



Analyse des séries temporelles

Introduction aux séries temporelles

- Préliminaires

- Exemples

Modélisation

- Modèles de séries temporelles

- Inférence statistique

- Inférer la non-stationnarité : détection de ruptures

Le problème de la prédiction

- Formalisation du problème de prédiction

- Processus des innovations



Introduction aux séries temporelles

Préliminaires

Exemples

Modélisation

Le problème de la prédiction



Introduction aux séries temporelles

Préliminaires

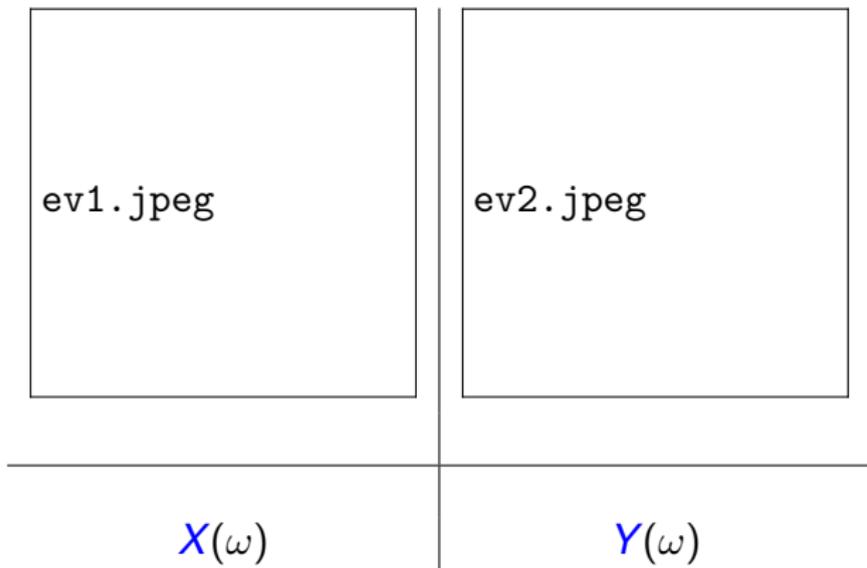
Exemples

Modélisation

Le problème de la prédiction



Qu'est-ce que l'information ?

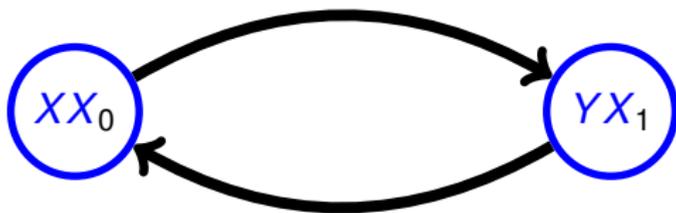


$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) ?$$



Qu'est-ce que la causalité ?

La notion de **causalité** n'est pas clairement définie uniquement à partir des **probabilités** :



$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B | X \in A)$$

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(Y \in B)\mathbb{P}(X \in A | Y \in B)$$

$$\mathbb{P}(X_0 \in A, X_1 \in B) = \mathbb{P}(X_0 \in A)\mathbb{P}(X_1 \in B | X_0 \in A)$$

D'où l'importance du temps $t = 0, 1$!



Introduction aux séries temporelles

Préliminaires

Exemples

Modélisation

Le problème de la prédiction

L'analyse des séries temporelles reposant sur la modélisation aléatoire trouve de nombreuses applications :

- ▷ Santé : analyse de signaux physiologiques (imagerie médicale).
- ▷ Ingénierie : surveillance, détection d'anomalies, localisation/poursuite.
- ▷ Données audio : analyse de la parole, synthèse, codage.
- ▷ Écologie : données climatiques, hydrologie.
- ▷ Économétrie : données économiques/financières.
- ▷ Assurance : analyse de risques.



Rythme cardiaque

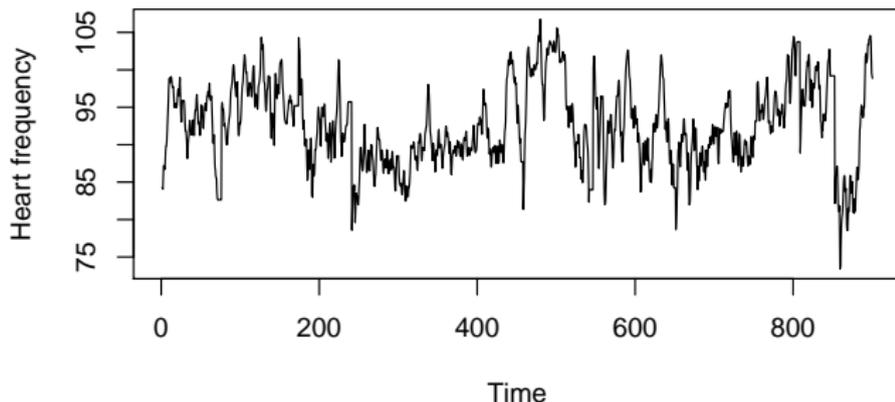


FIGURE: Rythme cardiaque d'une personne au repos au cours de 900 seconds. (nombre de battements par minute évalué toutes les 1/2 sec.)

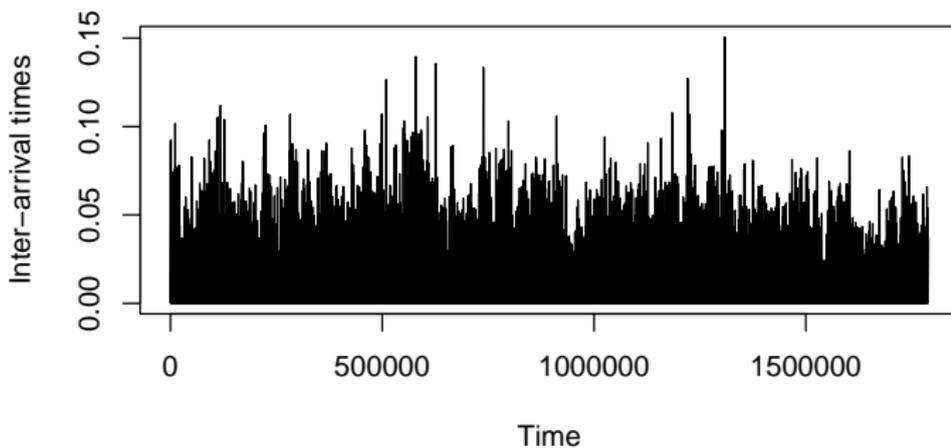


FIGURE: Temps d'inter-arrivées de paquets TCP (en seconde) durant 2 heures de trafic Internet sur un lien <http://ita.ee.lbl.gov/>.



Signal de parole

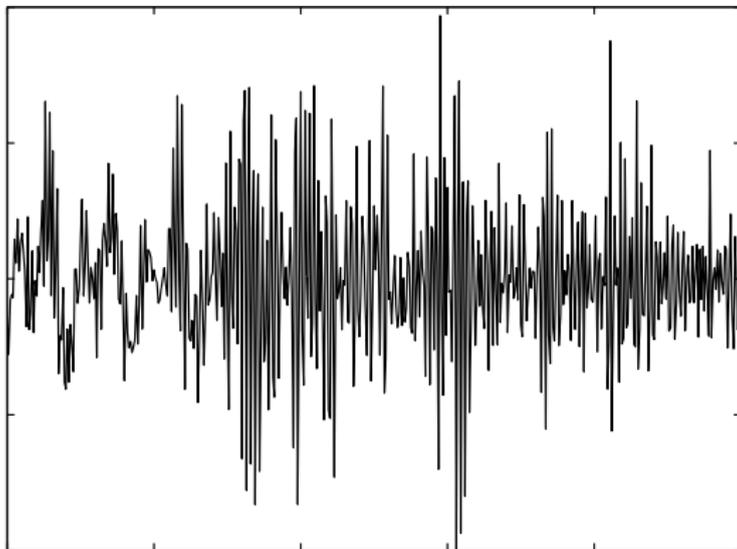


FIGURE: Un signal de parole échantillonné à 8000 Hz. Enregistrement du phonème sh (de sharp).



Données climatiques : vitesse du vent

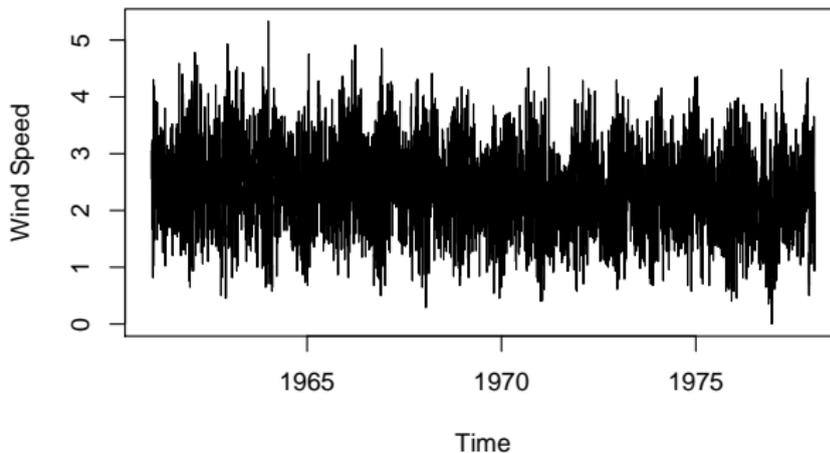


FIGURE: Enregistrement quotidien de la vitesse du vent à Kilkenny (Irlande) en noeuds (1 noeud = 0.5148 metres/second).



Données climatiques : indices de température

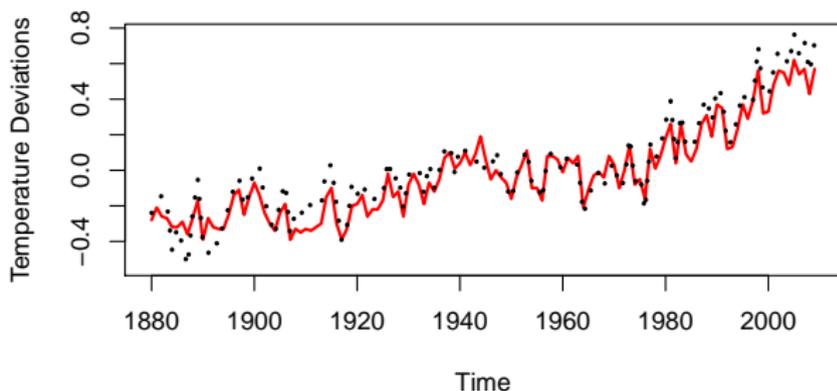


FIGURE: Indices moyens de température terre–océan (ligne rouge pleine) et surface–air (ligne noire pointillée).

<http://data.giss.nasa.gov/gistemp/graphs/>.



Produit national brut

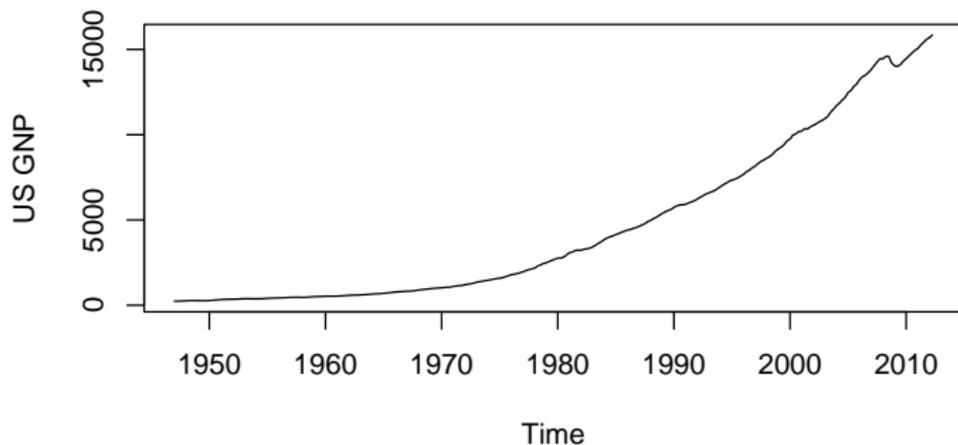


FIGURE: Produit national brut (PNB) des États-Unis en milliards de \$.
<http://research.stlouisfed.org/fred2/series/GNP>.



Taux trimestriel

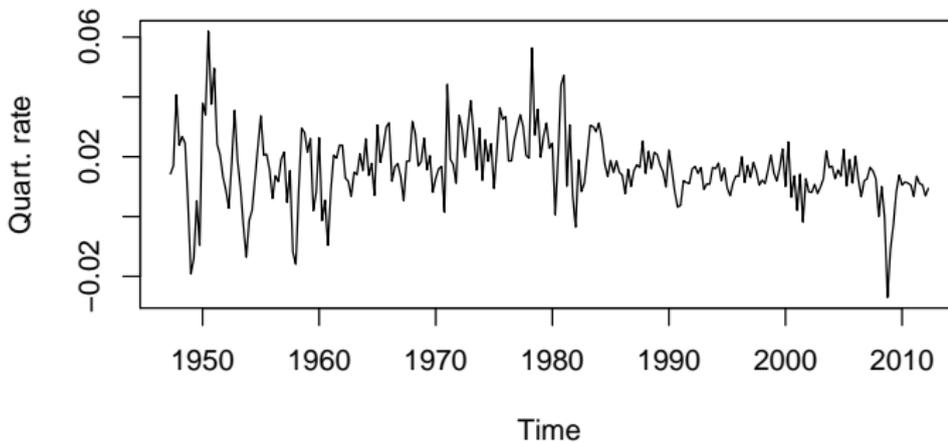


FIGURE: Taux trimestriel du PNB des États-Unis.



Indice financier

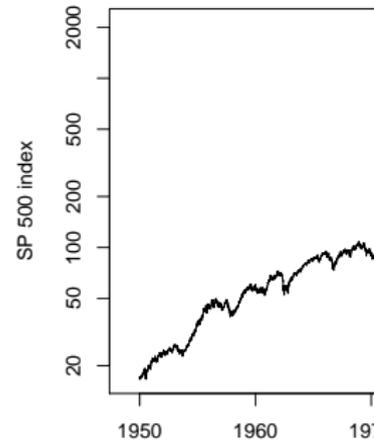
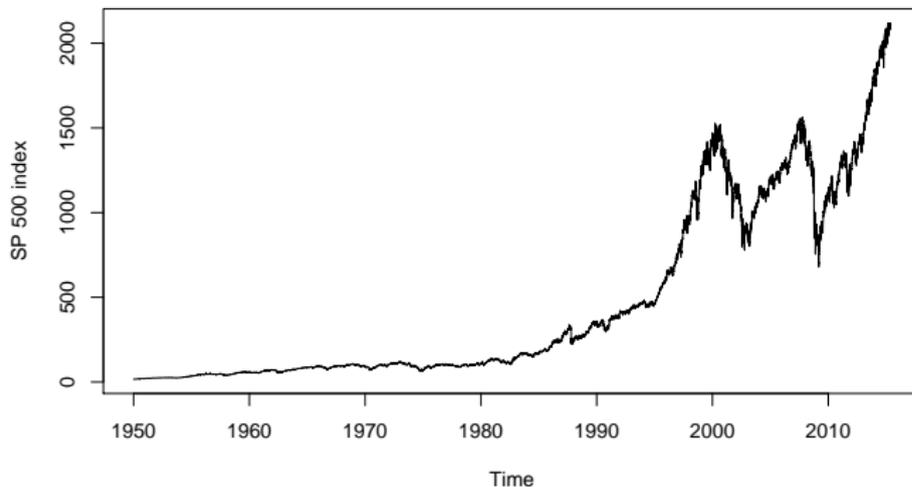


FIGURE: Valeur quotidienne à l'ouverture de l'indice Standard and Poor 500 (New York Stock Exchange (NYSE) et NASDAQ).



Rendements financiers

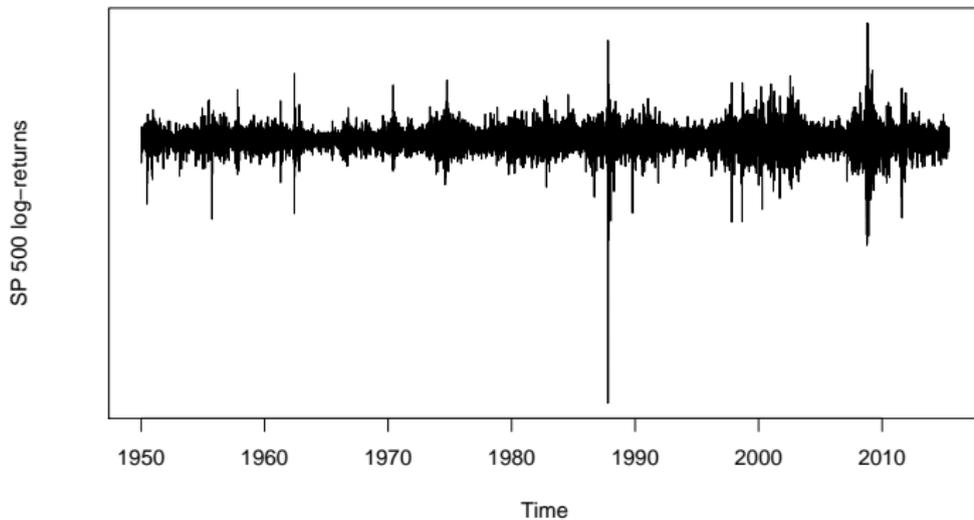


FIGURE: Log-rendements de l'indice SP500.



Buts principaux de l'analyse des séries temporelles

- ▷ **Modélisation aléatoire** : **tendance** (saisonnière, linéaire, . . .) + **bruit** (doté de “**propriétés structurelles**”).
- ▷ **Inférence Statistique** : **estimer** les paramètres du modèle, **tester** des hypothèses (**détecter** la présence d'une tendance, d'un signal, **classifier** des signaux).
- ▷ **Prédiction** : pour un modèle aléatoire donné, utiliser les données du passé pour “deviner” les valeurs futures.
- ▷ **Filtrage et poursuite** : estimer une quantité **cachée** (observée indirectement) et les suivre au cours du temps.
- ▷ **Détection d'un changement** : découvrir aussi rapidement que possible si la suite de valeurs observées a un comportement statistique qui s'est modifié au cours du temps (**détection d'anomalies**).



Introduction aux séries temporelles

Modélisation

Modèles de séries temporelles

Inférence statistique

Inférer la non-stationnarité : détection de ruptures

Le problème de la prédiction



Introduction aux séries temporelles

Modélisation

Modèles de séries temporelles

Inférence statistique

Inférer la non-stationnarité : détection de ruptures

Le problème de la prédiction



Stationnarité et covariance

- ▶ Une série temporelle est modélisée comme la réalisation d'un **processus stochastique** i.e. une suite $X = (X_t)_{t \in \mathbb{Z}}$ de v.a. définies sur le même **espace de probabilité** $(\Omega, \mathcal{F}, \mathbb{P})$.¹
- ▶ L'hypothèse de base en série temporelle est que $(X_t)_{t \in \mathbb{Z}}$ et $(X_{t+1})_{t \in \mathbb{Z}}$ ont la même loi : on dit que X est **stationnaire**.
- ▶ Alors si $X_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ (i.e. $\mathbb{E}[|X_t|^2] < \infty$), la **moyenne** $\mu = \mathbb{E}[X_t]$ ne dépend pas de t et la **covariance** $\gamma(s - t) = \text{Cov}(X_s, X_t)$ dépend uniquement de l'écart $s - t$.
- ▶ On dit que X est **stationnaire au second ordre**, si les propriétés du point précédent sont vérifiées (sans nécessairement supposer que X est stationnaire).

1. On peut alors munir $\mathbb{R}^{\mathbb{Z}}$ d'une tribu qui rend $\omega \mapsto X(\omega) = (X_t(\omega))_{t \in \mathbb{Z}}$ mesurable.



Exemples

- ▶ Si $X = (X_t)_{t \in \mathbb{Z}}$ est une suite de v.a. i.i.d. centrées et L^2 , on dit que X est un **bruit blanc fort**. Alors $\mu = 0$ et $\gamma(\tau) = 0$ pour $\tau \neq 0$.
- ▶ Si X est stationnaire au second ordre avec $\mu = 0$ et $\gamma(\tau) = 0$ pour $\tau \neq 0$, on dit que c'est un **bruit blanc faible**.
- ▶ Si X est stationnaire (au second ordre) alors Y défini par

$$Y_t = X_t + \sum_{k=1}^q \theta_k X_{t-k}, \quad t \in \mathbb{Z}, \text{ est stationnaire (au second ordre).}$$

- ▶ Si X est stationnaire au second ordre et $|\phi| \neq 1$, alors il existe un **unique** processus Y stationnaire au second ordre vérifiant

$$Y_t = \phi Y_{t-1} + X_t, \quad t \in \mathbb{Z}$$

- ▶ En particulier si $|\phi| < 1$, on peut écrire $Y_t = \sum_{k \geq 0} \phi^k X_{t-k}$.



Introduction aux séries temporelles

Modélisation

Modèles de séries temporelles

Inférence statistique

Inférer la non-stationnarité : détection de ruptures

Le problème de la prédiction



Méthode des moments

- ▷ La plupart des **modèles** repose sur un nombre fini de **paramètres** $\theta_1, \dots, \theta_p$ qu'il s'agit d'**estimer** à partir d'un historique de données X_1, \dots, X_n .
- ▷ La **méthode des moments** consiste à déduire des estimations de $\theta_1, \dots, \theta_p$ à partir d'estimations de μ et γ , en général :

$$\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n X_k,$$
$$\hat{\gamma}_n(t) = \frac{1}{n} \sum_{k=1}^{n-|t|} (X_k - \hat{\mu}_n) (X_{k+|t|} - \hat{\mu}_n).$$



Démo en R



Théorie asymptotique

- ▷ Comme X n'est en général pas i.i.d. les théorèmes habituels (loi des grands nombres, théorème de la limite centrale (TLC)) ne s'appliquent pas.
- ▷ On peut néanmoins avoir des résultats assez généraux, par exemple :
 - ▷ Si $\gamma(t) \rightarrow 0$ quand $t \rightarrow \infty$, alors $\hat{\mu}_n \xrightarrow{L^2} \mu$ quand $n \rightarrow \infty$.
 - ▷ Si $\sum_t |\gamma(t)| < \infty$, alors $\hat{\mu}_n \xrightarrow{\text{p.s.}} \mu$ quand $n \rightarrow \infty$.
 - ▷ TLC (si hypothèses plus fortes)...



Introduction aux séries temporelles

Modélisation

Modèles de séries temporelles

Inférence statistique

Inférer la non-stationnarité : détection de ruptures

Le problème de la prédiction



Du TLC au théorème de Donsker

Soit (X_k) suite de v.a. **i.i.d.** de moyenne μ et de variance σ^2 . On note

$$S_n(t) = \frac{1}{n} \sum_{k=1}^{n \lfloor nt \rfloor} X_k, \quad 0 \leq t \leq 1.$$

Alors le **théorème de la limite centrale** et le **théorème de Donsker** donne, quand $n \rightarrow \infty$,

$$W_n(t) := \frac{\sqrt{n}}{\sigma} (S_n(t) - t\mu) \Longrightarrow \mathcal{N}(0, 1) W(t)$$

où $W(t)$, $0 \leq t \leq 1$

est un mouvement brownien.



Comment se débarrasser de la moyenne

Pour se débarrasser de la moyenne μ , on écrit

$$\begin{aligned} W_n(t) - tW_n(1) &\implies W(t) - tW(1) \\ &\parallel \\ \frac{\sqrt{n}}{\sigma} (S_n(t) - t\mu) - t \frac{\sqrt{n}}{\sigma} (S_n(1) - \mu) &\implies W(t) - tW(1) \\ &\parallel \\ \frac{\sqrt{n}}{\sigma} (S_n(t) - tS_n(1)) &\implies W(t) - tW(1) \end{aligned}$$

$W(t) - tW(1)$, $0 \leq t \leq 1$, s'appelle un **pont brownien**.



Détection de rupture

Soit $\hat{\sigma}_n$ un estimateur consistant de σ , on en conclut :

$$\sup_{t \in [0,1]} \left| \frac{\sqrt{n}}{\hat{\sigma}_n} (\mathcal{S}_n(t) - t\mathcal{S}_n(1)) \right| \implies \sup_{t \in [0,1]} |W(t) - tW(1)| .$$

Si en revanche il y a une **rupture** dans la moyenne, *i.e.* il existe $\mu_1 \neq \mu_2$ et $0 < r < 1$ tels que

$$\mathbb{E}[X_k] = \begin{cases} \mu_1 & \text{si } 1 \leq k \leq [rn] \\ \mu_2 & \text{si } [rn] < k \leq n \end{cases} ,$$

alors on a

$$\sup_{t \in [0,1]} \left| \frac{\sqrt{n}}{\hat{\sigma}_n} (\mathcal{S}_n(t) - t\mathcal{S}_n(1)) \right| \sim \frac{\sqrt{n}}{\sigma} (1-r)r(\mu_1 - \mu_2) .$$



Détection de rupture

En effet, on peut écrire

$$\begin{aligned} S_n(r) - rS_n(1) &= (1-r-r)S_n(r) - r\{S_n(1) - S_n(r) - S_n(r)\} \\ &= (1-r)(S_n(r) - r\mu_1 - r\mu_1) \text{ (d'ordre } 1/\sqrt{n}) \\ &\quad - r\left(\{S_n(1) - S_n(r) - (1-r)\mu_2\} - (1-r)\mu_2\right) \text{ (d'ordre } 1/\sqrt{n}) \\ &\quad + \underline{(1-r)r(\mu_1 - \mu_2)}(1-r)r(\mu_1 - \mu_2) \end{aligned}$$



Démo en R



Introduction aux séries temporelles

Modélisation

Le problème de la prédiction

Formalisation du problème de prédiction

Processus des innovations



Introduction aux séries temporelles

Modélisation

Le problème de la prédiction

Formalisation du problème de prédiction

Processus des innovations



Espérance conditionnelle

La meilleure prédiction de Y sachant X est donné par

$$\hat{Y} = \mathbb{E}[Y | X] = \psi(X) ,$$

où (si $\mathbb{E}[|Y|^2] < \infty$) ψ est la fonction mesurable qui minimise

$$\phi \mapsto \mathbb{E} \left[(Y - \phi(X))^2 \right] .$$

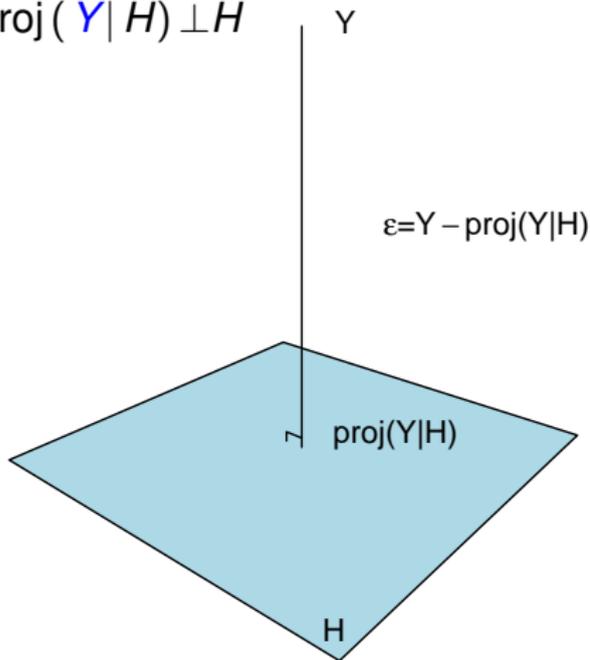
Il s'agit donc d'une **projection** $\hat{Y} = \text{proj}(Y | H)$ sur un sous-espace fermé $H = L^2(\Omega, \sigma(X), \mathbb{P})$ dans l'espace de Hilbert $L^2(\Omega, \mathcal{F}, \mathbb{P})$ muni du produit scalaire

$$\langle X, Y \rangle = \mathbb{E}[XY] .$$



Prédiction = projection

- (i) $\text{proj}(Y|H) \in H$
- (ii) $\epsilon = Y - \text{proj}(Y|H) \perp H$



On peut **simplifier** la résolution du problème de la prédiction en se contraignant à une **prédiction linéaire** :

$$H = \text{Vect}(1, X) = \{a + bX, a, b \in \mathbb{R}\}$$

(en supposant que $\mathbb{E}[|X|^2] < \infty$).

La solution est alors

$$\text{proj}(Y | \text{Vect}(1, X)) = \mathbb{E}[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - \mathbb{E}[X])$$



Introduction aux séries temporelles

Modélisation

Le problème de la prédiction

Formalisation du problème de prédiction

Processus des innovations



Définition

Soit $(X_t)_{t \in \mathbb{Z}}$ une série temporelle **stationnaire au second ordre centrée** de **covariance** γ . Son **passé "linéaire"** jusqu'au temps t est défini par

$$\mathcal{H}_t^X = \overline{\text{Vect}}(X_s, s \leq t) .$$

Prédiction linéaire et processus des innovations

- ▷ Le **meilleur prédicteur linéaire** est défini par

$$\text{proj}(X_t | \mathcal{H}_{t-1}^X) = \arg \min_{Y \in \mathcal{H}_{t-1}^X} \mathbb{E} [|X_t - Y|^2] .$$

- ▷ Le processus des innovations $(\epsilon_t)_{t \in \mathbb{Z}}$ de X est défini par

$$\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X), \quad t \in \mathbb{Z} .$$



Propriétés

On montre que $(\epsilon_t)_{t \in \mathbb{Z}}$ est un **bruit blanc faible**.

Étape 1 On a par définition $\text{proj}(X_t | \mathcal{H}_{t-1}^X) \in \mathcal{H}_{t-1}^X$, donc $\epsilon_t \in \mathcal{H}_t^X$.

Étape 2 En particulier $\mathbb{E}[\epsilon_t] = 0$.

Étape 3 De plus, pour tout $s < t$, on a alors $\epsilon_s \in \mathcal{H}_s^X \subseteq \mathcal{H}_{t-1}^X \perp \epsilon_t$. Donc $\langle \epsilon_s, \epsilon_t \rangle = 0$.

Étape 4 Il reste à montrer que $\text{Var}(\epsilon_t)$ ne dépend pas de t . Cela vient du fait que $(X_t)_{t \in \mathbb{Z}}$ et $(X_{t+1})_{t \in \mathbb{Z}}$ ont les mêmes covariances.



Exemples

- ▷ $X = (X_t)_{t \in \mathbb{Z}}$ est un **bruit blanc faible** si et seulement si $X = \epsilon$.
- ▷ Soit Z un **bruit blanc faible** et $|\phi| \neq 1$, et soit X l'**unique** processus stationnaire au second ordre vérifiant

$$X_t = \phi X_{t-1} + Z_t, t \in \mathbb{Z}$$

Alors, si $|\phi| < 1$, $\epsilon = Z$.

- ▷ En effet, $\phi X_{t-1} \in \mathcal{H}_{t-1}^X$ et comme on peut écrire $X_t = \sum_{k \geq 0} \phi^k Z_{t-k}$

on a $\mathcal{H}_{t-1}^X \subseteq \mathcal{H}_{t-1}^Z \perp Z_t$.

- ▷ Mais c'est faux si $|\phi| > 1$!



Prédiction en pratique

On dispose d'une suite X_1, \dots, X_T .

Étape 1 **Modélisation** (comment les données sont générées ?)

Étape 2 **Estimation** des paramètres du modèle.

Étape 3 Calcul du **meilleur prédicteur** correspondant.



Un petit plus : agrégation de N prédicteurs

Soient N prédicteurs $(\hat{X}_t^{(i)})_{t=1, \dots, T}$, $i = 1, \dots, N$. Soient pour un $\eta > 0$ et à tout instant $t = 1, \dots, T$, les poids

$$\hat{\alpha}_t^{(i)} \propto e^{-\eta \sum_{s=1}^{t-1} (\hat{X}_s^{(i)} - X_s)^2} \text{ qui somment à 1.}$$

Le **prédicteur agrégé** est défini par $\hat{X}_t = \sum_{i=1}^N \hat{\alpha}_t^{(i)} \hat{X}_t^{(i)}$, $t = 1, \dots, T$.

On fait l'hypothèse que $|\hat{X}_t^{(i)} - X_t| \leq C$ pour tout t et tout i . Alors, on peut trouver η suffisamment petit (en fonction de C) tel que

$$\frac{1}{T} \sum_{t=1}^T (\hat{X}_t - X_t)^2 \leq \inf_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T (\hat{X}_t^{(i)} - X_t)^2 + \frac{\ln N}{\eta T}.$$