Phi^{-1}(B)\Theta(B)z_t = \Psi(B)z_t \\ \text{Donne : } x_{t+h} &= \Psi(B)z_{t+h} \end{aligned}$$

  - Autoregressive après inversion de Θ et ré-écriture de z_t :

$$\begin{aligned} z_t &= \Theta^{-1}(B)\Phi(B)x_t = \Pi(B)x_t \\ \text{Donne : } x_{t+h} &= \Pi(B)x_{t+h-1} + z_{t+h} \end{aligned}$$

- Cf. TP !

Plan

1 Introduction

2 Stationarité

3 Prédiction linéaire

4 Processus ARMA

5 Prédiction

6 Conclusion

Conclusion

- Modélisation grâce au théorème de Wold
- Approches SOTA dans de nombreux domaines jusqu'aux années 2010
- Simple à mettre en oeuvre
- Toujours compétitif pour la prédiction à court terme
- Dépassé en performance par des approches basées apprentissage profond **si** suffisamment de données d'apprentissage disponibles

Références

- *The Analysis of Time Series - An introduction*, 6th edition, Chris Chatfield, *Chapman & HAll*
- *Forecasting Methods and Applications*, 3rd edition, Makridakis, Whellwright, Hyndman, *Wiley*
- *Time Series Analysis*, Hamilton, *Princeton University Press*