

Séries Temporelles - cours 1

RCP217

Clément Rambour

Plan

1 Introduction

2 Prédiction linéaire

3 Processus auto-régressifs (AR) et moving average (MA)

4 Filtre de Kalman

5 Travaux pratiques

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 série
 - 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 carrés si les caractéristiques du bruit sont constantes :

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

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 carrés si les caractéristiques du bruit sont constantes :

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

- Si s_t varie "lentement" → **tendance**
- Si s_t périodique → **composante saisonnière**
- n_t → **bruit**

Analyse de séries temporelles vs. régression

Particularités

- Les mêmes données peuvent être présentées différemment selon le lieu et l'époque
- Les données sont (très) corrélées
 - Le passé permet de prédire le futur
 - Les données ne sont **pas iid**
- Les structures peuvent être visibles sur des axes temporels 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 $\forall h \in \mathbb{Z}, n \in \mathbb{N}$, $(x_{t_1}, \dots, x_{t_n})$ et $(x_{t_1+h}, \dots, x_{t_n+h})$ 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 $\{x_t\}_{t \in \mathbb{Z}}$ en $t \in \mathbb{Z}$ est donnée par

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

Covariance et Autocovariance

La covariance de $\{x_t\}_{t \in \mathbb{Z}}$ et $\{y_t\}_{t \in \mathbb{Z}}$ 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 $\{x_t\}_{t \in \mathbb{Z}}$ 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

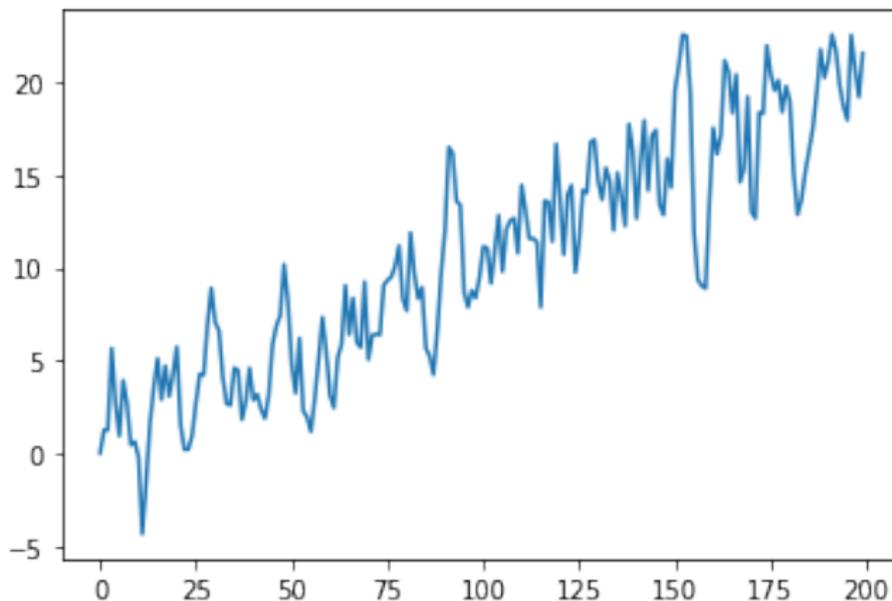
La moyenne et la covariance sont généralement estimées sur un ensemble de N échantillons. On note :

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

La covariance empirique est calculée comme :

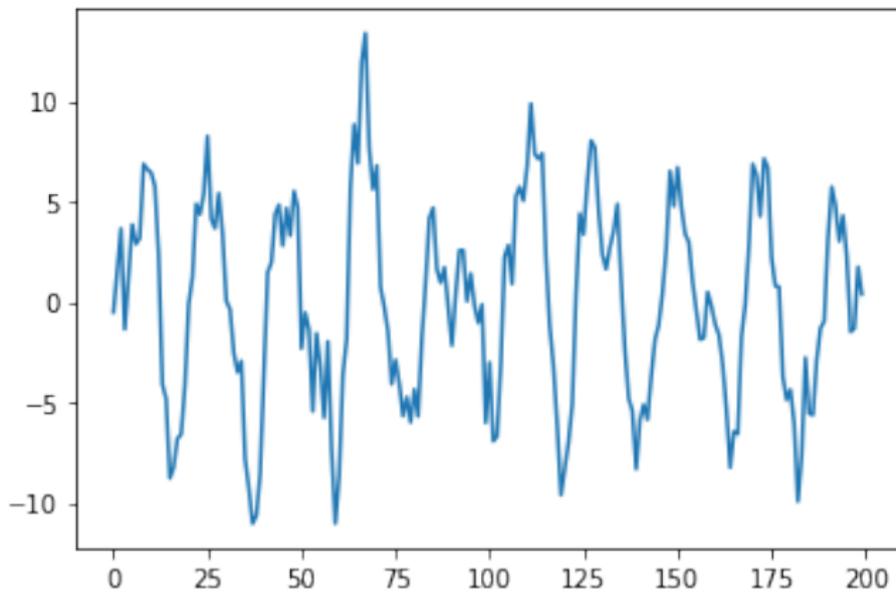
$$\hat{\gamma}_{xy}(h) = \sum_{n=1}^N (x_n - \bar{x})(y_{n+h} - \bar{y}).$$

Exemples



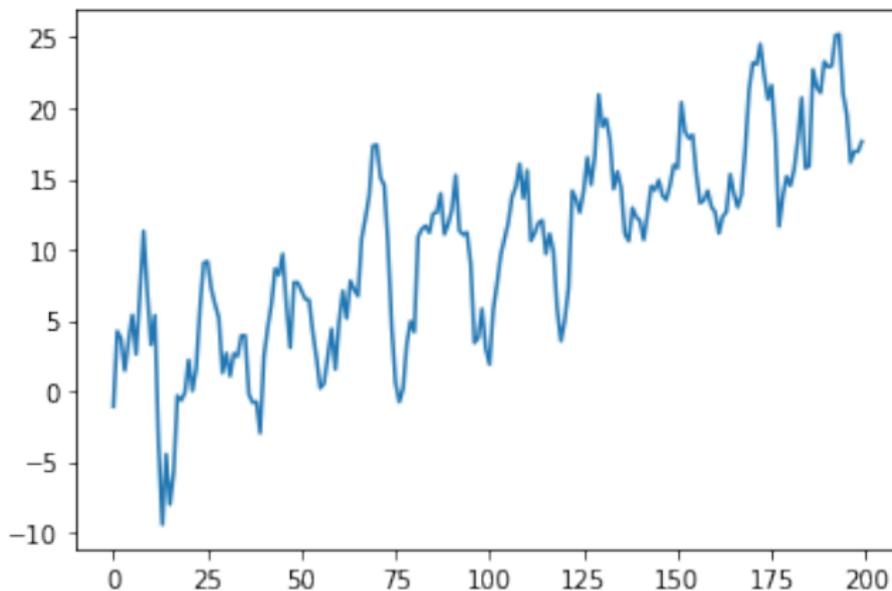
⇒ Non stationnaire car tendance croissante

Exemples



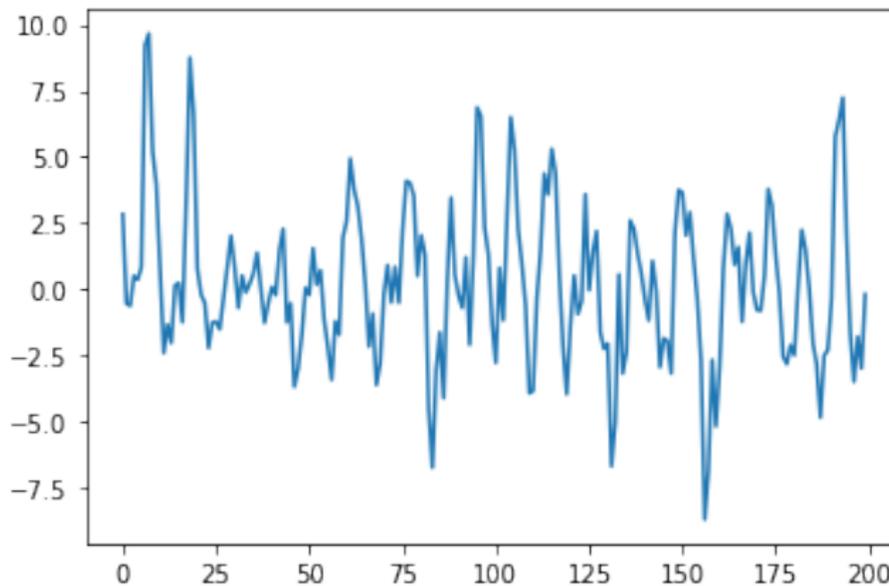
⇒ Non stationnaire car périodicité

Exemples



⇒ Non stationnaire car tendance croissante et périodicité

Exemples



⇒ Stationnaire

Corrélation

Cross-Corrélation

Soit la cross-corrélation entre deux times series x et y 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 et y :

- 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 :

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

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 décalée d'un lag de h .

Plan

1 Introduction

2 Prédiction linéaire

3 Processus auto-régressifs (AR) et moving average (MA)

4 Filtre de Kalman

5 Travaux pratiques

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

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

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

Prédicteur linéaire optimal

Idée : régression locale ?

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

Prédiction progressive

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

où $\mathbf{x} = [x_{t-1}, \dots, x_{t-p}]$

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

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

Prédicteur linéaire optimal

Équation de Yule-Walker

Pour $\mathbf{x} = [x_{t-1}, \dots, x_{t-p}]$,

$$\mathbf{\Gamma}_p \boldsymbol{\varphi}_p = \gamma_p$$

avec $\mathbf{\Gamma}_{p_{i,j}} = \gamma_{xx}(i-j)$ et $\boldsymbol{\gamma}_p = [\gamma_{xx}(1), \dots, \gamma_{xx}(p)]^T$.

Mean Square Error

L'erreur d'approximation est donnée par

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

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

Estimation récursive

Algorithme de Gram-Schmidt ou des innovations

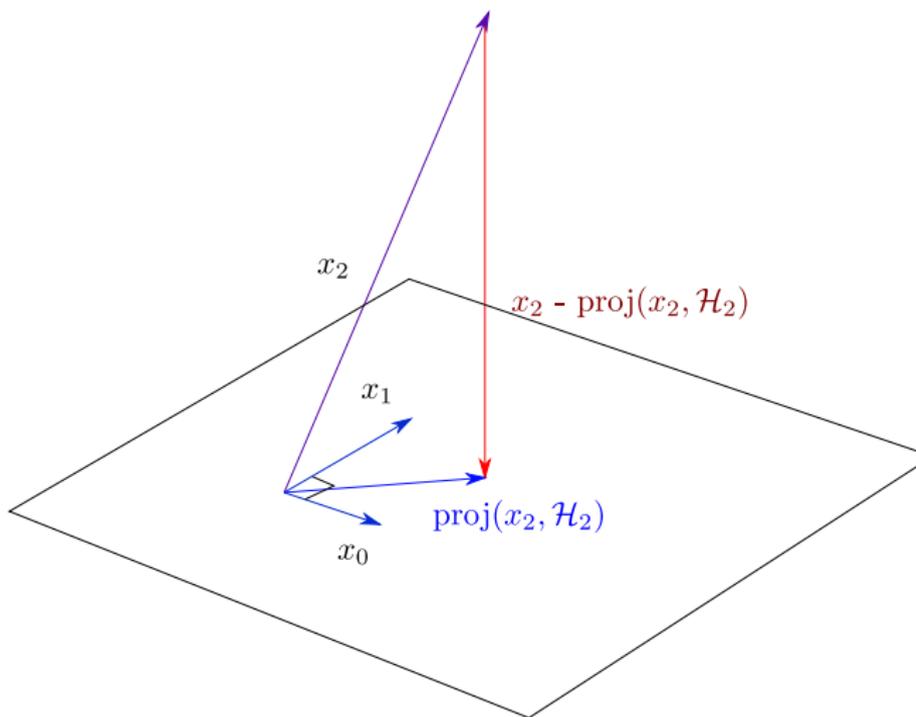
Intuition : si on connaît une base orthonormale décrivant l'ensemble des observations rencontrées, une nouvelle observation peut se décomposer simplement sur cette base.

On considère pour cela le produit scalaire :

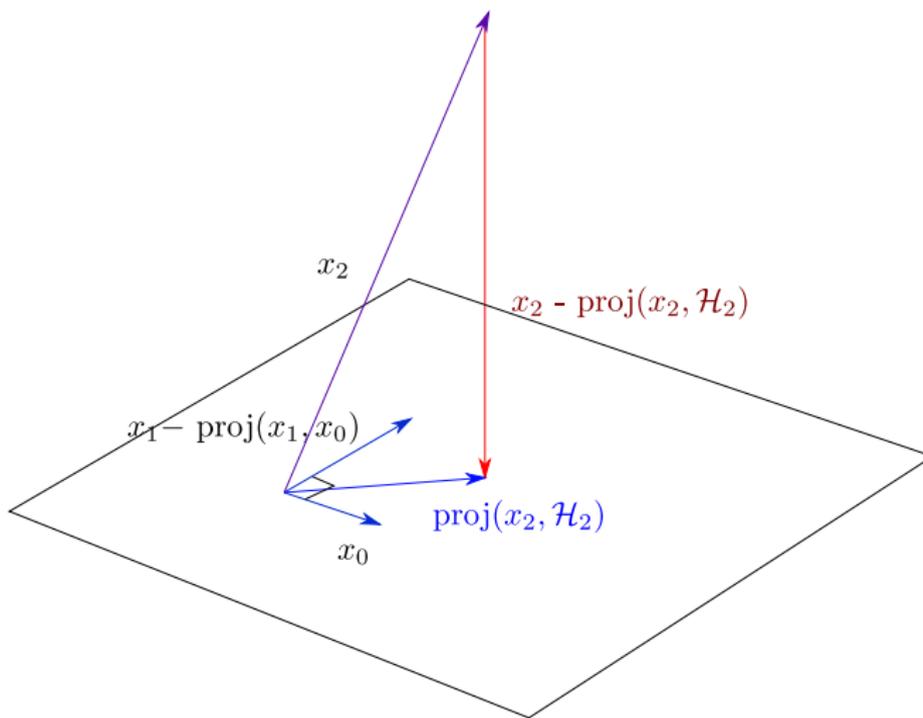
$$\langle x, y \rangle = \mathbb{E}(xy)$$

attention à soustraire la norme, si $\mathbb{E}(x)$ et $\mathbb{E}(y) \neq 0$

Estimation récursive



Estimation récursive



Estimation récursive

Algorithme de Gram-Schmidt ou des innovations

Motivation : décomposer x_{n+1} sur la base orthogonale \mathcal{H}_n donnée par la récurrence : $[x_1, x_2 - \text{proj}(x_1, x_2), \dots, x_n - \text{proj}(x_n, \mathcal{H}_{n-1})]$ où $\text{proj}(y, z) = \mathbb{E}(yz)z$

On a alors :

$$\forall n \geq 1, \quad \hat{x}_{n+1} = \sum_{k=1}^n \theta_{n,k} (x_{n+1-k} - \hat{x}_{n+1-k})$$

$$\theta_{n,n-k} = \frac{\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2}{\sigma_k^2}$$

$$\sigma_n^2 = \gamma(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \sigma_j^2$$

Estimation récursive

Algorithme de Gram-Schmidt ou des innovations

Motivation : décomposer x_{n+1} sur la base orthogonale \mathcal{H}_n donnée par la récurrence : $[x_1, x_2 - \text{proj}(x_1, x_2), \dots, x_n - \text{proj}(x_n, \mathcal{H}_{n-1})]$ où $\text{proj}(y, z) = \mathbb{E}(yz)z$

On a alors :

$$\forall n \geq 1, \quad \hat{x}_{n+1} = \sum_{k=1}^n \theta_{n,k} (x_{n+1-k} - \hat{x}_{n+1-k})$$

$$\theta_{n,n-k} = \frac{\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2}{\sigma_k^2}$$

$$-2 \quad \dots \quad (n+1, n+1) \quad \sum_{j=0}^{n-1} \theta_{k,k-j}^2 \quad -2$$

Lorsque, $\{x_t\}_{t \in \mathbb{Z}}$ est stationnaire et inversible, on peut également utiliser l'algorithme de Levinson-Durbin.

Plan

1 Introduction

2 Prédiction linéaire

3 Processus auto-régressifs (AR) et moving average (MA)

4 Filtre de Kalman

5 Travaux pratiques

Processus Auto-Régressifs (AR)

Formalisation de la récurrence locale vu précédemment ?

- Modèle linéaire d'estimation en fonction du passé
- Le choix de se porte sur la mémoire utilisée

Processus Auto-Régressifs (AR)

Formalisation de la récurrence locale vu précédemment ?

- Modèle linéaire d'estimation en fonction du passé
- Le choix de se porte sur la mémoire utilisée

Modèle Auto-Régressif

Un processus est auto-régressif d'ordre $0 \leq p$ s'il est du stationnaire au sens large et solution de l'équation de récurrence

$$X_t = Z_t + \sum_{k=1}^p \varphi_k X_{t-k}$$

où $Z_t \sim \mathcal{N}(0, \sigma^2)$ et $\forall k, \varphi_k \in \mathbb{R}$.

Exemple : AR(1)

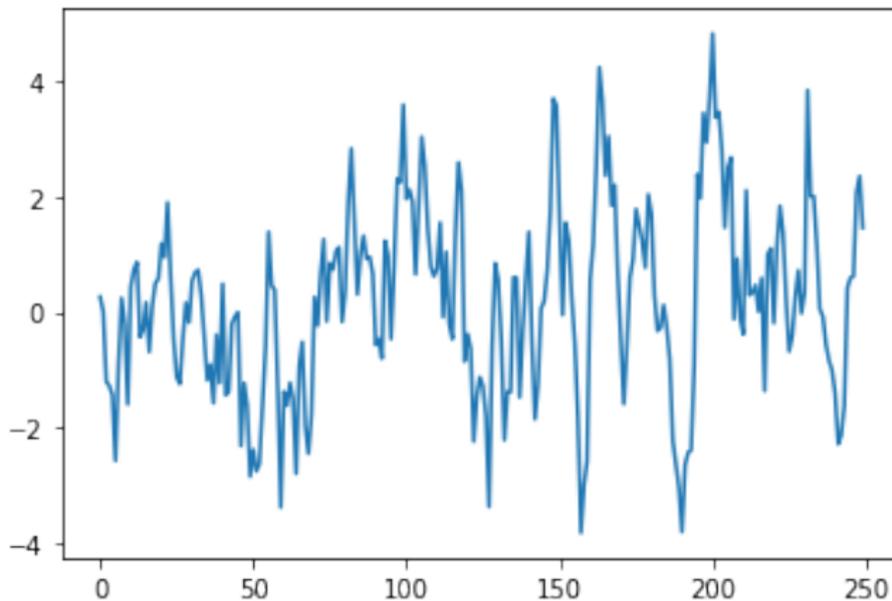


FIGURE – $\varphi < 1$

Exemple : AR(1)

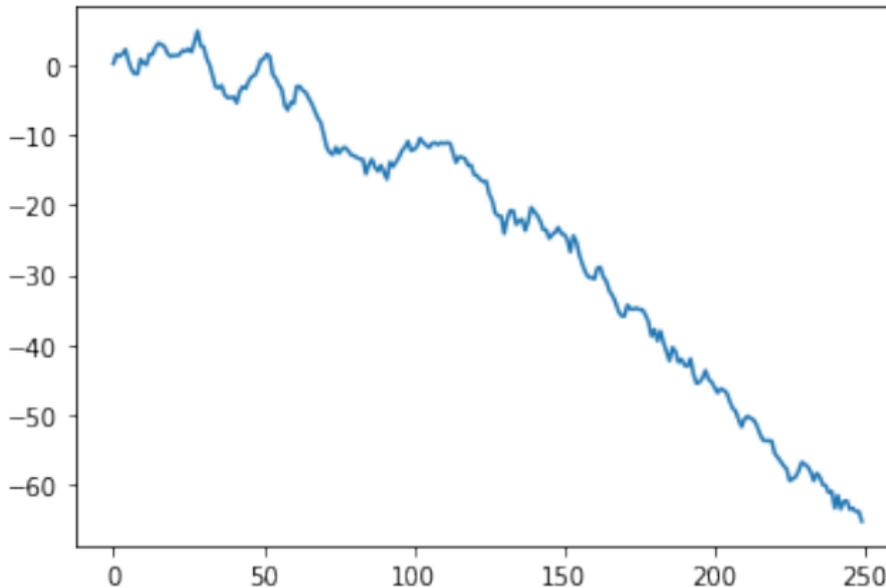


FIGURE - $\varphi > 1$

Exemple : AR(1)

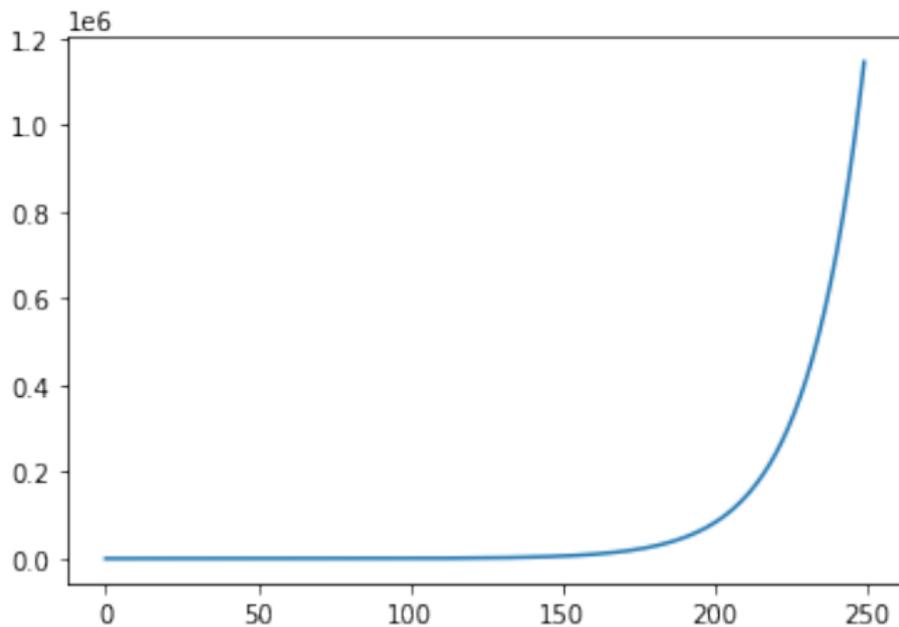


FIGURE – $\varphi > 1$

Moving Average ou Moyenne Mobile (MA)

Que fait on des erreurs des erreurs d'estimation ?

- On les propage dans l'estimation du futur
- \simeq intégrer le bruit dans le temps

Moving Average ou Moyenne Mobile (MA)

Que fait-on des erreurs des erreurs d'estimation ?

- On les propage dans l'estimation du futur
- \simeq intégrer le bruit dans le temps

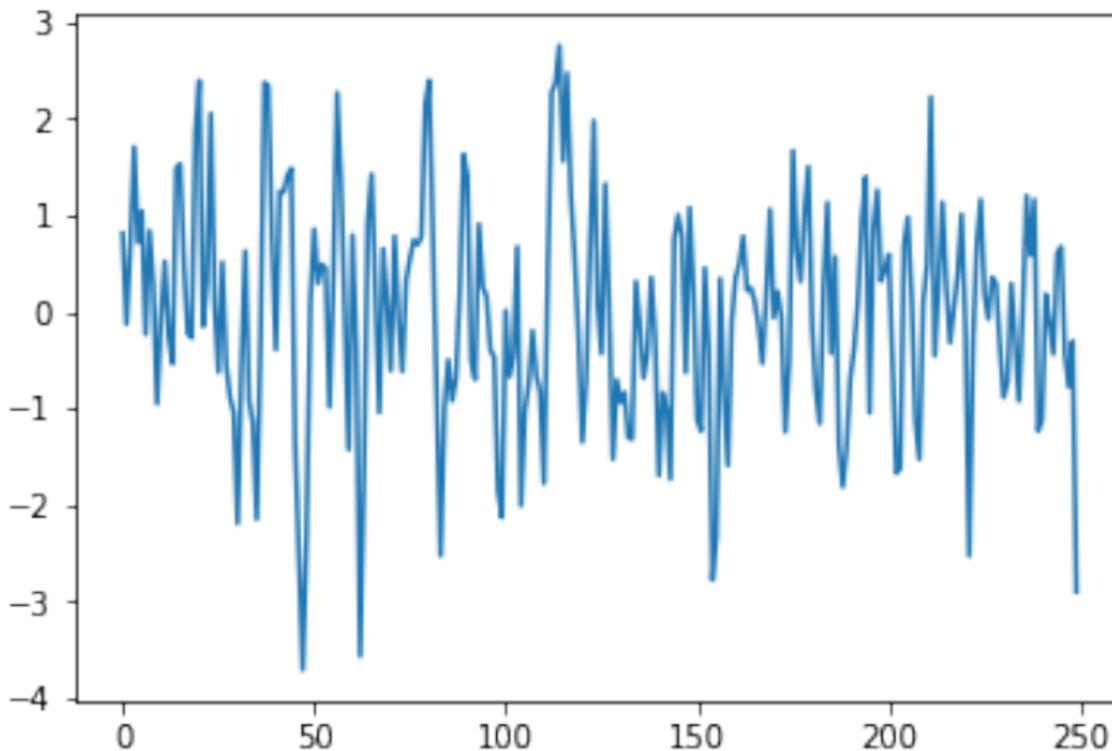
Processus Moving Average (MA)

Un processus est auto-régressif d'ordre $0 \leq q$ s'il est du stationnaire au sens large et solution de l'équation de récurrence

$$X_t = Z_t + \sum_{k=1}^q \theta_k Z_{t-k}$$

où $\forall t, Z_t \sim \mathcal{N}(0, \sigma^2)$ et $\forall k, \theta_k \in \mathbb{R}$.

Exemple : MA(1)



ARMA

Auto Régressif Moving Average Processus (ARMA)

Un processus est ARMA d'ordre $(p, q) \in \mathbb{N}^2$ s'il est du stationnaire au sens large et solution de l'équation de récurrence

$$X_t = Z_t + \sum_{k=1}^p \varphi_k X_{t-k} + \sum_{k=1}^q \theta_k Z_{t-k}$$

où $\forall t, Z_t \sim \mathcal{N}(0, \sigma^2)$ et $\forall i, j, \varphi_i$ and $\theta_j \in \mathbb{R}$.

Existence, causalité et inversibilité

Existence

Si le polynôme $\Phi(z) = \sum_{k=1}^p \varphi_k z^k$ n'a pas de racine de module 1 alors ARMA(p,q) possède une unique solution stationnaire $\iff \{x_t\}_{t \in \mathbb{Z}}$ peut s'écrire comme une combinaison linéaire de $\{z_t\}_{t \in \mathbb{Z}}$:

$$x_t = \sum_{k \in \mathbb{Z}} \alpha_k z_{t-k}$$

Causalité

Si $\Phi(z)$ n'a pas de racine $|z| < 1$ alors la solution est causale *ie* ne dépend que du passé : $\{k | \alpha_k \neq 0\} \subset \mathbb{N}$

Inversibilité

Si $\Theta(z)$ n'a pas de racine $|z| < 1$ alors la solution est inversible *ie* $\{z_t\}_{t \in \mathbb{Z}}$ admet une décomposition causale de $\{x_t\}_{t \in \mathbb{Z}}$

Estimation des paramètres

Pour un processus AR, les modèles de prédiction linéaires s'appliquent et l'algorithme de Yule-Walker peut être utilisé pour estimer $\{\varphi_1, \dots, \varphi_p\}$

Maximum de vraisemblance

Estimation par Gram-Schmidt de $\{\hat{x}_1, \dots, \hat{x}_n\}$. On a alors la vraisemblance :

$$\mathcal{L}(\varphi, \theta, \sigma) = \frac{1}{(2\pi\sigma^2)^n} \exp\left(\sum_{j=1}^n \frac{(x_j(\varphi, \theta) - \hat{x}_j)^2}{2\sigma^2}\right)$$

On a donc,

$$(\hat{\varphi}, \hat{\theta}, \hat{\sigma}) = \underset{\varphi, \theta, \sigma}{\operatorname{argmin}} \mathcal{L}(\varphi, \theta, \sigma)$$

Plan

- 1 Introduction
- 2 Prédiction linéaire
- 3 Processus auto-régressifs (AR) et moving average (MA)
- 4 Filtre de Kalman**
- 5 Travaux pratiques

Principe

- Estimer l'état d'un système à partir de mesures incomplètes
- Dans une problématique de prédiction : état = valeurs suivantes du processus
- Pouvoir estimer l'erreur
- Pouvoir générer des estimation online (de façon récursive)

Intuition

L'idée à la base du filtre de Kalman se décompose en deux étapes

Prédiction

D'après un modèle (physique, statistique, signal...), on émet une prédiction

Correction

Une fois la mesure disponible, on règle un paramètre de d'ajustement

Pré-requis (rappel)

Estimateur linéaire non biaisé

Soit \mathbf{x} et \mathbf{y} , deux vecteurs aléatoires. Il existe un unique estimateur linéaire non biaisé $\hat{\mathbf{x}}$ à partir de \mathbf{y} donné par :

$$\hat{\mathbf{x}} = \mathbb{E}(\mathbf{x}) + \mathbf{K}(\mathbf{y} - \mathbb{E}(\mathbf{y}))$$

$$\mathbf{K} = \mathbf{\Gamma}_{xy} \mathbf{\Gamma}_y^{-1}$$

Équations d'état

$$\begin{cases} x_{t+1} &= \mathbf{A}x_t + \mathbf{u}_t + \mathbf{w}_t \\ y_{t+1} &= \mathbf{C}x_t + n_t \end{cases}$$

- \mathbf{w}_t et n_t : bruits blanc gaussiens
- \mathbf{x}_t : état du système (position, vitesse, valeur boursière, potentiel...)
- \mathbf{y}_t : mesure du système (capteurs, vidéos...)
- Les matrices \mathbf{A} et \mathbf{C} peuvent également varier dans le temps mais sont supposées connues

Filtre de Kalman

Objectif : obtenir la meilleure façon d'estimer \mathbf{x}_t à partir de l'estimation précédente $\hat{\mathbf{x}}_{t-1}$ et de la mesure courante \mathbf{y}_t

Intuition

En prenant $\hat{\mathbf{x}}_{t-1}$ comme un estimateur de la moyenne de \mathbf{x}_t , on peut faire l'estimation suivante :

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t-1} + \mathbf{K}_t(\mathbf{y} - \mathbf{C}\mathbf{x}_{n|n-1})$$

où la notation $n|n-1$ correspond à l'estimation de n sachant $[x_1, \dots, x_{n-1}]$.

Filtre de Kalman

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{u}_t$$

$$\mathbf{\Gamma}_{t|t-1} = \mathbf{A}\mathbf{\Gamma}_{t-1|t-1}\mathbf{A}^T + \mathbf{\Gamma}_w$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbb{E}(\mathbf{y}_t))$$

$$\hat{\mathbf{\Gamma}}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{C})\mathbf{\Gamma}_{t|t-1}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{\Gamma}_{t|t-1}\mathbf{C}^T + \mathbf{\Gamma}_n$$

$$\mathbf{K}_t = \mathbf{\Gamma}_{t|t-1}\mathbf{C}^T\mathbf{S}_t^{-1}$$

Plan

- 1 Introduction
- 2 Prédiction linéaire
- 3 Processus auto-régressifs (AR) et moving average (MA)
- 4 Filtre de Kalman
- 5 Travaux pratiques

- Implémenter le *1-step* forecasting à l'aide des équations Yule-Walker
- Implémenter l'algorithme des innovations pour le forecasting online
- Analyse de processus ARMA
- Développement d'un filtre de Kalman