Introduction aux séries temporelles

Paul Bernard (Paul.Bernard@inria.fr)

Equipe Mnemosyne - Inria/IMN/LaBRI

Sommaire

Séries temporelles

- I. Introduction et définitions
- II. Modélisation et analyse
- III. Évaluation des modèles
- IV. Ouvertures

(cet après midi : **Réseaux de neurones récurrents**)

Introduction
et définitions

Introduction: *Timeseries in the wild*

Nature des séries temporelles

Processus aléatoire (discret ou continu)

$\{\dots, X_t, X_{t+1}, \dots, X_{t+n}, \dots\}$

Observation et échantillonnage $h(X_t)$ $\Rightarrow$

Séries temporelles (discret)

$\{x_0, x_1, \dots, x_t, x_{t+1}, \dots, x_{t+m}\}$

déf. Un tirage (ou observation) d'un **processus aléatoire** $\{\dots, X_{tn}, X_{t(n+1)}, \dots, X_{t(n+m)}, \dots\}$ défini par un ensemble **(discret)** de **variables aléatoires indexées dans le temps** est une ***série temporelle***.

Introduction: *Timeseries in the wild*

Domaines d'application

Processus aléatoire (discret ou continu)

$\{\dots, X_t, X_{t+1}, \dots, X_{t+n}, \dots\}$

Observation et échantillonnage

$h(X_t)$

$\Longrightarrow$

Séries temporelles (discret)

$\{x_0, x_1, \dots, x_t, x_{t+1}, \dots, x_{t+m}\}$

ex.

- Systèmes physiques : météo, mouvement des océans, contrôle du rythme cardiaque, rayonnements lumineux et thermiques…
- Sources anthropiques : indicateurs économiques et financiers, langage, comportement…

- Données de terrain ou d'étude : température sur un an, rythme cardiaque sur une journée, vidéos et enregistrements audios, texte…
- **Données simulées : modèles géophysiques, économétriques, du langage…**

Introduction: *Timeseries in the wild*

Un problème difficile

Processus aléatoire (discret ou continu)

$$\{\ldots, X_t, X_{t+1}, \ldots, X_{t+n}, \ldots\}$$

Souvent inconnu/continu/caché

Observation et échantillonnage $h(X_t)$

$\Longrightarrow$

Souvent partiel(le)

Séries temporelles (discret)

$$\{x_0, x_1, \ldots, x_t, x_{t+1}, \ldots, x_{t+m}\}$$

Souvent incomplet/sans répétition

ex.

- Systèmes physiques : météo, mouvement des océans, contrôle du rythme cardiaque, rayonnements lumineux et thermiques…
- Sources anthropiques : indicateurs économiques et financiers, langage, comportement…

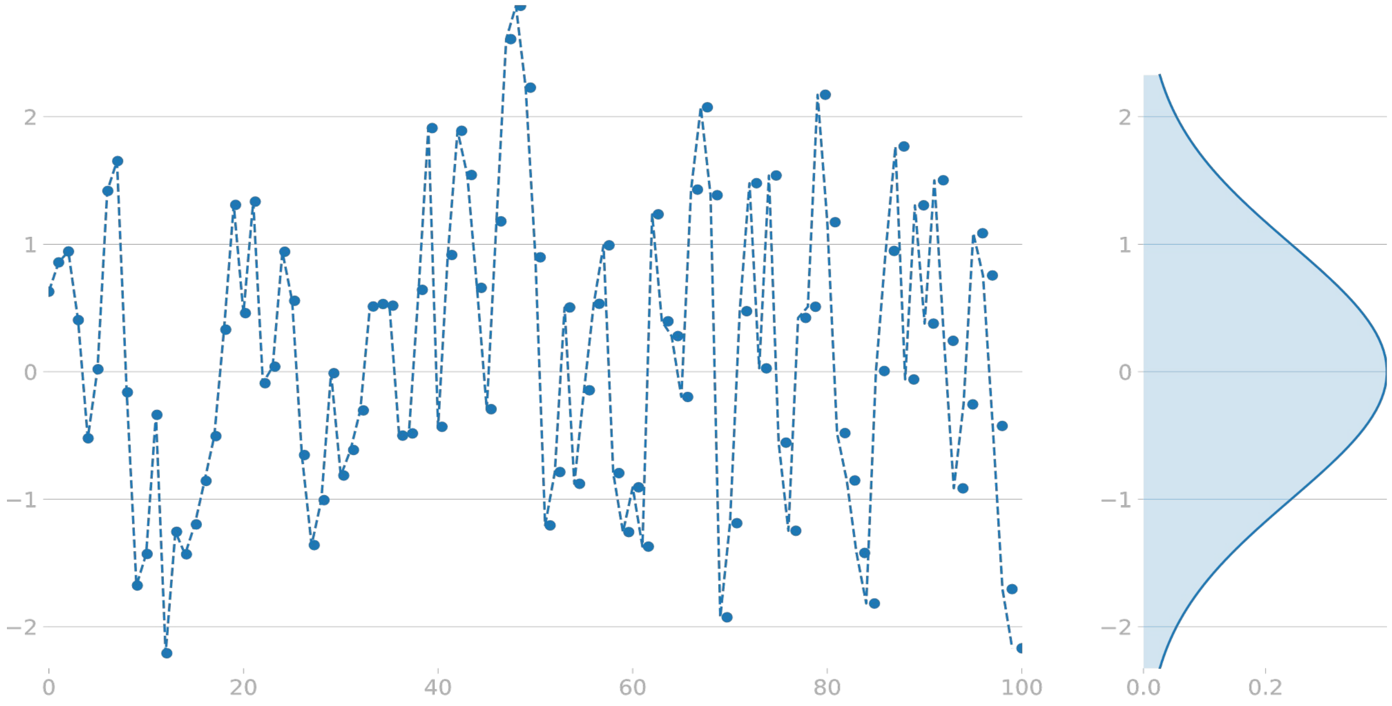
- Données de terrain ou d'étude : température sur un an, rythme cardiaque sur une journée, vidéos et enregistrements audios, texte…
- Données simulées : modèles géophysiques, économétriques, du langage…

Problématique

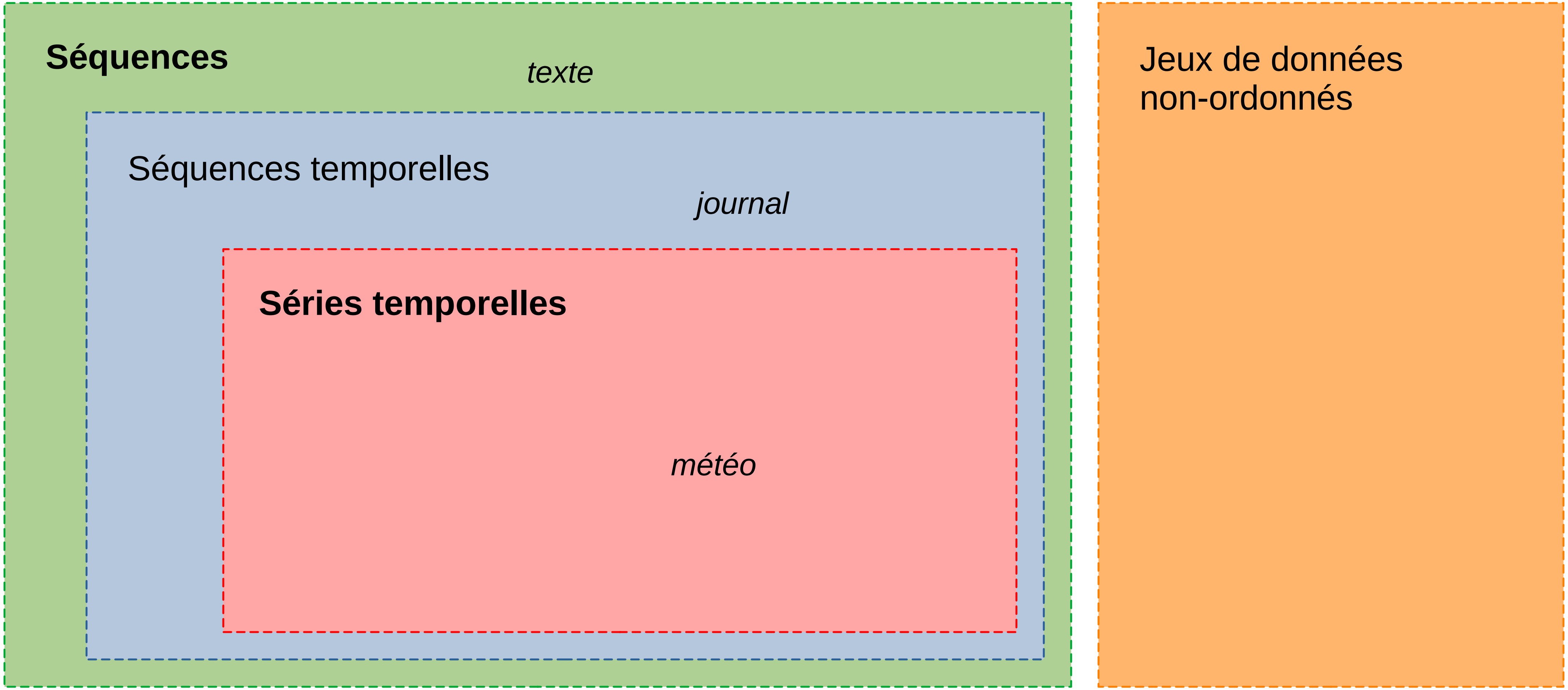
En quoi un processus aléatoire est différent d’une série de distributions d’événements aléatoires ?

Est-ce que ce n’est pas juste du bruit gaussien, mais horizontalement ?

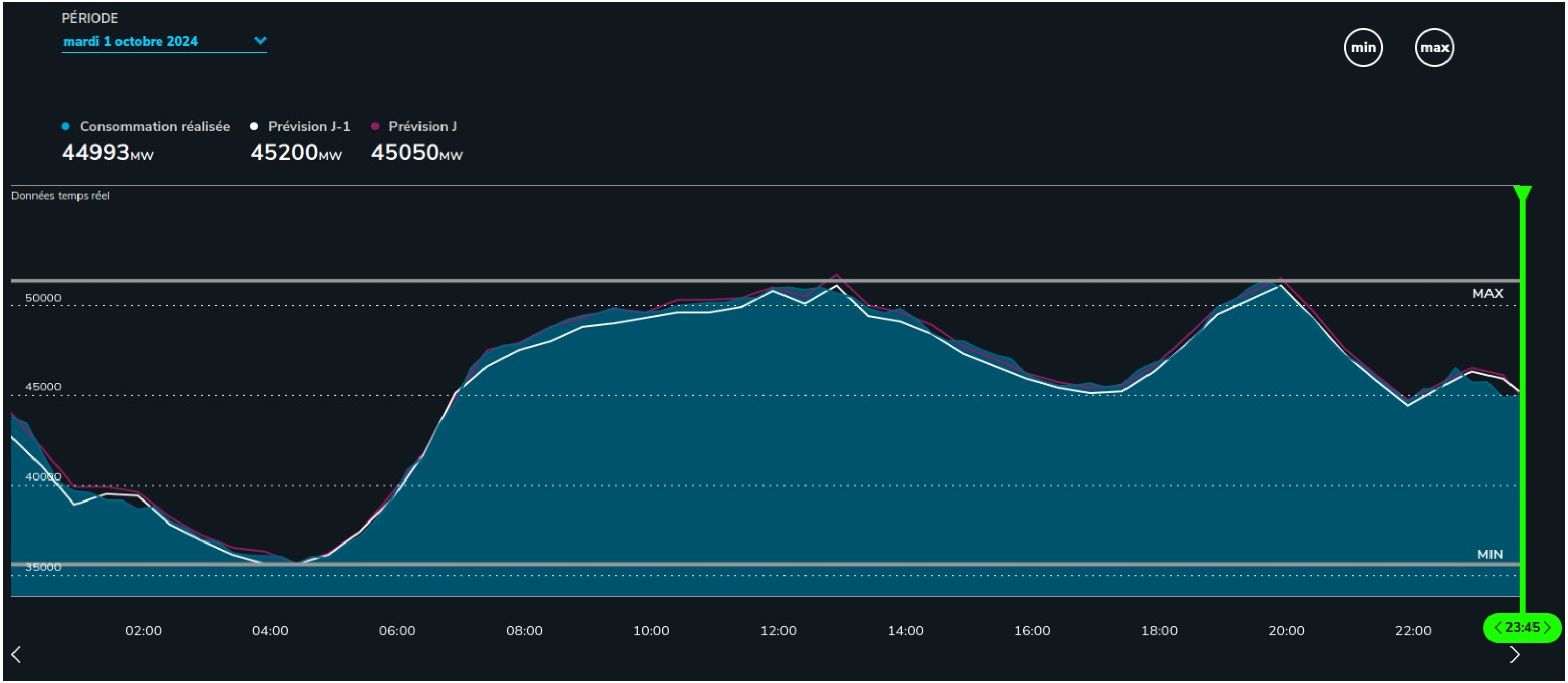
(spoilers: non)



Séquences et séries temporelles



Prédiction


PÉRIODE
mardi 1 octobre 2024
min
max
Consommation réalisée
44993MW
Prévision J-1
45200MW
Prévision J
45050MW
Données temps réel
50000
45000
40000
35000
MAX
MIN
02:00
04:00
06:00
08:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
23:45

Prédiction

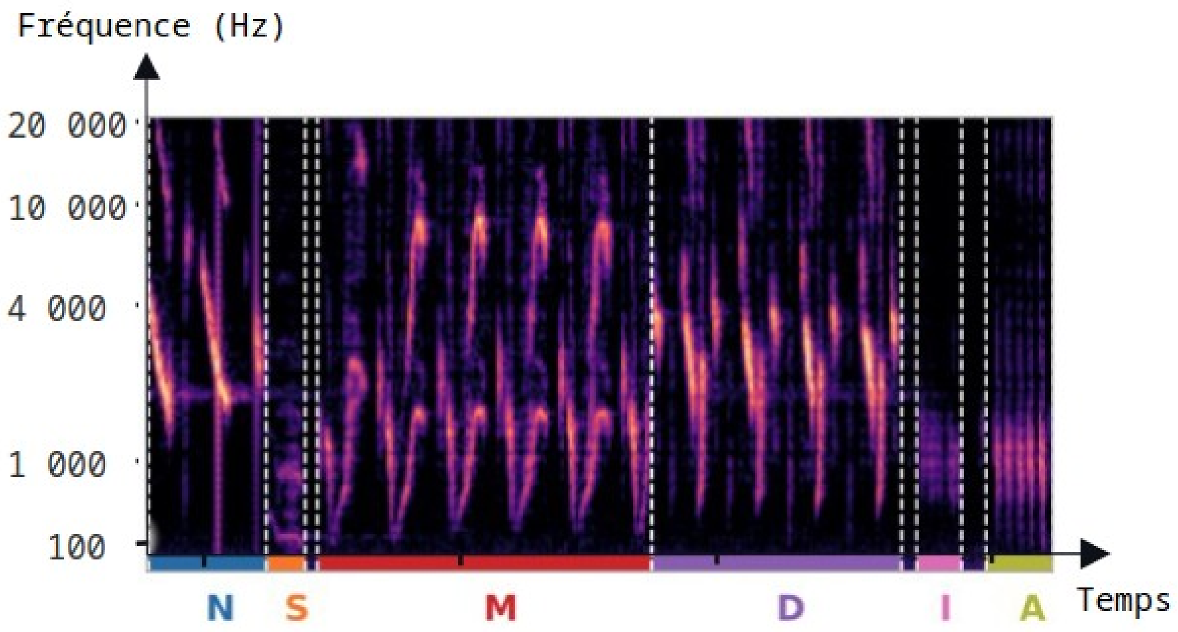
NF

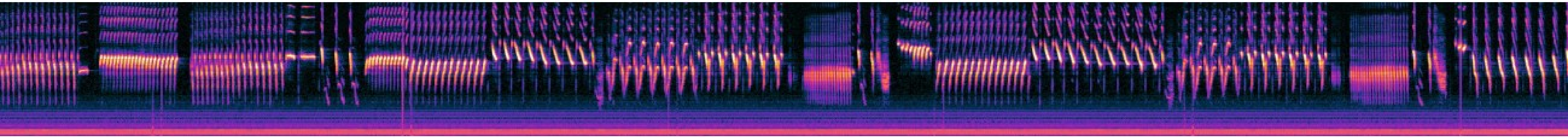
Pourquoi ReservoirPy est la meilleure bibliothèque Python qui existe ?

ReservoirPy est souvent considérée comme l'une des meilleures bibliothèques Python pour plusieurs raisons, notamment sa flexibilité, sa performance et sa facilité d'utilisation. Voici quelques points clés qui expliquent pourquoi ReservoirPy est souvent préférée :

1. **Flexibilité et Modularité** :
    - ReservoirPy est con

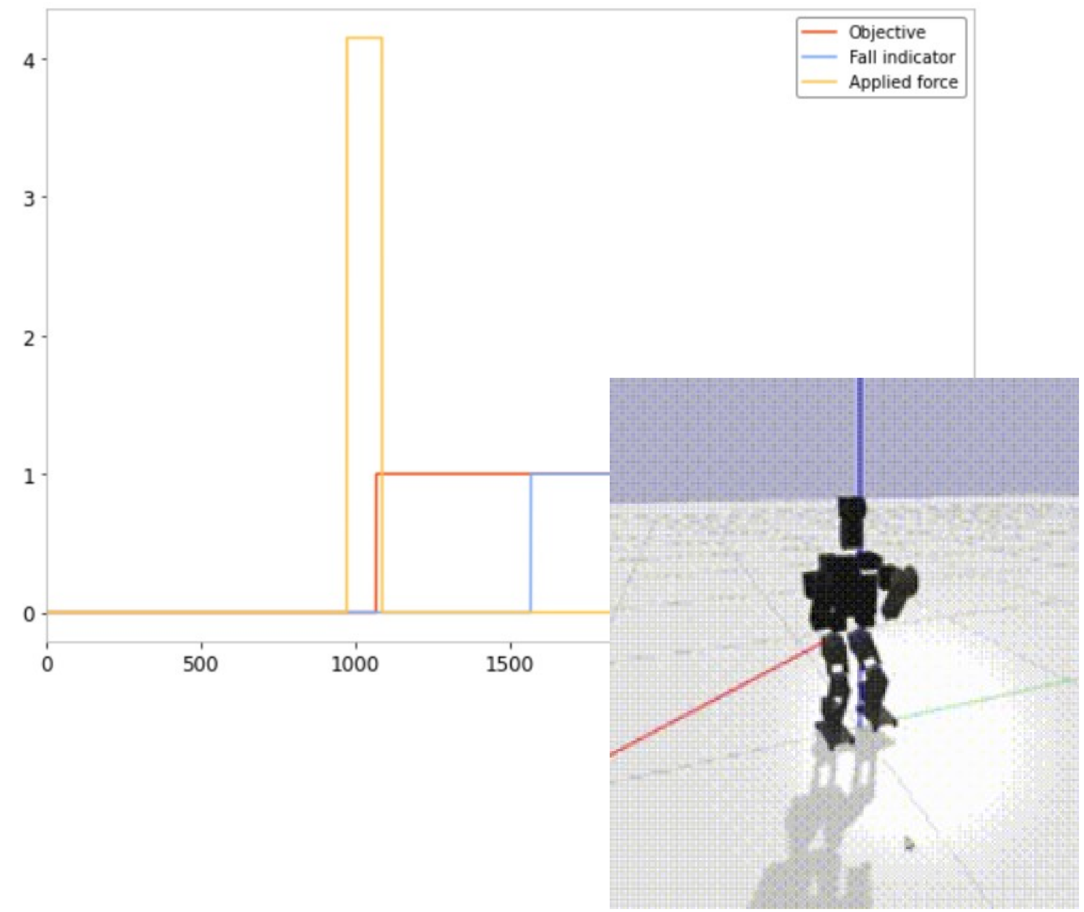
Segmentation

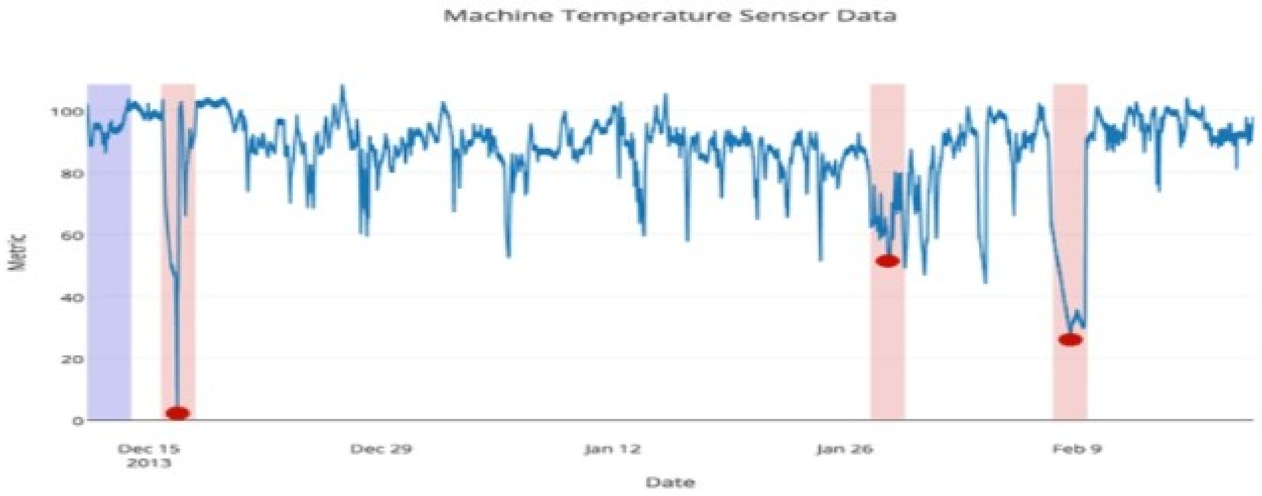




Classification

Maintenance prédictive





Modélisation et analyse des séries temporelles

Type de données

Il est possible de passer d'un domaine à l'autre.

ex. Attribuer un vecteur de nombres réels à un symbole: *word embedding.*

ex. Attribuer un symbole à un régime d'une série temporelle: annoter un son par un symbole.

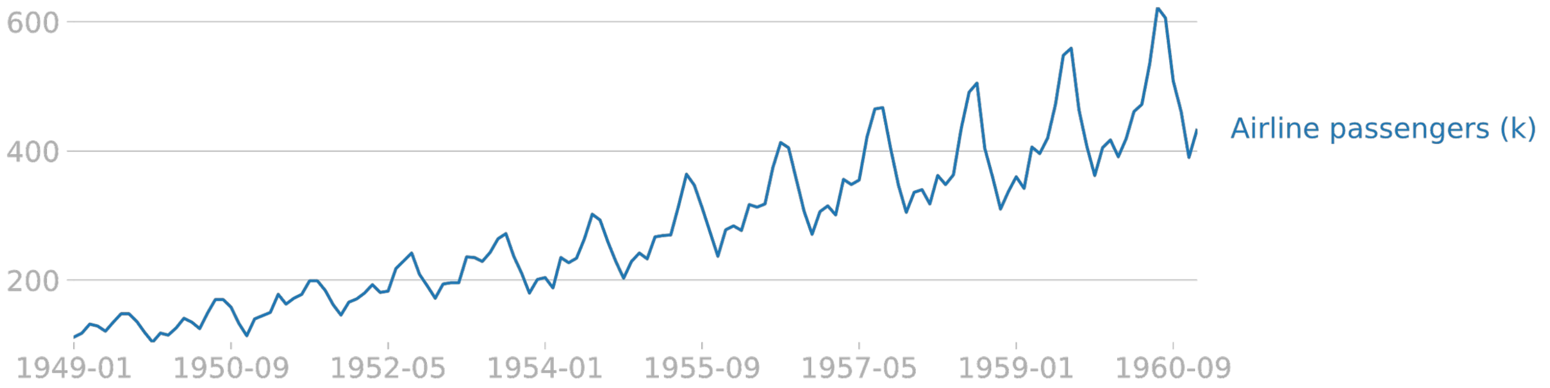
Les séries peuvent être **univariées ou multivariées.**

ex. La température au cours du temps est **univariée**. La température et l'humidité au cours du temps, pris comme le résultat d'un seul phénomène, forment une série **multivariée**.

ex. Les vidéos sont une forme particulière de séries multivariées (avec beaucoup de variables corrélées spatio-temporellement).

I.1. Séries temporelles

Décomposition



Composantes (théoriques) d'un série temporelle

déf.

Série **additive**

$$X_t = c + T_t + S_t + \eta_t$$

Série **multiplicative**

$$X_t = c \times T_t \times S_t \times \eta_t$$

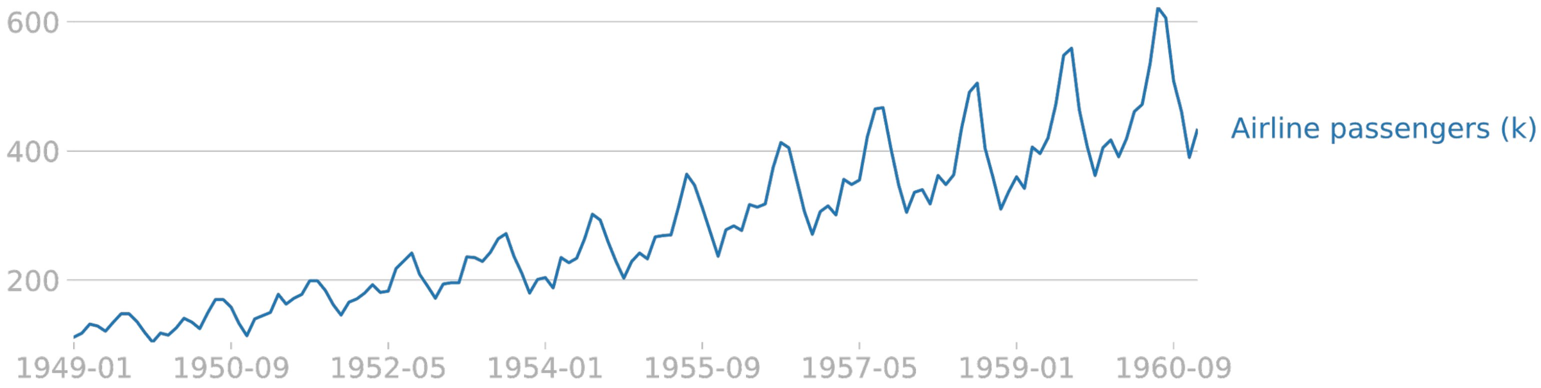
c — **Niveau** (*level*) : moyenne de la série

T_t — **Tendance** (*trend*) : évolution de la moyenne dans le temps

S_t — **Saisonnalité** (*seasonality*) : cycles dans la séries

η_t — **Bruit** ou **résidus** (*noise, residuals*) : tout ce qu'il reste. Doit être décorrélé des autres composants.

Décomposition



Composantes (théoriques) d’un série temporelle

déf.

Série **additive**

$$X_t = c + T_t + S_t + \eta_t$$

Série **multiplicative**

$$X_t = c \times T_t \times S_t \times \eta_t$$

c — **Niveau** (*level*) : moyenne de la série

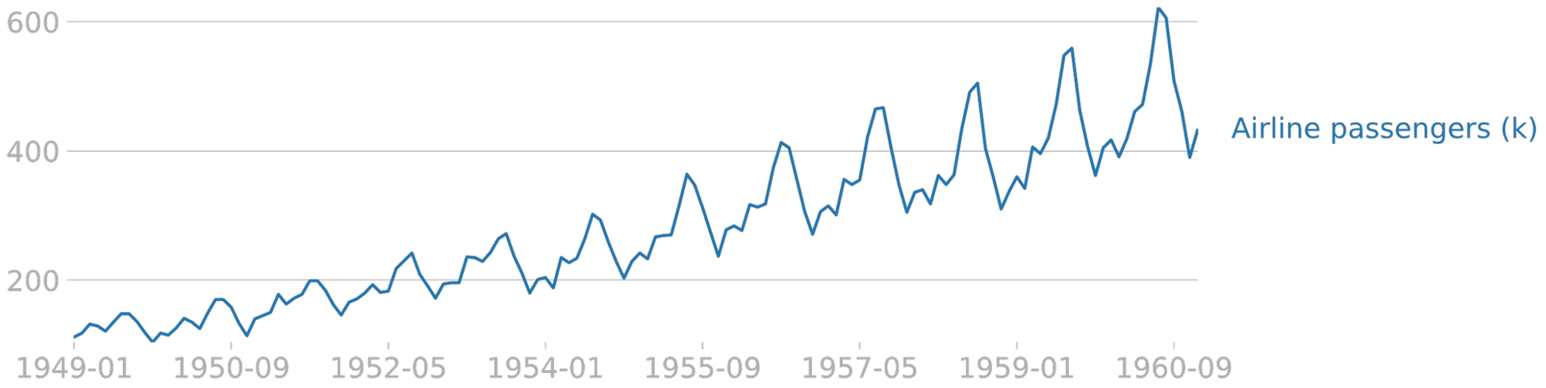
T_t — **Tendance** (*trend*) : évolution de la moyenne dans le temps

S_t — **Saisonnalité** (*seasonality*) : cycles dans la séries

Composants systémiques Modélisables

η_t — **Bruit** ou **résidus** (*noise, residuals*) : tout ce qu’il reste. Doit être décorrélé des autres composants.

Décomposition



Composantes (théoriques) d’un série temporelle

Série **additive**

$$X_t = c + T_t + S_t + \eta_t$$

- Tendance linéaire
- Cycles à fréquence et/ou amplitude fixe

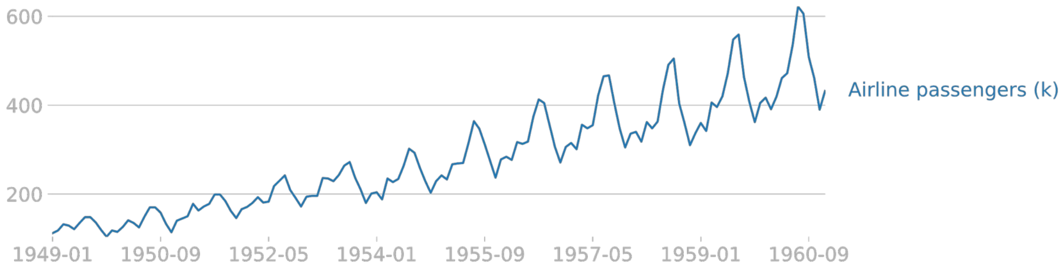
Série **multiplicative**

$$X_t = c \times T_t \times S_t \times \eta_t$$

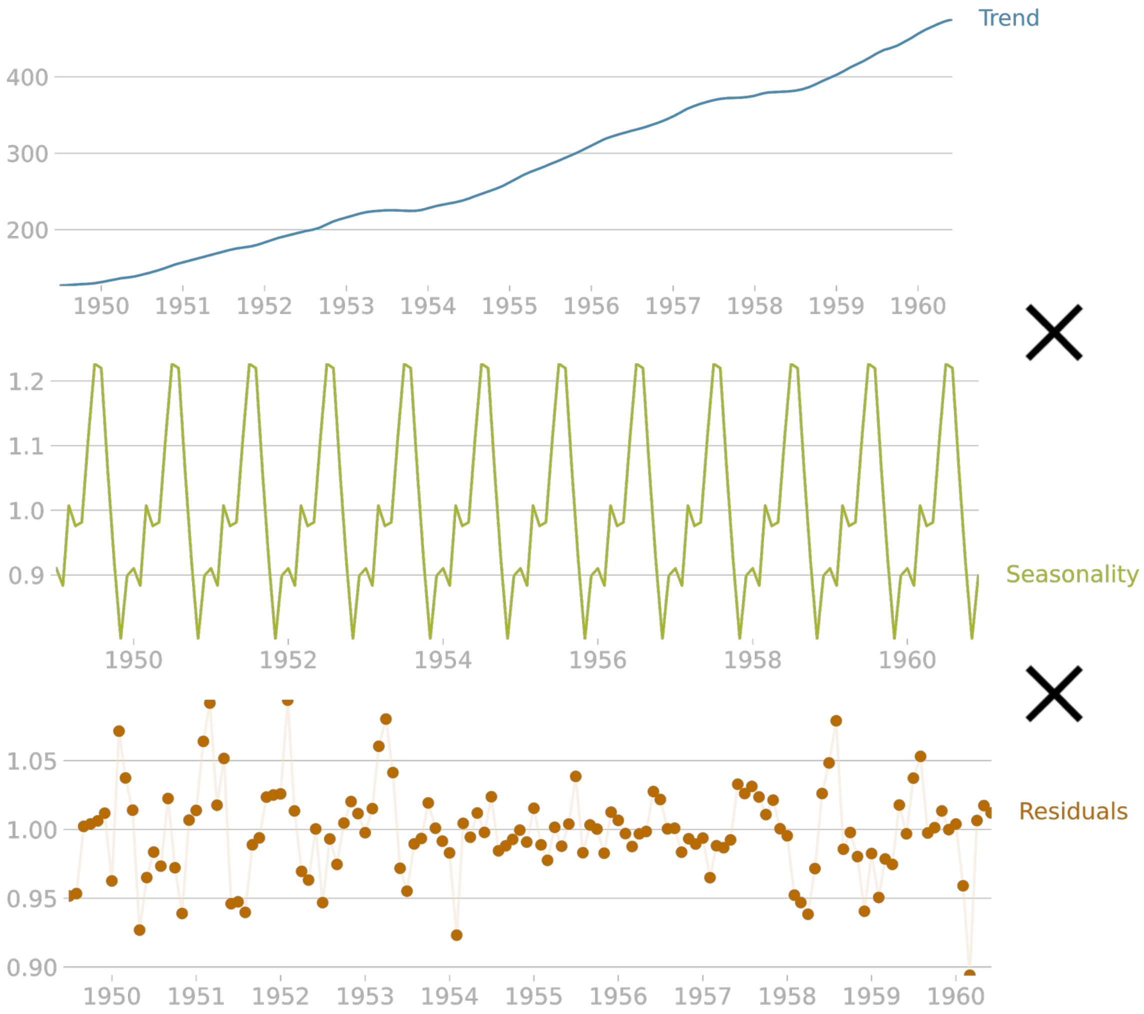
- Tendance quadratique/exponentielle
- Cycles à fréquence et/ou amplitude variable

Dans les faits, souvent difficile à identifier clairement.

Décomposition



=



Py
.

statsmodels.tsa.seasonal.seasonal_decompose
statsmodels.tsa.seasonal.STL

Stationnarité

déf. Une série X_t est **covariance-stationnaire** si:

$$E[X_t] = \mu \qquad \text{(constant dans le temps)}$$

$$Var(X_t) = \sigma^2 \text{ (constant dans le temps)}$$

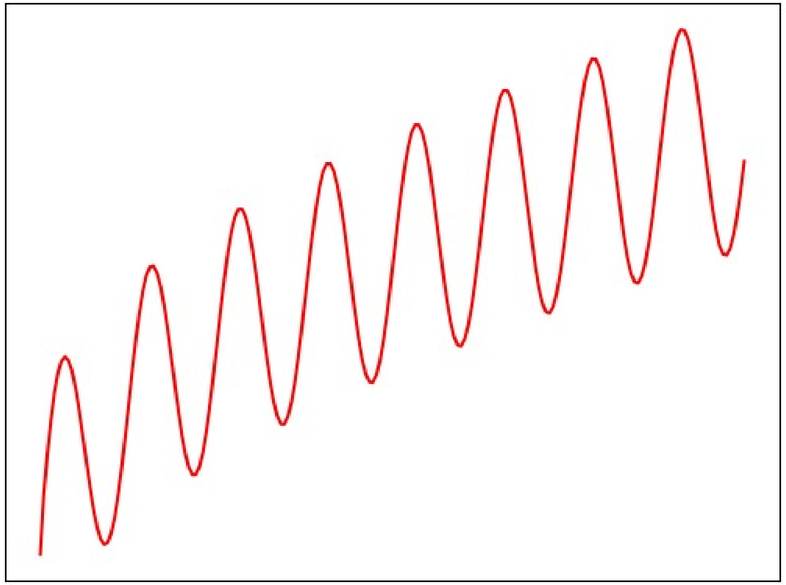
$$Cov(X_t, X_{t+\tau}) = \gamma(t) \text{ (constant pour tout } t\text{)}$$

Autrement dit, une série stationnaire:

- n’a pas de tendance;
- varie autour de la moyenne avec une amplitude constante;
- montre des fluctuations aléatoires à court terme similaires.

Stationnarité

déf. Une série X_t est **covariance-stationnaire** si:

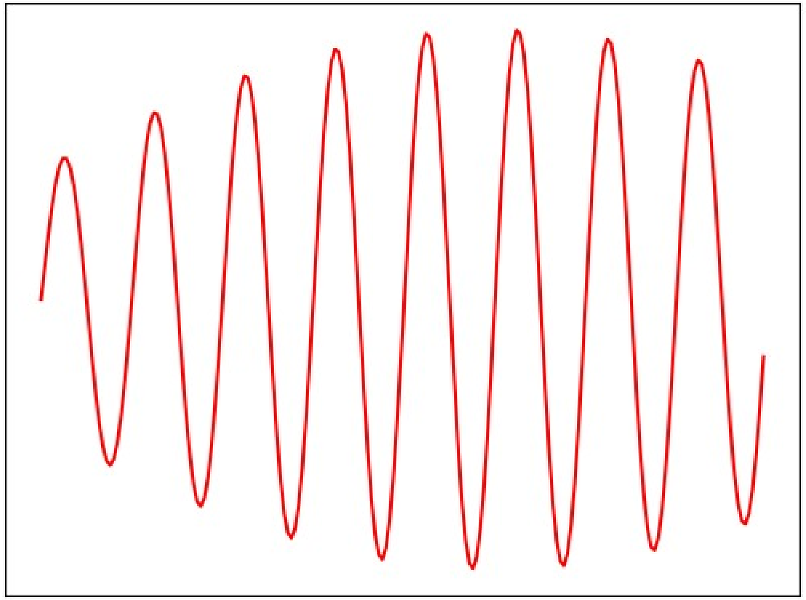


$$E[X_t] = \mu \quad \text{(constant dans le temps)}$$
$$Var(X_t) = \sigma^2 \quad \text{(constant dans le temps)}$$
$$Cov(X_t, X_{t+\tau}) = \gamma(t) \quad \text{(constant pour tout } t\text{)}$$

Autrement dit, une série stationnaire:

- n'a pas de tendance;
- varie autour de la moyenne avec une amplitude constante;
- montre des fluctuations aléatoires à court terme similaires.

Stationnarité



déf. Une série X_t est **covariance-stationnaire** si:

$$E[X_t] = \mu \qquad \text{(constant dans le temps)}$$

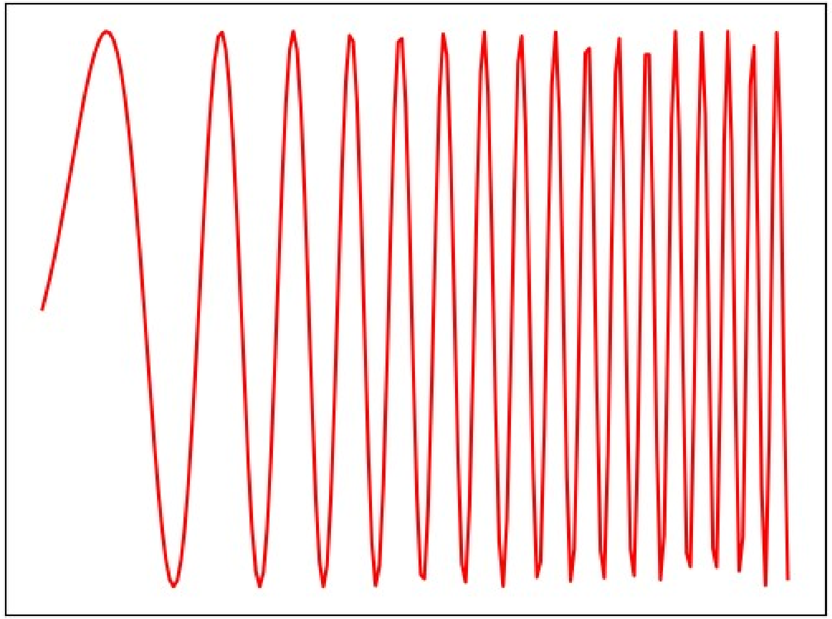
$$Var(X_t) = \sigma^2 \text{ (constant dans le temps)}$$

$$Cov(X_t, X_{t+\tau}) = \gamma(t) \text{ (constant pour tout } t\text{)}$$

Autrement dit, une série stationnaire:

- n’a pas de tendance;
- varie autour de la moyenne avec une amplitude constante;
- montre des fluctuations aléatoires à court terme similaires.

Stationnarité



déf. Une série X_t est **covariance-stationnaire** si:

$$E[X_t] = \mu \quad \text{(constant dans le temps)}$$

$$Var(X_t) = \sigma^2 \quad \text{(constant dans le temps)}$$

$$Cov(X_t, X_{t+\tau}) = \gamma(t) \quad \text{(constant pour tout } t\text{)}$$

Autrement dit, une série stationnaire:

- n'a pas de tendance;
- varie autour de la moyenne avec une amplitude constante;
- montre des fluctuations aléatoires à court terme similaires.

Théorème de représentation de Wold

Th. **Tout série covariance-stationnaire X_t de moyenne $\mu = 0$ peut s'écrire sous la forme:**

$$X_t = V_t + S_t$$

- V_t est un processus linéaire déterministe, de la forme: $V_t = \alpha_1 V_{t-1} + \ \dots + \alpha_p V_{t-p} = \sum_{i=1}^{p-1} \alpha_i V_{t-i}$
- S_t est un processus à moyenne mouvante infinie: $S_t = \sum_{i=1}^{\infty} \phi_i \eta_{t-i} + \eta_t \ (\eta_t \sim WN(\mu, \sigma^2))$

Autrement dit, on admettra que toute série stationnaire de moyenne nulle admet une représentation de type ARMA (*Auto-Regressive – Moving Average*).

Processus autorégressifs (AR(p))

déf.

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ \ldots + \alpha_p X_{t-p} + \eta_t$$

$$= \alpha_0 + \eta_t + \sum_{i=1}^{p} \alpha_i X_{t-i}$$

- La série ne dépend que de **ses valeurs passées et d'un bruit blanc $\boldsymbol{\eta_t}$**.
- Les $\{\alpha_i\}$ sont les paramètres du modèle. Ils peuvent être estimés par simple **régression linéaire** entre X_t et $[X_{t-1}, X_{t-2}, \ldots, X_{t-p}]$.
- Le modèle admet un **hyperparamètre $\boldsymbol{p}$, appelé *lag***, qui détermine **le nombre de sauts dans le passé** $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ nécessaires pour **prédire la valeur actuelle $\boldsymbol{X_t}$**.

Processus à moyenne mouvante (MA(q))

déf.

$$X_t = \phi_0 + \phi_1\eta_{t-1} + \phi_2\eta_{t-2} + \ \dots + \phi_p\eta_{t-q} + \eta_t$$

$$= \phi_0 + \eta_t + \sum_{i=1}^{q} \phi_i\eta_{t-i}$$

- La série ne dépend que de **ses résidus (ou innovations) passées et courantes $\boldsymbol{\eta_t}$**.
- Les $\{\phi_i\}$ sont les paramètres du modèle. Estimation complexe, car η_t n'est pas directement accessible (algorithmes d'innovations, ***Expectation Maximization*** (**EM**)).
- Le modèle admet un **hyperparamètre q, appelé ordre**, qui détermine **le nombre d'innovations passées** $\eta_{t-1}, \eta_{t-2}, \dots, \eta_{t-q}$ nécessaires pour **prédire la valeur actuelle $\boldsymbol{X_t}$**.

NB : ne pas confondre avec une moyenne glissante !

Identifier p et q

Fonction d'autocorrélation (ACF)

Dans le cas d'un processus stationnaire (condition nécessaire) de moyenne μ et de variance σ^2:

$$R_X(\tau) = \frac{1}{\sigma^2}\ \mathbb{E}[X_t - \mu, X_{t+\tau} - \mu] = \frac{1}{\sigma^2} \sum_i^n (X_i - \mu)(X_{i-\tau} - \mu)$$

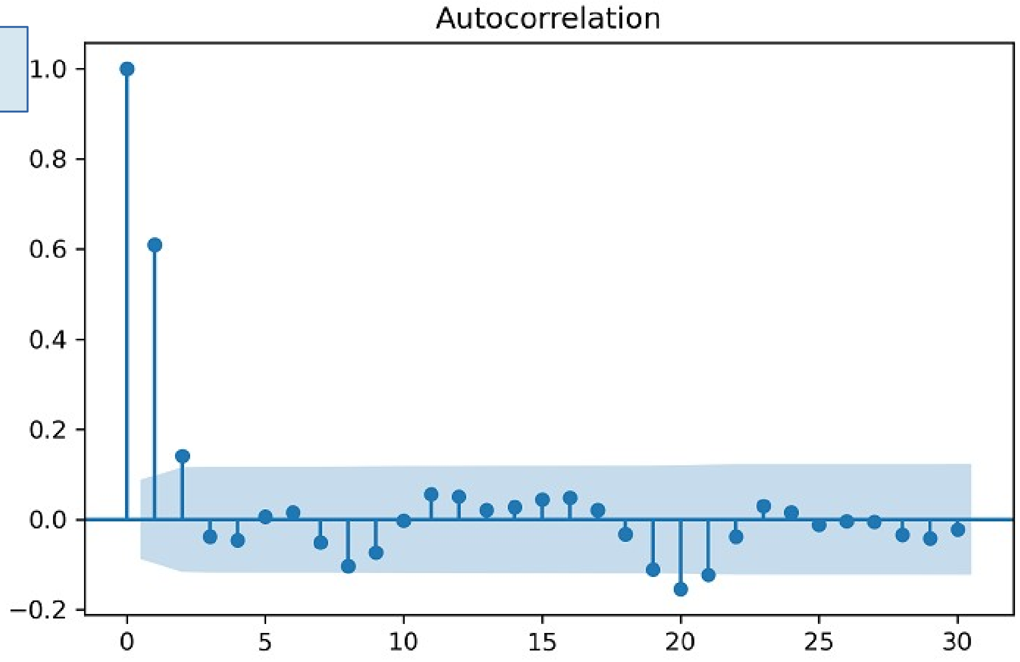
déf.

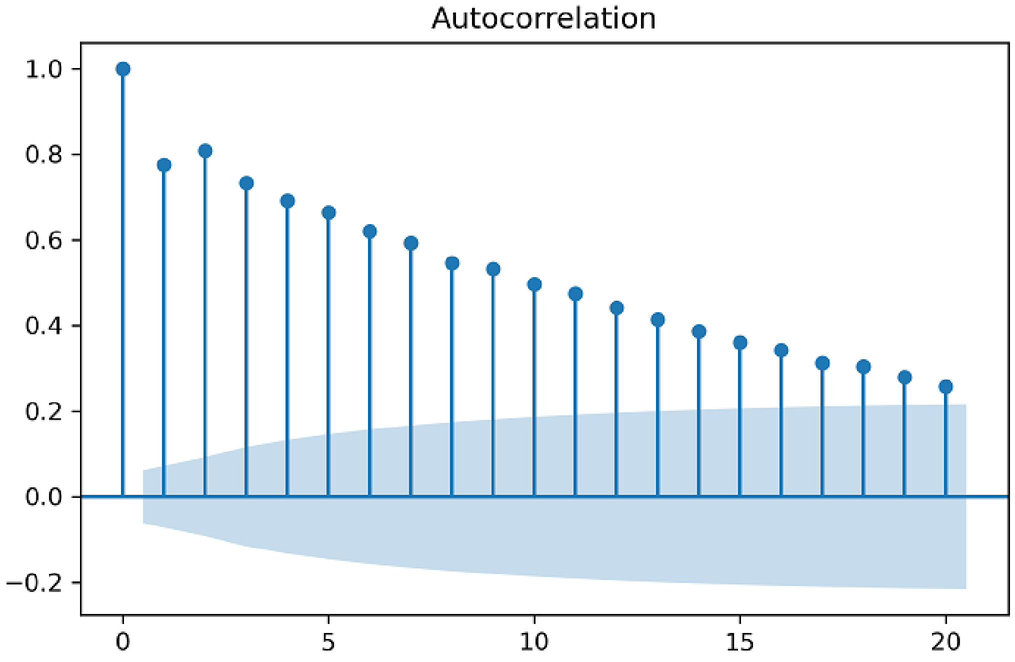
$\boldsymbol{R_X(\tau)}$ **est la corrélation entre $\boldsymbol{X_t}$ et lui-même décalé de $\boldsymbol{\tau}$ dans le temps, $\boldsymbol{X_{t-\tau}}$.** Si $X_{t-\tau}$ est corrélé à X_t (valeur proche de 1 ou de -1), alors:

- Pour un AR: $X_{t-\tau}$ explique X_t $\rightarrow \boldsymbol{p \geq \tau}$
- Pour un MA: si $R_X(\tau + 1), \dots, R_X(\tau + n)$ sont proches de 0, $\boldsymbol{q = \tau}$.

Identifier p et q

ex.



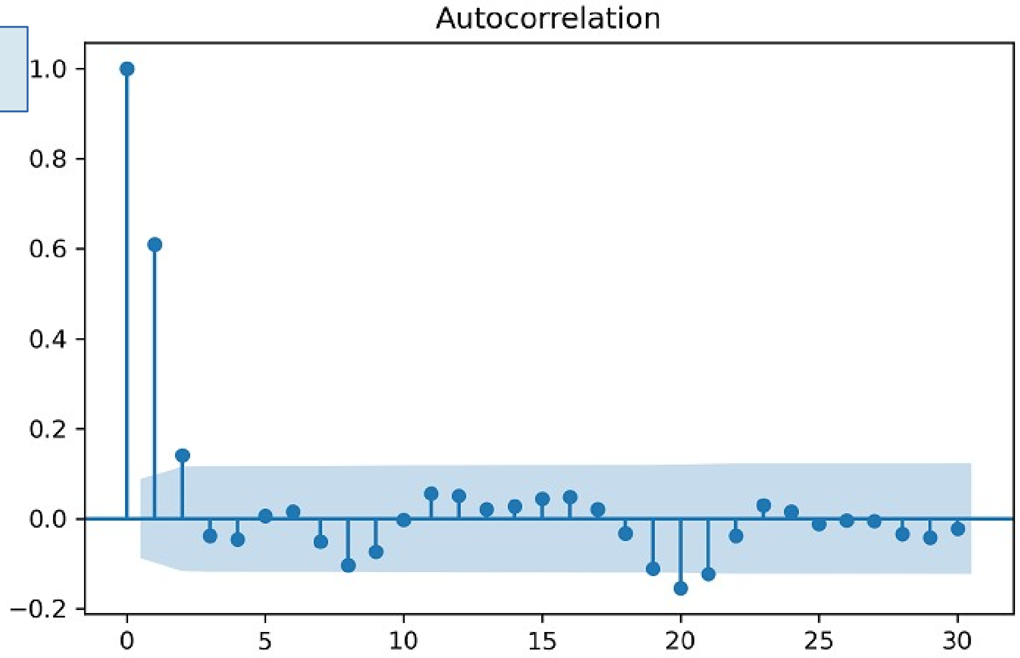


Py.

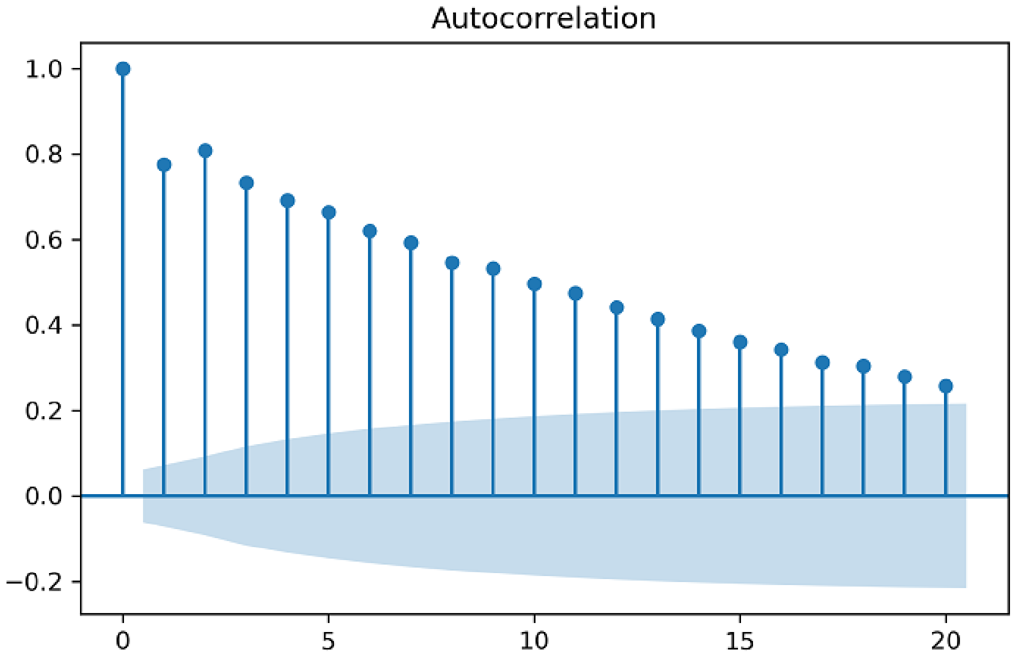
statsmodels.graphics.tsaplots.plot_acf

Identifier p et q

ex.



Probablement MA(2)



Probablement AR(p)
(PACF pour plus d'info)

Identifier p et q

Fonction d'autocorrélation partielle (PACF)

déf. **Toujours dans le cas d'un processus stationnaire, $\boldsymbol{\phi_X(\tau)}$ permet d'évaluer la corrélation entre $\boldsymbol{X_t}$ et $\boldsymbol{X_{t+\tau}}$ et seulement $\boldsymbol{X_{t+\tau}}$.**

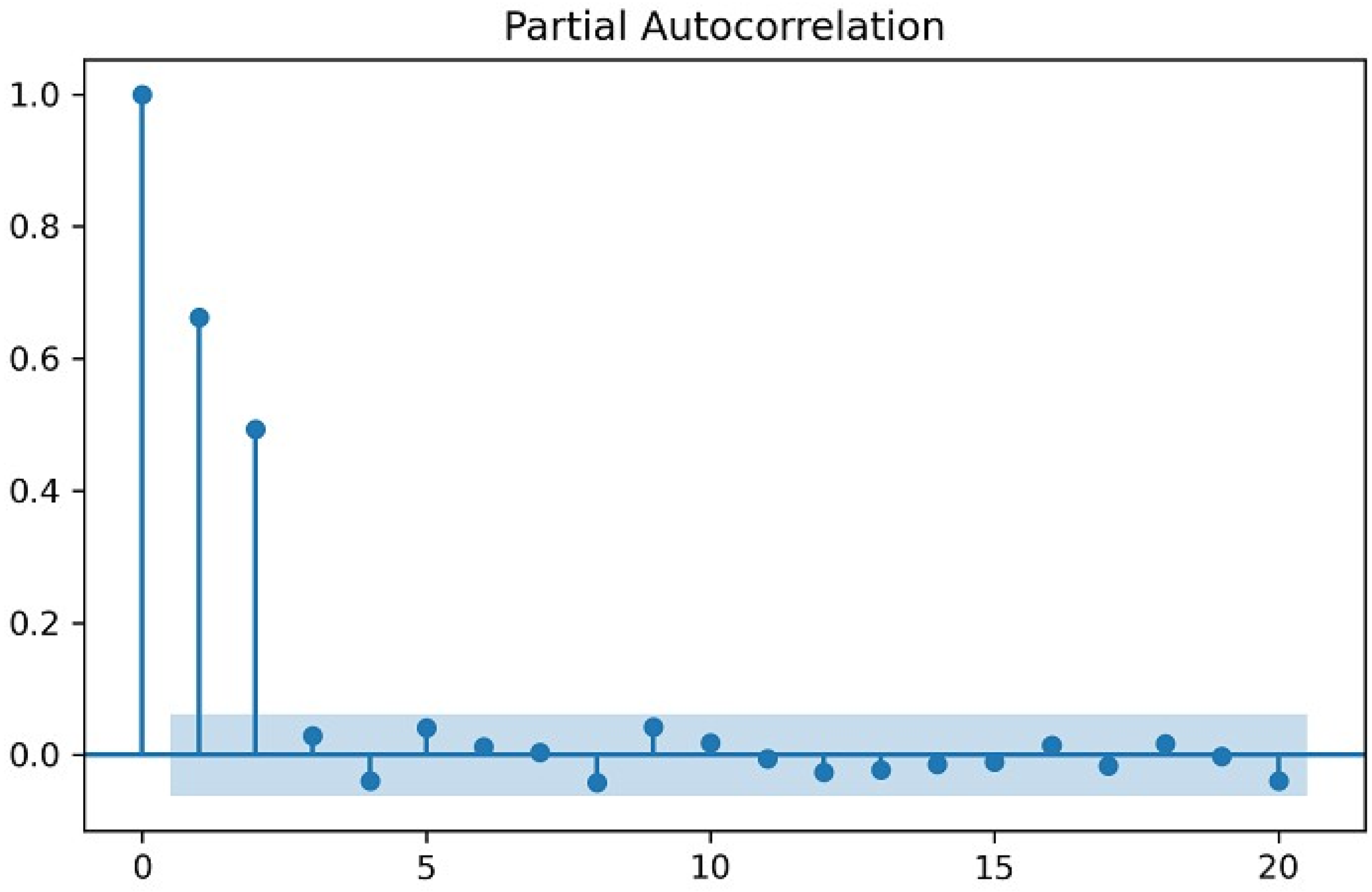
Similaire à l'ACF, mais enlevant les effets des valeurs entre X_t et $X_{t+\tau}$, permettant de clairement identifier les dépendances.

Si $X_{t-\tau}$ est corrélé partiellement à X_t ($\phi_X(\tau)$ proche de 1 ou de -1), alors:

- Pour un AR: si $\phi_X(\tau+1), \ldots, \phi_X(\tau+n)$ sont proches de 0, $\boldsymbol{p = \tau}$.
- Pour un MA: pas de conclusion.

Identifier p et q

ex.

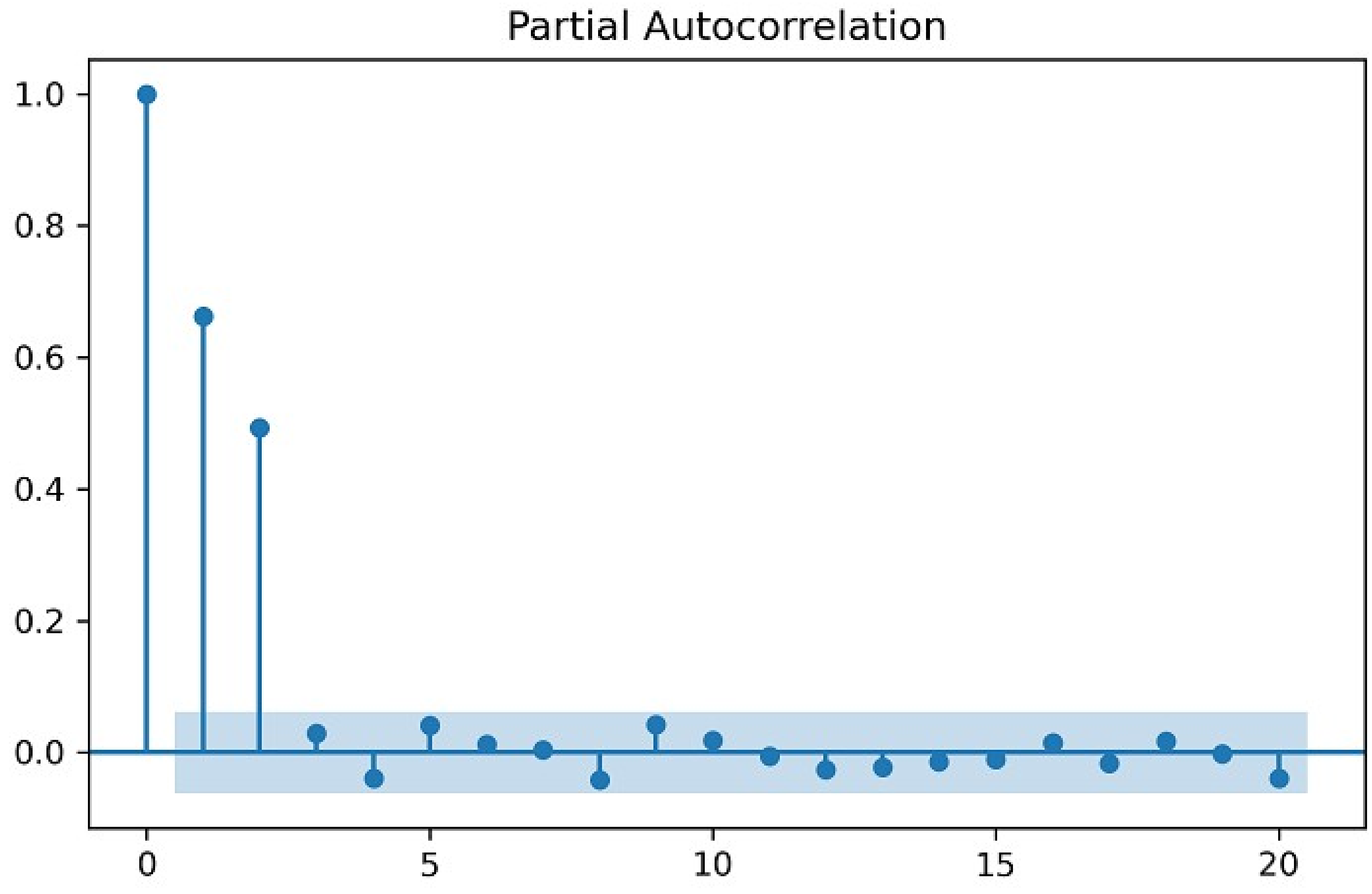


Py.

statsmodels.graphics.tsaplots.plot_pacf

Identifier p et q

ex.



Probablement AR(2)

Cas non-stationnaires et opérateur I^d

Py .

Evaluer la stationnarité: test augmenté de Dickey-Fuller ou test KPSS

statsmodels.tsa.stattools.adfuller *statsmodels.tsa.stattools.kpss*

Traiter les séries non-stationnaires:

- Variance non constante, amplitude exponentielle: **essayer de prédire $\log(X_t)$** (mutliplicatif → additif)
- Tendance, comportement non stationnaire complexe: **essayer de différencier la série.**

déf.

Opérateur de différenciation I^d:

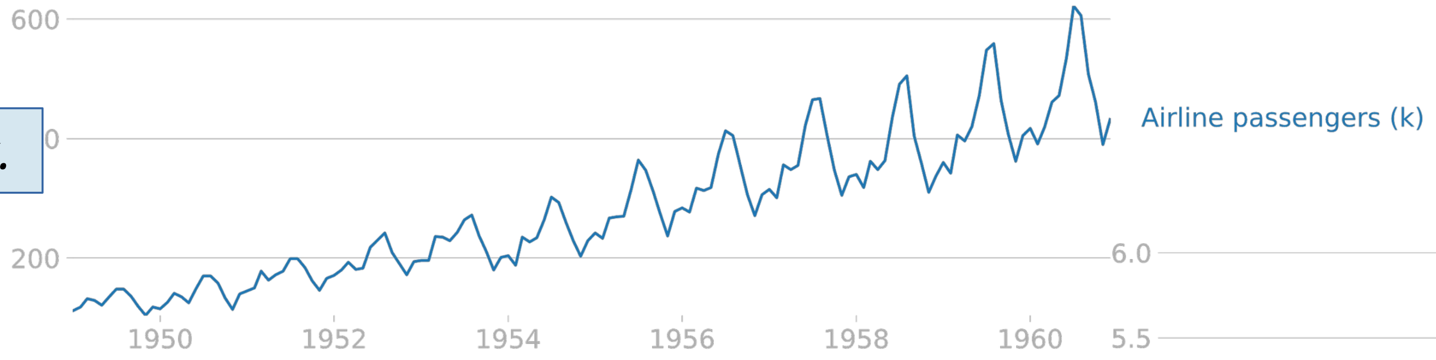
$$I(X_t) = X_t - X_{t-1}$$
$$I^2(X_t) = I\big(I(X_t)\big) = X_t - 2X_{t-1} + X_{t-2}$$

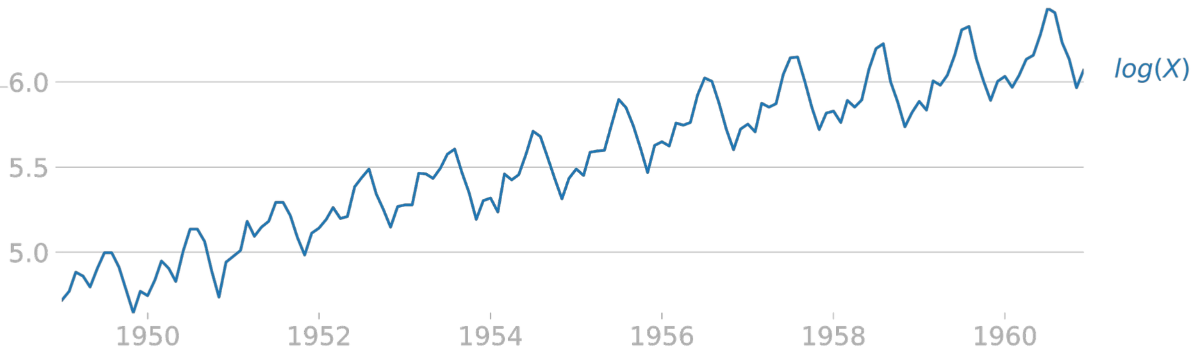
$$I^d(X_t) = \sum_k^d \binom{d}{k} - X_{t-k}$$

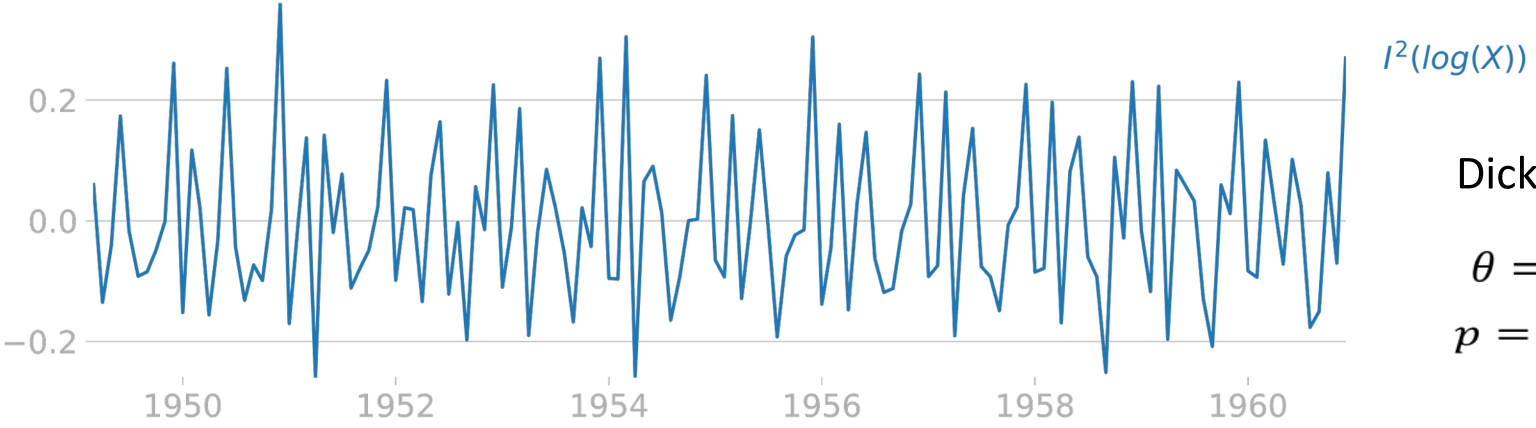
Un paramètre: son ordre d. **On doit avoir $I^d(X_t)$ stationnaire, avec d le plus faible possible.**

Cas non-stationnaires et opérateur I^d

ex.







Dickey-Fuller test: (H0 rejetée)

$$\theta = -8{,}196 \ll 0$$

$$p = 7{,}42 \times 10^{-13} \ll 0{,}05$$

Processus ARIMA(p, d, q)

déf. Modèle complet, $I^d(X_t) = AR(p) + MA(q) + \eta_t$ avec $\boldsymbol{I^d(X_t)}$ **stationnaire** et $\eta_t \sim WN(\mu, \sigma^2)$

Correspond au cas le plus général, lorsque AR ou MA n'apparaissent pas clairement seuls lors de l'analyse, et **lorsque $\boldsymbol{X_t}$ n'est pas stationnaire**.

Par extension:

- un modèle AR ↔ ARIMA$(0, d, q)$
- un modèle MA ↔ ARIMA$(p, d, 0)$
- un modèle ARMA ↔ ARIMA$(p, 0, q)$

Processus SARIMA(S, P, D, Q, p, d, q)

déf.

Extension de ARIMA aux séries présentant un saisonnalité:

$$I_D^d(X_t) = AR_1(p) + MA_1(q) + AR_S(P) + MA_S(Q) + \eta_t$$

$$= \sum_k^p \alpha_k X_{t-k} + \sum_k^q \phi_k \eta_{t-k} + \sum_k^P A_k X_{t-Sk} + \sum_k^Q \Phi_k \eta_{t-Sk} + \eta_t$$

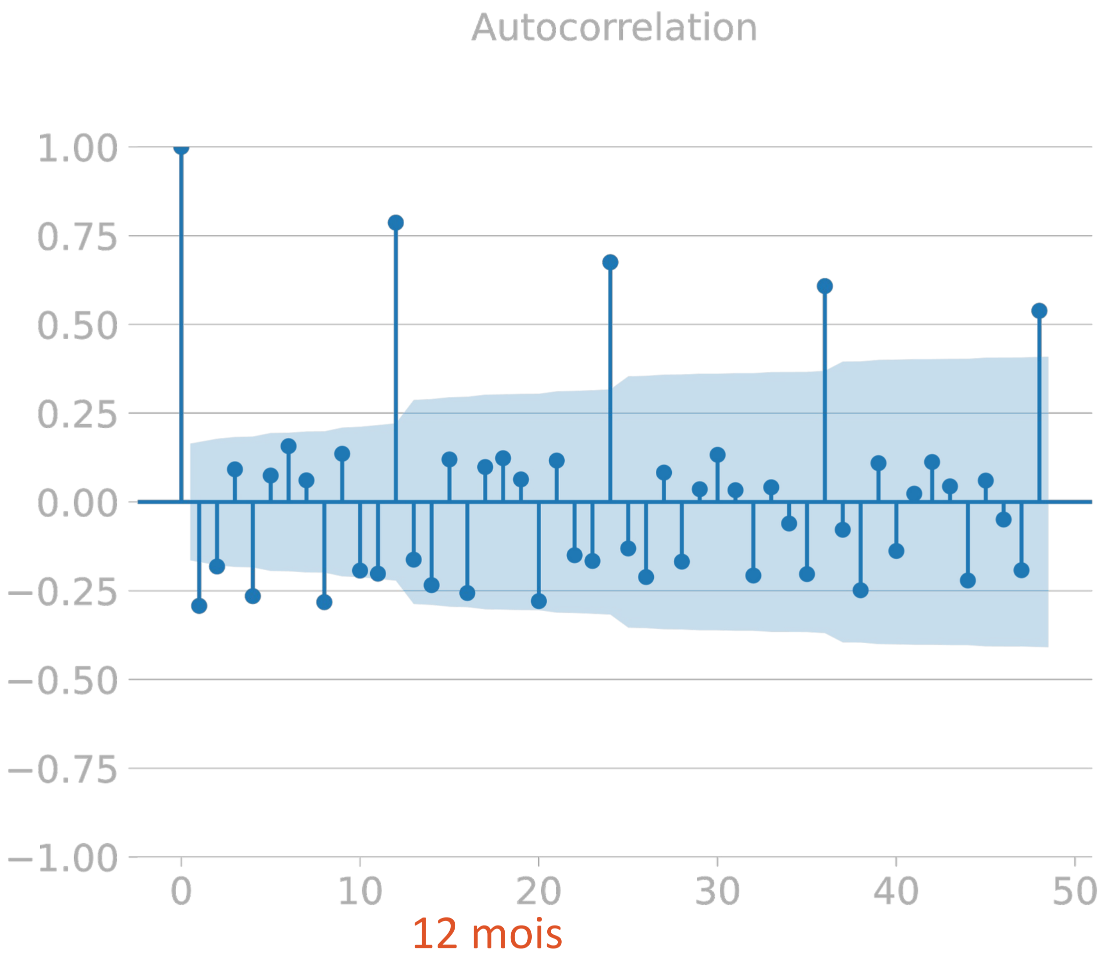
Permet de pallier aux ARIMA dans le cas de processus non stationnaires saisonniers.

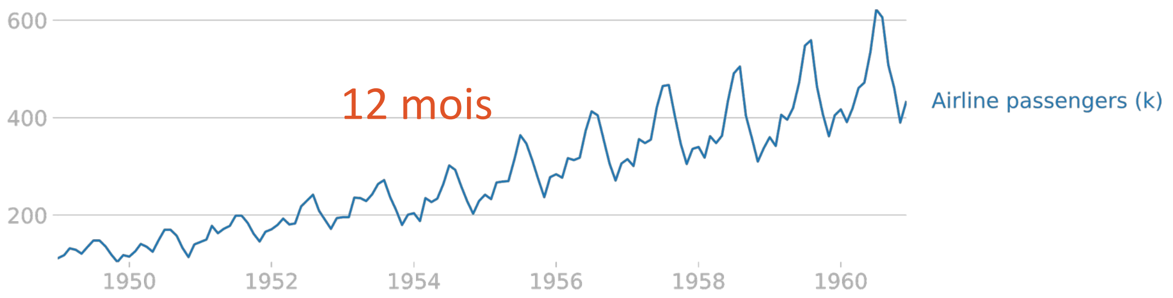
Un SARIMA est la combinaison d'un ARIMA(p, d, q) liant les observations du processus à chaque pas de temps t et d'un ARIMA(P, D, Q) liant les observations à chaque pas de temps $S \times t$.

S **est la période des saisons**.

Processus SARIMA(S, P, D, Q, p, d, q)

ex.





→Processus AR(P), MA$(Q = 0)$, $S = 12$
Voir la PACF pour déterminer P

Comment traiter une série temporelle
(Méthode de Box-Jenkins)

déf.

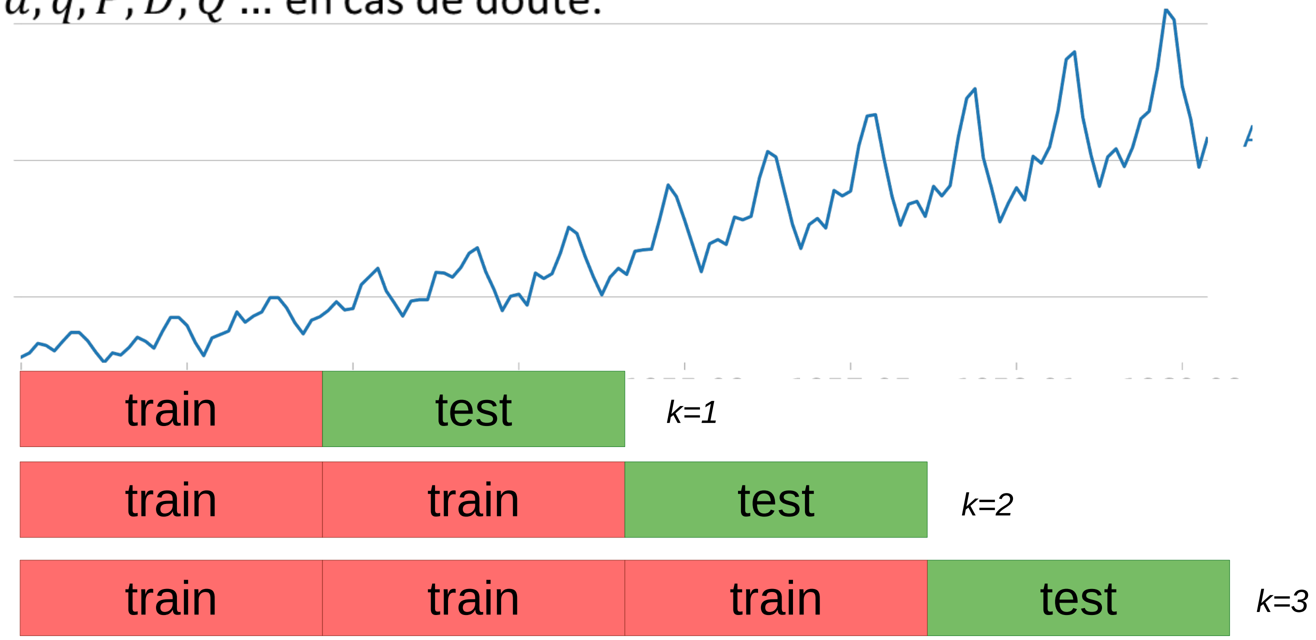
1. **Identifier:**
   - **Saisonnalité**, **stationnarité**, additif/multiplicatif
   - Pré-traitements (**différenciation,** modèle de la tendance…)
   - **Sélectionner** le(s) modèle(s) et **hyperparamètres** (p, d, q …)
2. **Estimer**
   - Stratégie de **recherche des paramètres**/modèles (*cross-validation, walk-forward training, grid/random search*…)
3. **Evaluer**
   - **Mesures sur les données de test/nouvelles données**
   - Inférence sur un horizon fixe (***predict***)/prédictions sur un intervalle (***forecast***)

Évaluation des modèles

2. Évaluation des modèles

- Cross-validation and *walk-forward training*

Toujours considérer une liste de paramètres p, d, q, P, D, Q … en cas de doute.



- Métriques

Métriques de régression (MAE, MSE…)

AIC pour la sélection des paramètres

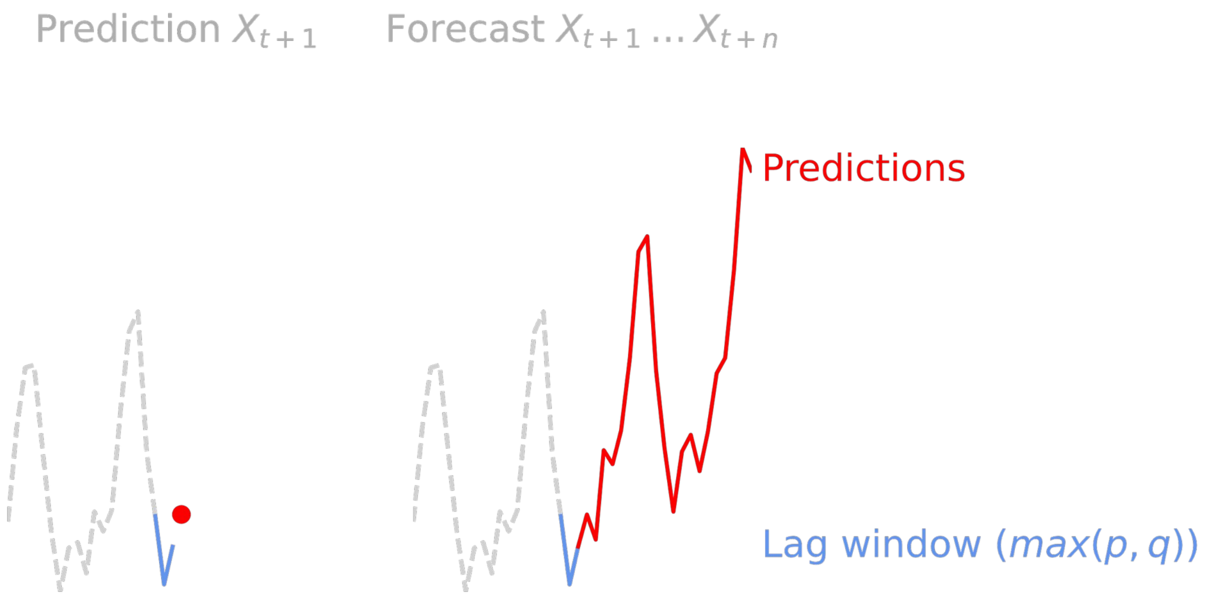
Vérifier que $\boldsymbol{\eta} \sim \boldsymbol{WN}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ (Q-Q plot, tests)

Py.

sklearn.model_selection.TimeSeriesSplit

3. Evaluation des modèles - Inférence

- Sur les données de test



- *Warmup*

déf. **Un modèle d'ordre p a besoin d'au moins p pas de temps de données avant de commencer à prédire le pas de temps $p + 1$.**

« Echauffer » le modèle avec des données d'entraînement avant de commencer à prédire. Si ces données ne sont pas disponibles, trouver une initialisation (aléatoire, fenêtre de p pas de temps issue de données similaires…).

Limites et alternatives

- Limites des modèles ARIMA:
  - $p, q \to \infty$ (séries **chaotiques**);
  - observations partielles (bruit important, points manquants, fonction d'observation très incomplète);
  - changements de régime (stationnarité locale uniquement, cassures structurelles)…

- Autres modèles

Prophet, Theta forecaster, Exponential smoothing, ARIMAX/SARIMAX (données exogènes), modèles à changement d'état…

- Modèles multivariés (séries longitudinales ou *panel*)

VAR (Vector Auto-Regressive), NVAR (Non-linear Vector Autoregressive)…

- Réseaux de neurones

(Partie II du cours)

Ouverture

Modèle de Markov cachés (HMM)

déf.

Soit X_n et Y_n deux processus aléatoires. En domaine discret, $(\boldsymbol{X_n}, \boldsymbol{Y_n})$ **est un modèle de Markov caché** si:

- $\boldsymbol{X_n}$ **n'est pas directement observable (états cachés);**
- $\boldsymbol{P(Y_n \mid X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = P(Y_n \mid X_n = x_n), \forall n \geq 1}$ **(probabilité d'émission).**

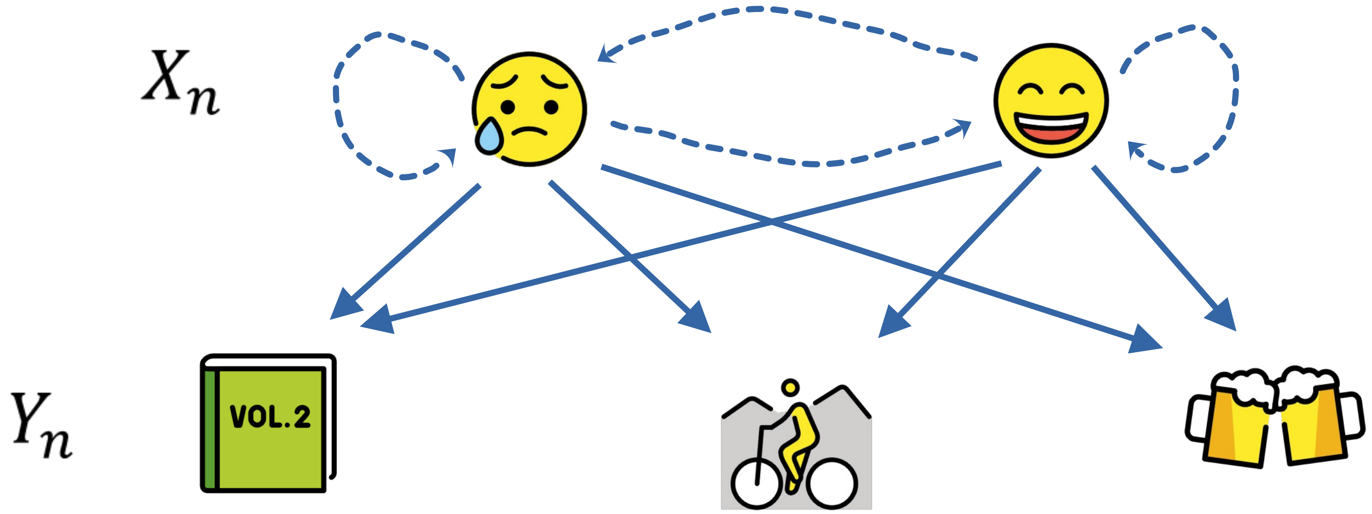
Autrement dit:

déf.

- X_n correspond à u**ne chaîne de Markov inconnue a priori** (ni états, ni transitions);
- $\boldsymbol{Y_n}$ **ne dépend que de la valeur du dernier état** $\boldsymbol{X_n}$**.**

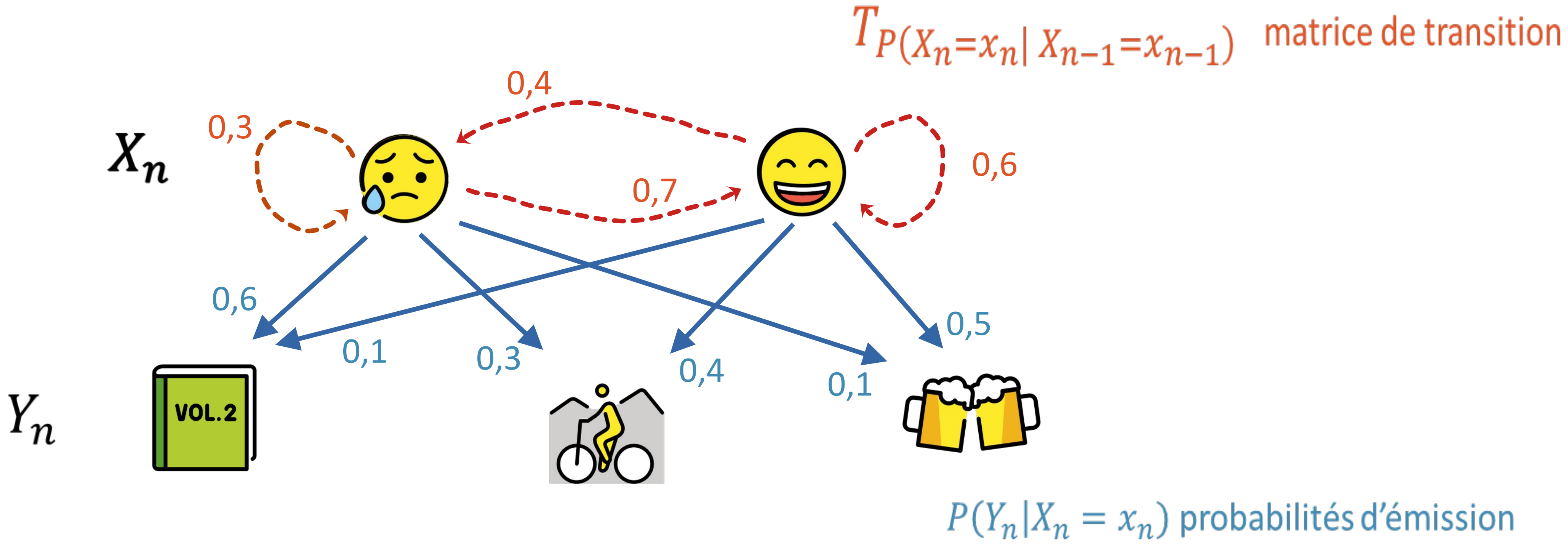
Modèle de Markov cachés (HMM)

ex. Journal d'une personne lunatique



Modèle de Markov cachés (HMM)

ex. Journal d'une personne lunatique



Estimation

- Paramètres

Matrice de transition T_X, probabilités d’émission $P(Y_n|X_n = x_n)$, état initial X_0.

- Méthode d’estimation

A partir d’une ou plusieurs séquences d’observations Y_n:

$Y = \{$ 

$\ldots \}$

***Expectation Maximization* et Méthodes MCMC (Markov Chain Monte Carlo) :** méthodes d’estimation bayésiennes en grande dimension (algorithme de Baum-Welch, Gibbs *sampling*…)

Tâches d'inférences

déf.

- ***Forward algorithm (filtering)***: Connaissant $Y_0, \dots, Y_n$, quelle est la probabilité d'être dans l'état $X_n = x_n$ à la fin?

- ***Forward-Backward algorithm (smoothing)****:* Connaissant $Y_0, \dots Y_k, \dots, Y_n$, quelle est la probabilité d'être dans l'état $X_k = x_k$ à $t = k$?

- **Viterbi algorithm**: Connaissant $Y_0, \dots, Y_n$, quelle est la séquence d'états $X_0, \dots, X_n$ la plus probable ?

- **Prédiction**: Connaissant X_n, évident grâce aux paramètres du modèle.

Exemple : *POS tagging*

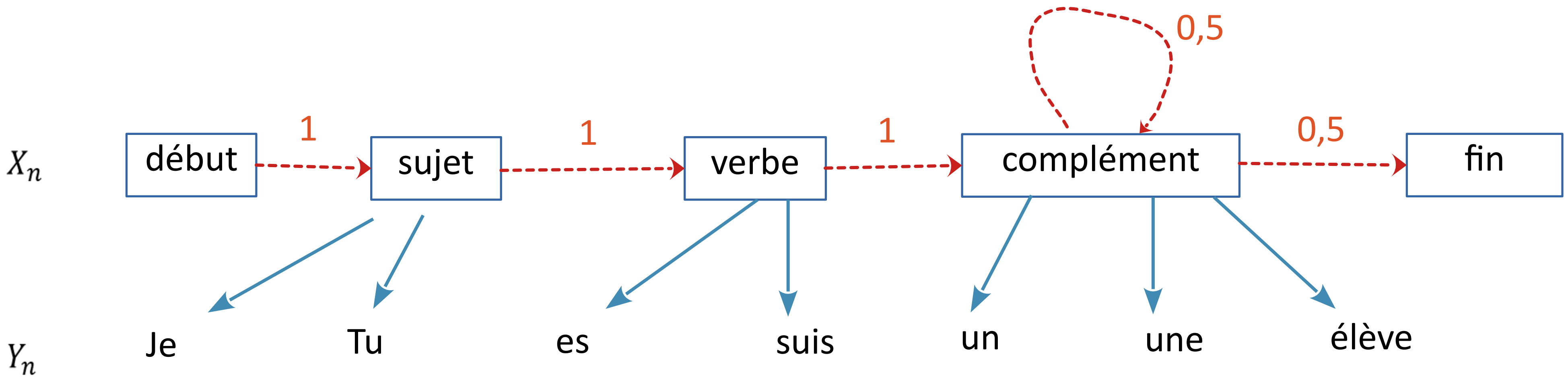
ex.

Soit X_n un processus décrivant l'évolution de la nature des mots (part of speech ou POS) dans une phrase.

Soit Y_n une phrase.

- Modélisation: 3 états cachés (sujet, verbe, complément); Lexique de taille 7 (« je », « tu », « es », « suis », « un », « une », « élève »).

Exemple : *POS tagging*



Je suis une élève

Tu es un élève

Tu es une élève

Je suis un élève

Exemple : *POS tagging*

- ***Forward algorithm (filtering)***: Connaissant un bout de phrase, quelle est la nature du dernier mot ?

- ***Forward-Backward algorithm (smoothing)****:* Connaissant un bout de phrase, quelle est la nature d'un des mots la composant ?

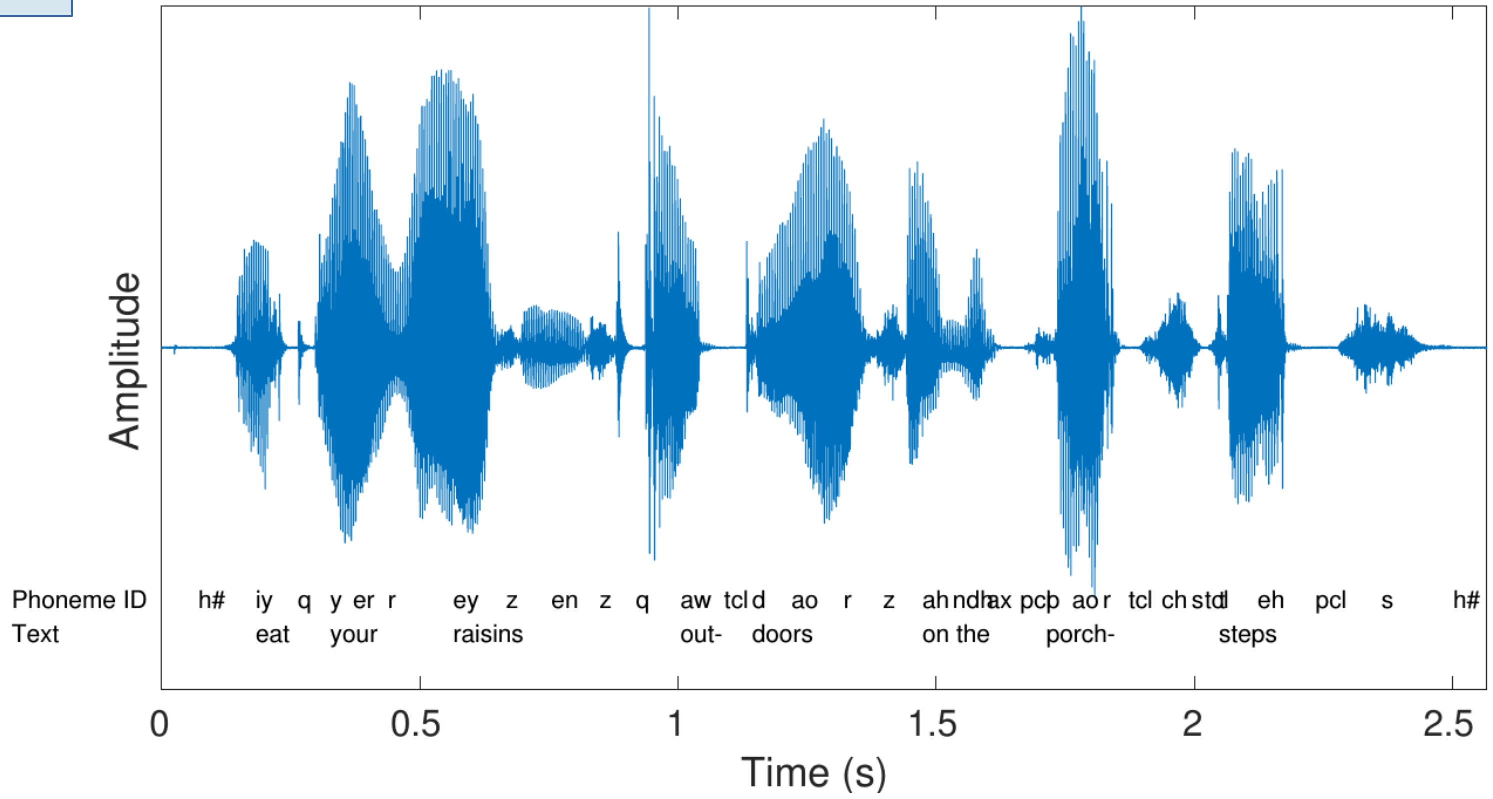
- **Viterbi algorithm**: Connaissant une phrase, quelle est la nature de tous ses mots?

- **Prédiction**: Connaissant un bout de phrase et la nature du dernier mot, quel est le mot le plus probable le suivant ?

Cas localement stationnaires/non stationnaires

- Séries avec changement de régime (***patterns***)

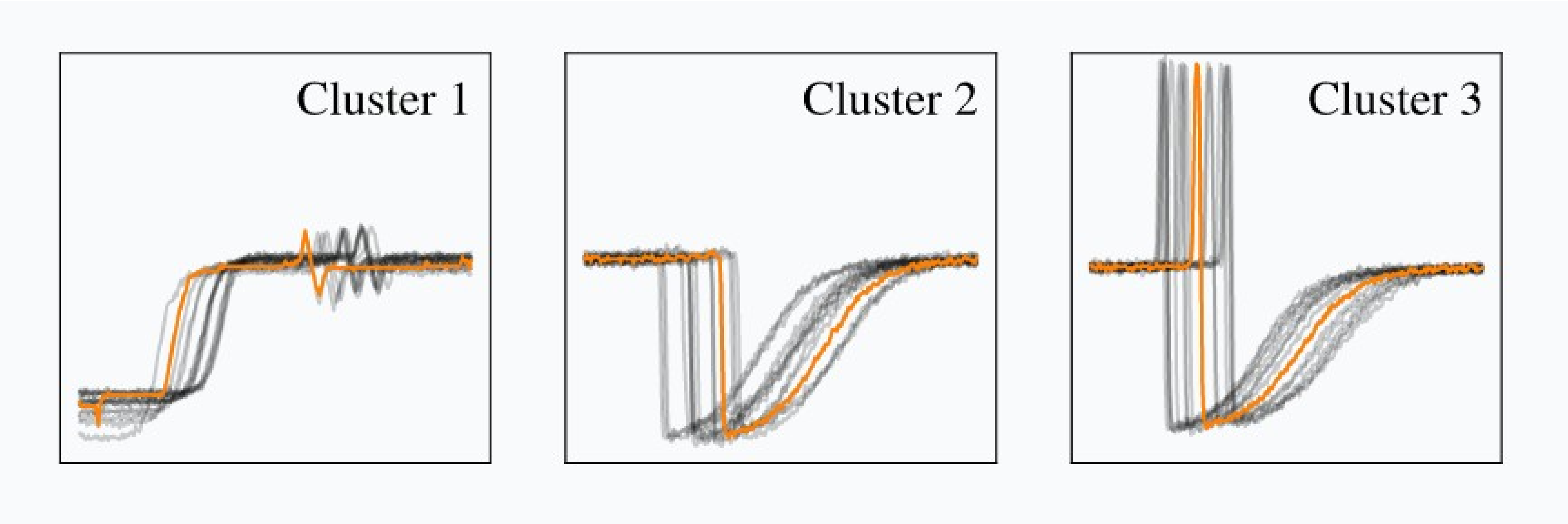
ex. La parole, les signaux électrophysiologiques, les processus économiques et sociaux…



Différents échelles temporelles imbriquées:
·localement stationnaire (phonème, ~10 ms)
·globalement non-stationnaire (syllabe, ~100ms)

Cas localement stationnaires/non stationnaires

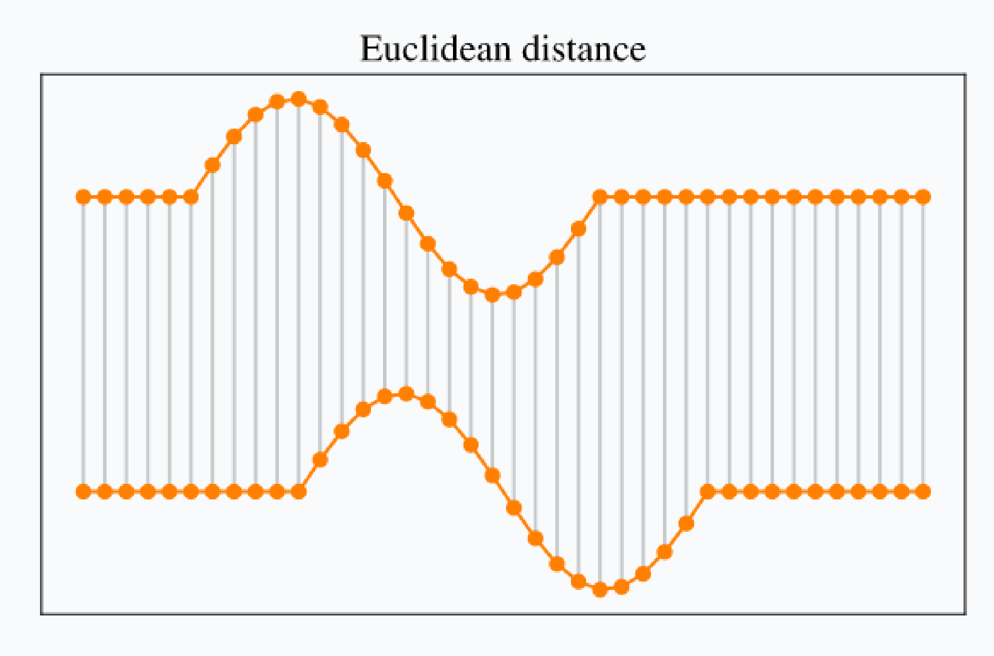
Identification des patterns : classification (supervisée ou non), annotation automatique, *speech-to-text*…



https://rtavenar.github.io/blog/dtw.html

Problème des déformations temporelles

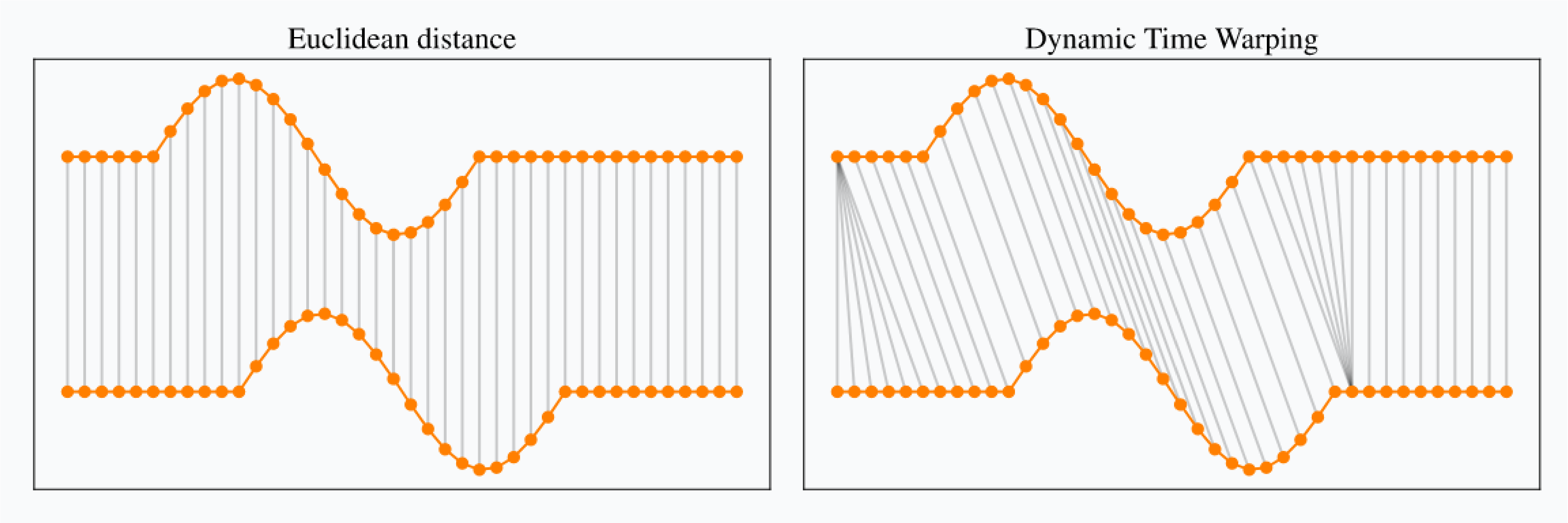
La classification implique de pouvoir **mesurer la similarité entre des instances d'une classe**.



La plupart des métriques ne prend pas en compte les déformations temporelles (translation, étirement).

https://rtavenar.github.io/blog/dtw.html

Dynamic Time Warping (DTW)



https://rtavenar.github.io/blog/dtw.html

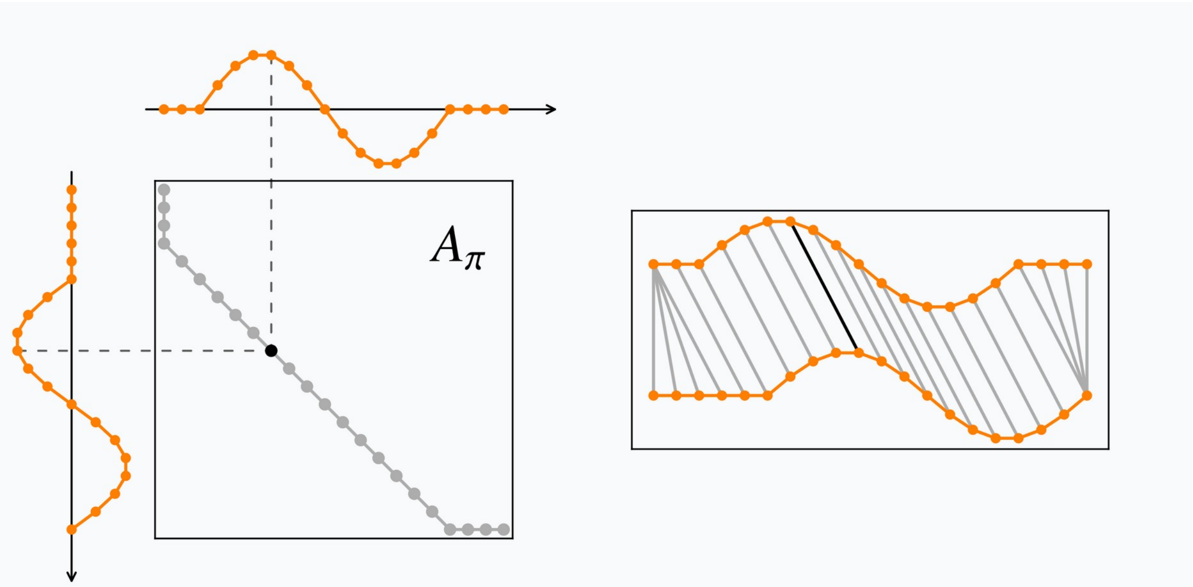
Dynamic Time Warping (DTW)

$$DTW_q(x, x') = \min_{\pi \in A(x,x')} \left(\sum_{(i,j) \in \pi} d(x_i, x'_j)^q \right)^{\frac{1}{q}}$$

déf.

$A(x, x')$ est l'ensembles des **chemins d'alignement** admissibles entre les N points de x et les M points de x'.

A_π est la matrice représentant ces chemins. Le calcul se fait par programmation dynamique.

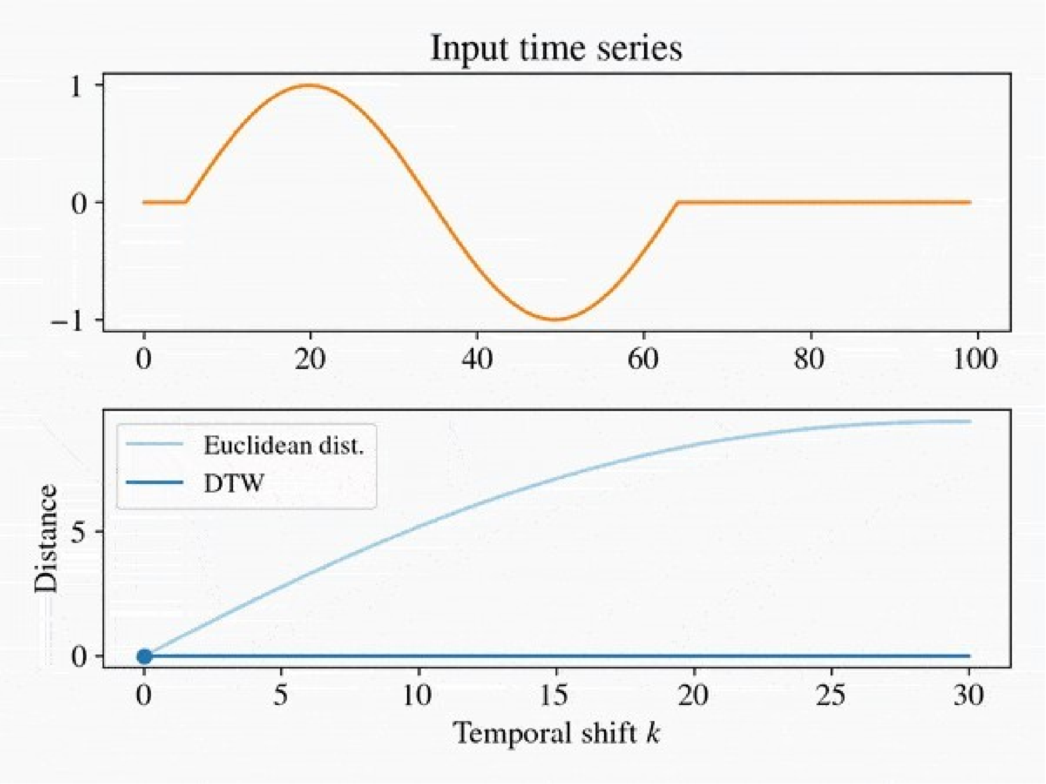


https://rtavenar.github.io/blog/dtw.html

Dynamic Time Warping (DTW)

$$DTW_q(x, x') = \min_{\pi \in A(x,x')} \left(\sum_{(i,j)\in\pi} d(x_i, x'_j)^q \right)^{\frac{1}{q}}$$

déf.



Mesure invariante aux déformations temporelles.

https://rtavenar.github.io/blog/dtw.html

À vous !

https://github.com/PAUL-BERNARD/ENSC-IA-2024