

Introduction à l'analyse des séries temporelles

Felix Cheysson
Chargé de recherche CNRS

Université Gustave Eiffel, CNRS, UMR 8050, LAMA
felix.cheysson@univ-eiffel.fr

Année universitaire 2022-2023

Enseignement

- 2 × 3h de cours théoriques (ici, faculté de médecine du KB).
- 4 × 3h de TPs pratiques sous R (ici, KB-TP3).
 - TPs dédoublés en demi-groupes ?
- Supports de cours, exercices corrigés, annales et TPs disponibles sur <https://fcheysson.github.io/serc.html>.

Evaluation

- Quiz (questions de cours) au premier TP : 20 minutes, ~ 3 points.
 - Contenu indispensable signalé par l'icône 📖.
 - 1 feuille manuscrite recto-verso acceptée.
- Devoir collaboratif au dernier TP : 1 heure, ~ 7 points.
 - Tous documents + ordinateur acceptés.
- Devoir maison : 1 mois, ~ 10 points.

Rappels : variable aléatoire

- **Formellement.** Une variable aléatoire X est une application d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ dans un espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- **Intuitivement.** Une variable aléatoire est une variable dont la valeur dépend d'événements inconnus.
- On caractérise une variable aléatoire par :
 - les différentes valeurs qu'elle peut prendre ;
 - la probabilité qu'elle prenne ces valeurs.

- Fonction de répartition F_X :

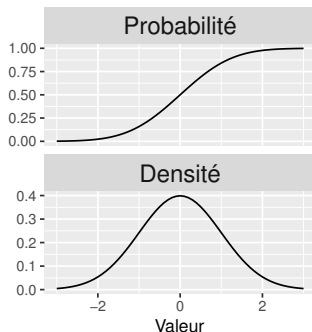
$$F_X(t) = \mathbb{P}(X \leq t);$$

- Fonction de densité f_X :

$$F_X(t) = \int_{-\infty}^t f_X(u) du.$$

- Fonction quantile :

$$Q(p) = \inf\{x \in \mathbb{R}; p \leq F_X(x)\} \approx F_X^{-1}(p)$$



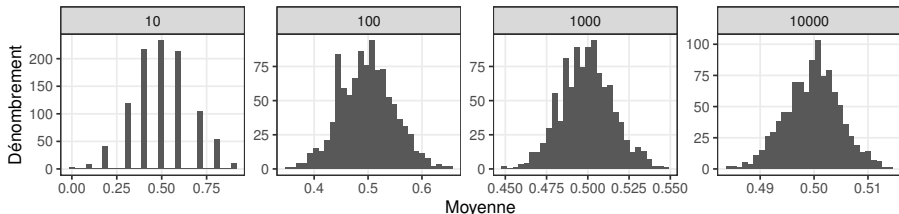
Rappels : loi normale $\mathcal{N}(\mu, \sigma^2)$

- Fonction de densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}.$$

- Caractérisée complètement par son espérance μ et sa variance σ^2 .
- Apparaît dans le **théorème central limite** : Soit une suite de variables aléatoires X_1, X_2, \dots indépendantes et identiquement distribuées (*i.i.d.*), d'espérance μ et de variance $\sigma^2 < \infty$. Alors

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$



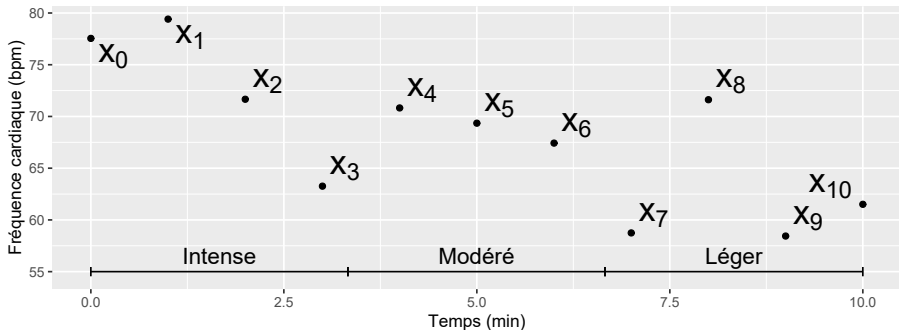
Définition : série temporelle

- **Définition.** On appelle série temporelle une collection de variables aléatoires indexées par le temps :

$$(X_t)_{t \in \mathbb{N}} = (X_0, X_1, X_2, \dots).$$

- **Exemple.** On mesure la fréquence cardiaque d'un individu toutes les minutes pendant 10 minutes au cours d'un effort :

$$(X_t)_{0 \leq t \leq 10} = (X_0, X_1, \dots, X_{10}).$$



- Lors d'un test d'effort, on cherche à exprimer la fréquence cardiaque d'un individu en fonction de l'effort réalisé par ce dernier.
- **Naïvement.** On effectue une ANOVA en fonction de l'effort :

$$X_t = \mu + \alpha_{\text{effort}} + \varepsilon_t,$$

où ε_t désigne une variable d'erreur gaussienne : $\varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

Question

Quel est le problème d'une telle modélisation ? Quelle hypothèse n'est pas vérifiée dans ce modèle ?

RESEARCH ARTICLE

ICU admissions and in-hospital deaths linked to COVID-19 in the Paris region are correlated with previously observed ambient temperature

Mehdi Mejdoubi^{1*}, Xavier Kyndt², Mehdi Djennaoui²

1 Department of Radiology, Hospital of Valenciennes, Valenciennes, France, **2** Department of Public Health, Hospital of Valenciennes, Valenciennes, France

* mejdoubi-m@ch-valenciennes.fr



Abstract

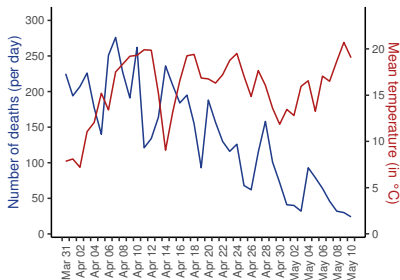
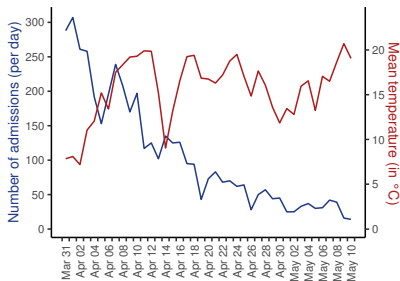
The purpose of this ecological study was to explore the association of weather with severity indicators of coronavirus disease 2019 (COVID-19). Daily COVID-19-related intensive care unit (ICU) admissions and in-hospital deaths in the Paris region and the daily weather characteristics of Paris midtown were correlated with a time lag. We assessed different study periods (41, 45, 50, 55, and 62 days) beginning from 31 March 2020. Daily ICU admissions and in-hospital deaths were strongly and negatively correlated to ambient temperatures (minimal, average, and maximal). The highest Pearson correlation coefficients and statistically significant *p* values were found 8 days before the occurrence of ICU admissions and 15 days before deaths. Partial correlations with adjustment on days since lockdown showed similar significant results. The study findings show a negative correlation of previously observed ambient temperature with severity indicators of COVID-19 that could partly explain the death toll discrepancies between and within countries.

Introduction

The ongoing major epidemic of coronavirus disease 2019 (COVID-19) infection began in cold areas of temperate countries of the northern hemisphere [1], which raises the question of weather influence.

Continental France has been severely hit by this outbreak of COVID-19, with around 30 000 deaths, including more than 19 000 in-hospital deaths (as of 15 June 2020). The main affected regions were its eastern region (mainly Alsace) and the Paris region (Île-de-France), with more than 3500 and 7200 in-hospital deaths, respectively.

Monitoring the outbreak of COVID-19 with biological testing is imprecise as the infection can be asymptomatic [2]. Data such as deaths or intensive care unit (ICU) admissions (reflecting critically ill patients) are more reliable, especially in highly developed countries. In the natural course of infection with COVID-19, a pulmonary worsening may appear around 5 to 9 days after infection onset [2, 3], with an immunological mechanism known as “cytokine



OPEN ACCESS

Citation: Mejdoubi M, Kyndt X, Djennaoui M (2020) ICU admissions and in-hospital deaths linked to COVID-19 in the Paris region are correlated with previously observed ambient temperature. *PLoS ONE* 15(11): e0242088. <https://doi.org/10.1371/journal.pone.0242088>

Editor: Corinne den Uyl, Erasmus Medical Center, NETHERLANDS

Received: August 3, 2020

Accepted: October 22, 2020

Published: November 20, 2020

Copyright: © 2020 Mejdoubi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this study are public and freely accessible. Ambient temperature data are available from <http://www.météofrance.com> and daily numbers of COVID-19 deaths are available from <https://data.covid19.fr>.

Funding: The author(s) received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.



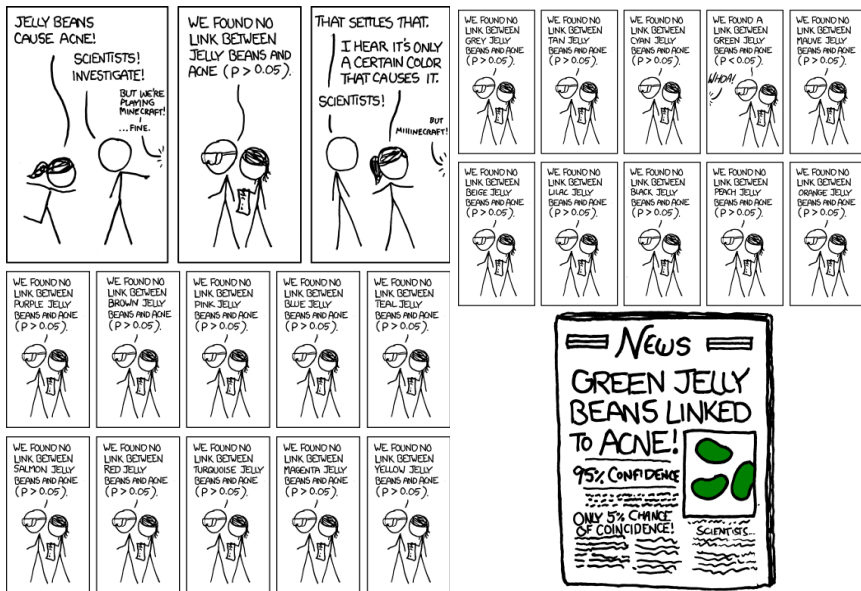
Procédure de décision permettant de rejeter une hypothèse sur la base de l'étude statistique d'un échantillon (X_1, \dots, X_n) .

- Hypothèse nulle \mathcal{H}_0 :
 - En général choisie comme l'inverse de ce que l'on veut montrer.
 - Hypothèse pour laquelle les calculs statistiques sont explicites.
- Statistique de test et distribution sous H_0 .
 - Détermine si l'échantillon a de fortes chances d'être observé sous \mathcal{H}_0 .
 - Ne représente qu'un cliché des caractéristiques de l'échantillon.
- Région de rejet et p -valeur.
 - Permettent de conclure quant au rejet ou non de \mathcal{H}_0 .

Risque de première espèce α

Probabilité de rejeter \mathcal{H}_0 à tort : attention aux tests multiples, aux fausses interprétations, aux idées reçues, etc.

p-value : erreurs d'interprétation (credits to xkcd)



Exemple : test de Kolmogorov-Smirnov

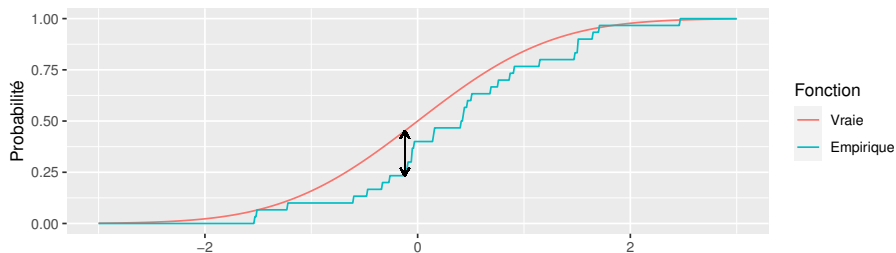
Permet de tester si l'échantillon provient d'une distribution F déterminée.

- Hypothèse nulle $\mathcal{H}_0 : \forall x \in \mathbb{R}, F_n(x) = F(x)$.
- Statistique de test : $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$.
- Loi sous \mathcal{H}_0 :

$$\sqrt{n}D_n \xrightarrow[n \rightarrow \infty]{d} K = \sup_{t \in [0,1]} |B(F(t))|,$$

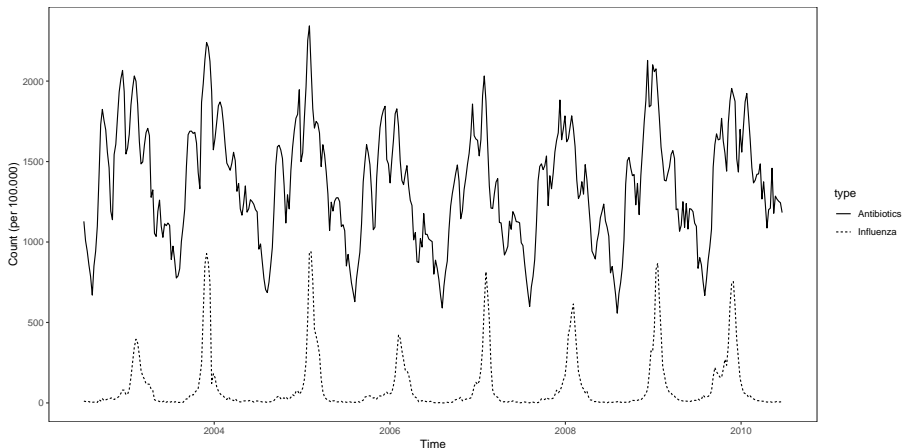
où $B(\cdot)$ correspond au pont brownien.

- Région de rejet au niveau de risque $\alpha : \sqrt{n}D_n > K_{1-\alpha}$, où $K_{1-\alpha}$ représente le quantile de niveau $1 - \alpha$ de la variable K .



Exemple de données de surveillance

- Nombre de remboursements des prescriptions d'antibiotiques en ville.
 - Données du SNIIRAM (CNAMTS + RSI) : 90% de la population.
- Nombre de syndromes grippaux.
 - Données du Réseau Sentinelles : représentatives à l'échelle de la France.



Qualité de l'indicateur

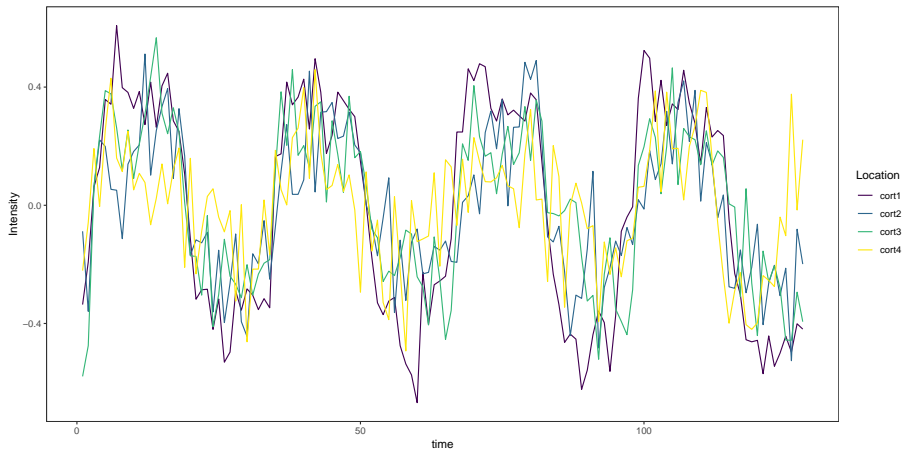
- **Représentatif ?**
 - Surveillance passive : sous-estimation.
 - Biais de recrutement.
- **Stabilité temporelle ?**
 - Evolution de l'indicateur dans le temps.

Qualité de la modélisation

- Caractérisation simple de la série (tendance, saison, variations accidentelles, points de rupture, ...).
- Représentation de la série par un modèle statistique (ARIMA).
- Prédiction de la série à partir de ses valeurs passées.

Exemple de données biologiques

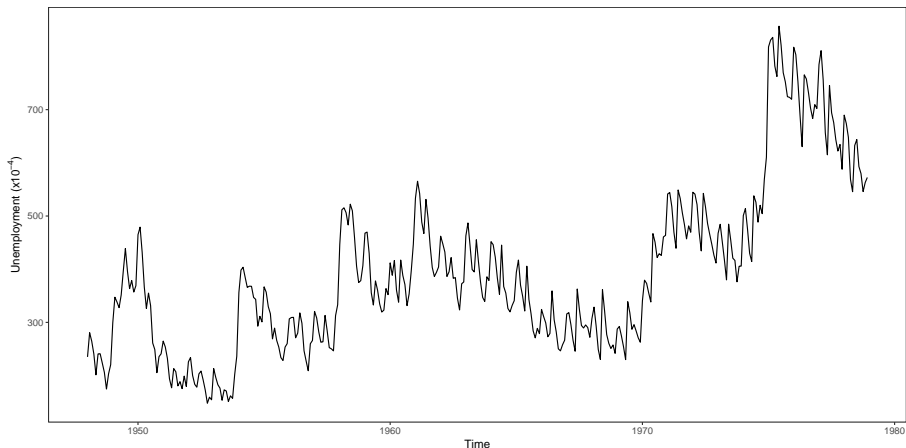
Intensité du signal BOLD (Blood Oxygenation-Level Dependent) dans le cortex d'un patient soumis à des stimuli, collectée par fMRI.



Exemple de données économiques

Nombre de personnes sans emploi aux Etats-Unis entre 1948 et 1978.

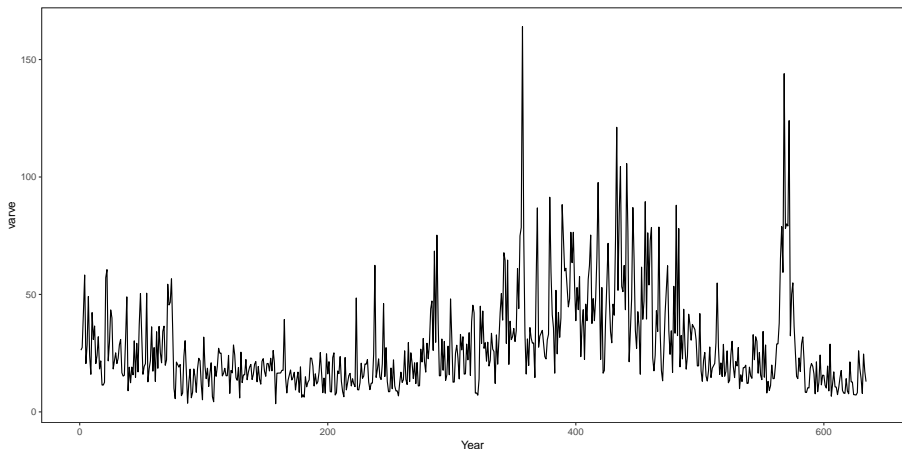
- Impact du choc pétrolier de 1973.



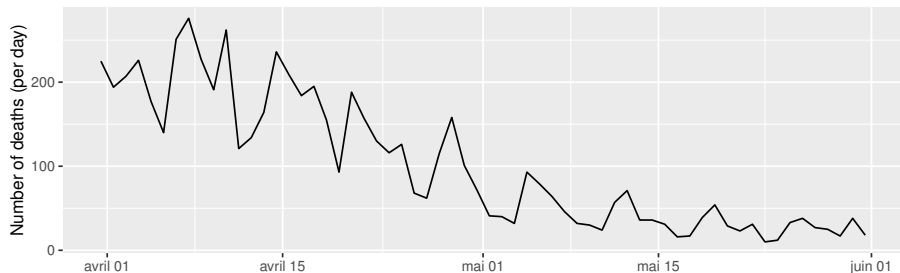
Exemple de données géologiques

Épaisseur de la couche sédimentaire.

- Recueillie au Massachusetts.
- Représentative de la température il y a 11.834 ans.



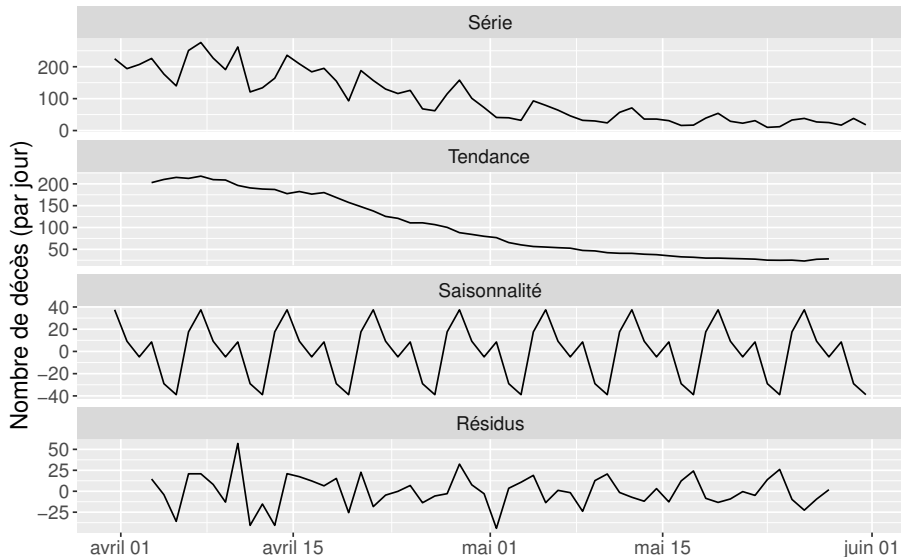
Composantes d'une série temporelle



On distingue généralement 3 composantes d'une série temporelle :

- Sa tendance (= évolution à long terme) ; peut être récupérée par une moyenne glissante ou un modèle de régression.
- Sa saisonnalité (= évolution périodique) ; peut être récupérée par une moyenne saisonnière ou par analyse spectrale de la série.
- Ses résidus (= ce qu'il reste).

Décomposition d'une série temporelle



On dispose d'une ou de deux séries que l'on veut modéliser :

$$\begin{aligned}X_t &= T_t + S_t + U_t \\Y_t &= T'_t + S'_t + \beta X_t + U'_t,\end{aligned}$$

où

- T_t et T'_t désignent les **tendances** (évolution à long terme) ;
- S_t et S'_t désignent les **saisonnalités** (évolution périodique) ;
- U_t et U'_t désignent les variables **résiduelles autodépendantes**, *i.e.*

$$\forall s \neq t, \quad U_s \not\perp U_t \quad \text{et} \quad U'_s \not\perp U'_t.$$

Problématique

Prendre en compte la structure d'autodépendance des innovations.

Applications

- Identifier la nécessité d'un modèle de séries temporelles.
- Modéliser les séries temporelles simples (ARIMA).
 - Nomenclature, démarche, *etc.*
 - Programmation en R.
- Connaître ses limites.

Théoriques

- Connaître les notions statistiques sous-jacentes aux modèles de séries temporelles.
- Comprendre les résultats mathématiques permettant l'analyse des séries temporelles.
 - Stationnarité d'une série.
 - Théorème de représentation de Wold.

- 1 Notions théoriques essentielles
- 2 Exemples filés
- 3 Stationnarité
- 4 Estimateurs de la corrélation
- 5 Données non stationnaires
- 6 Modèles ARIMA
- 7 Modèles saisonniers SARIMA
- 8 Procédure de Box et Jenkins
- 9 Pour aller plus loin : introduction à l'analyse spectrale

- 1 Notions théoriques essentielles
 - Fonction d'autocovariance
 - Fonction d'autocorrélation



- **Espérance.**

(ex. 2, 4, 5)

$$\mu_t = \mathbb{E}[X_t]$$

- **Fonction d'autocovariance** ou produit moment d'ordre 2. (ex. 1)

$$\begin{aligned}\gamma(s, t) &= \text{Cov}(X_s, X_t) \\ &= \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)]\end{aligned}$$

- *Remarque.* La fonction d'autocovariance désigne la dépendance linéaire entre deux temps.
- *Remarque.* $\gamma(t, t) = \text{Var}(X_t)$.



- **Propriétés.** Soient A, B et C trois variables aléatoires et α un réel.
(ex. 2, 3, 4)

$$\text{Cov}(A, B) = \text{Cov}(B, A)$$

$$\text{Cov}(A + B, C) = \text{Cov}(A, C) + \text{Cov}(B, C)$$

$$\text{Cov}(\alpha A, B) = \alpha \text{Cov}(A, B)$$

- **Fonction d'autocorrélation (ACF).** (ex. 2, 3)

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

- *Remarque.* La fonction d'autocorrélation caractérise la prédictabilité linéaire de la série au temps t sachant la valeur de X_s .

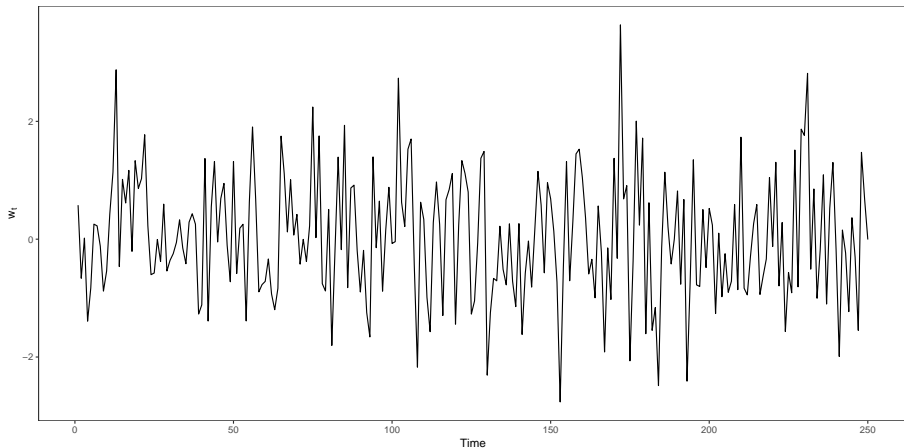
2 Exemples filés

- Bruit blanc Gaussien
- Moyenne mobile
- Processus autorégressif
- Marche aléatoire avec dérive
- Signal bruité

Exemple 1.1 : Bruit blanc Gaussien

Collection de variables aléatoires normales *indépendantes* et *identiquement distribuées* :

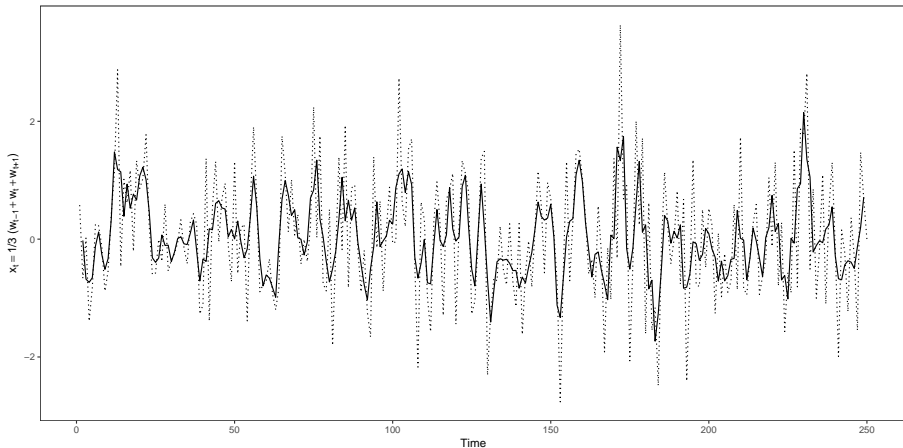
$$W_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$$



Exemple 1.2 : Moyenne mobile

Combinaison linéaire d'un bruit blanc Gaussien :

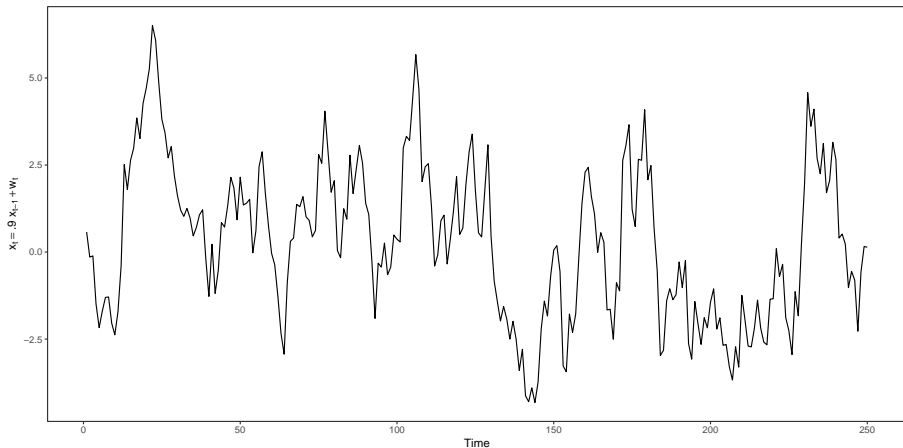
$$X_t = \frac{1}{3} (W_{t-1} + W_t + W_{t+1})$$



Exemple 1.3 : Processus autorégressif

Régression sur les valeurs passées de la série :

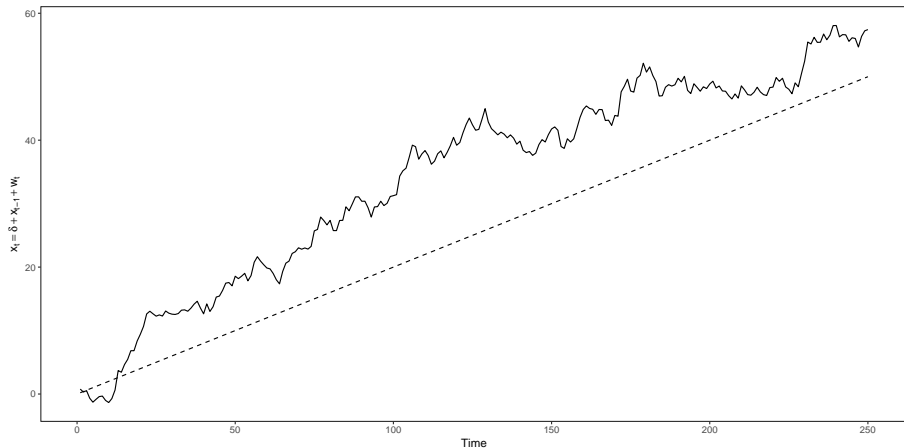
$$X_t = 0.9 X_{t-1} + W_t$$



Exemple 1.4 : Marche aléatoire avec dérive

Succession de pas aléatoire avec dérive δ :

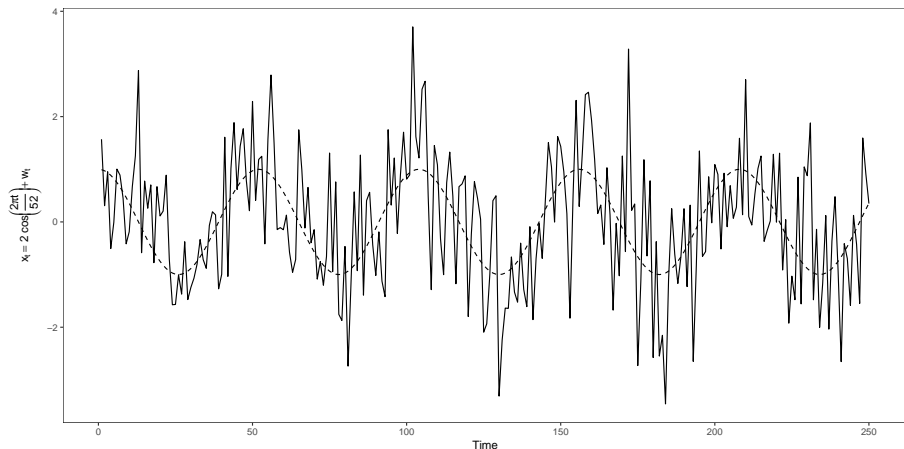
$$X_t = \delta + X_{t-1} + W_t$$



Exemple 1.5 : Signal bruité

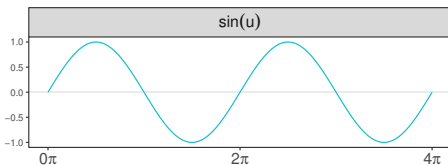
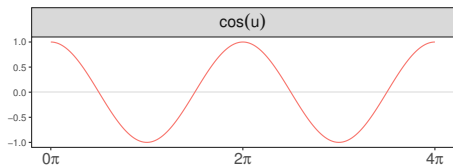
Signal périodique bruité :

$$X_t = 2 \cos\left(\frac{2\pi t}{52}\right) + W_t$$

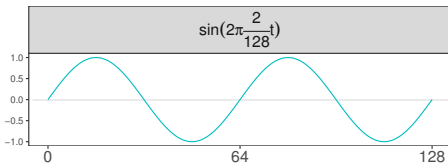
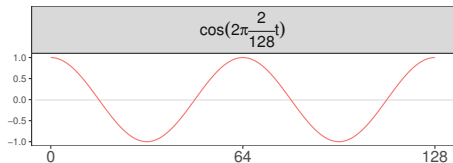


Sinus et cosinus

Les fonctions \sin and \cos sont 2π -périodiques : pour $u \in \mathbb{R}$ et $k \in \mathbb{Z}$,
$$\cos(u + 2k\pi) = \cos(u).$$

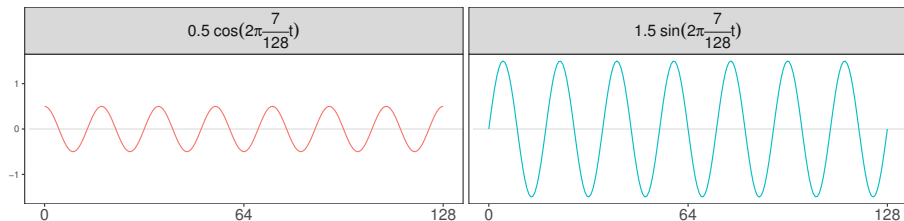


Soit $u = 2\pi \frac{2}{128}t$ pour $t = 1, 2, \dots, 128$.



On peut interpréter $\frac{2}{128}$ comme 2 cycles pour 128 unités de temps.

Similairement, avec $u = 2\pi \frac{7}{128}t$ pour $t = 1, 2, \dots, 128$.

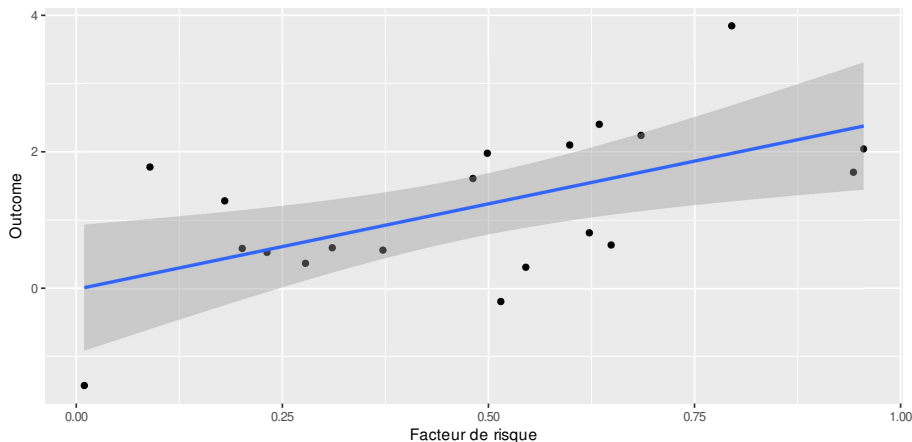


- $\cos(2\pi \frac{k}{n}t)$ et $\sin(2\pi \frac{k}{n}t)$ possèdent k cycles pour n unités de temps.
- La quantité $f = \frac{k}{n}$ est appelée la *fréquence* du sinus ou du cosinus.
 - Elle représente le nombre (ou la fraction) de cycles par unité de temps.
 - Si f est faible (resp. élevée), le sinus est dit de basse (haute) fréquence.
- La *période* $T = \frac{1}{f}$ dénombre les unités de temps nécessaires à un cycle.
- L'*amplitude* est l'étendue maximale de la variation et est égale à 1 pour les fonctions sin et cos.

- 3 Stationnarité
 - Problématique
 - Stationnarité au sens strict
 - Stationnarité au sens faible

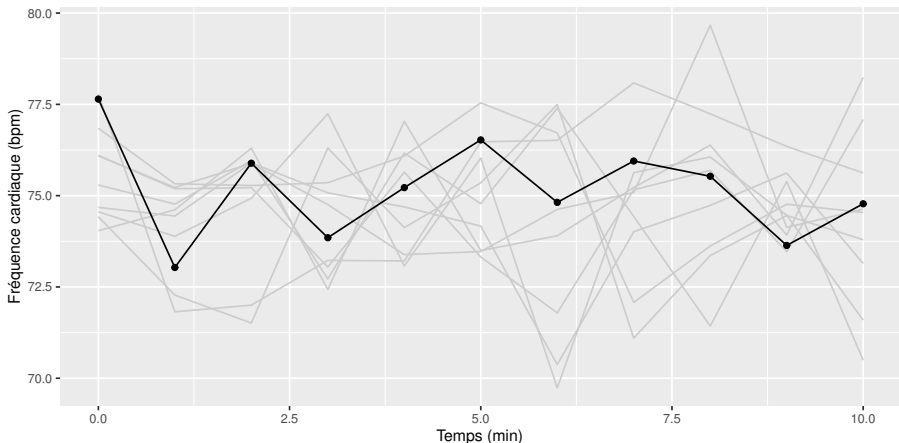
Retour sur le cas simple : indépendance des observations

- Echantillon **indépendant** $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$.
- Inférence de la relation $Y = f(X)$ en combinant l'information.
- Par exemple, cas de la régression linéaire :



De l'inférence avec une seule trajectoire ?

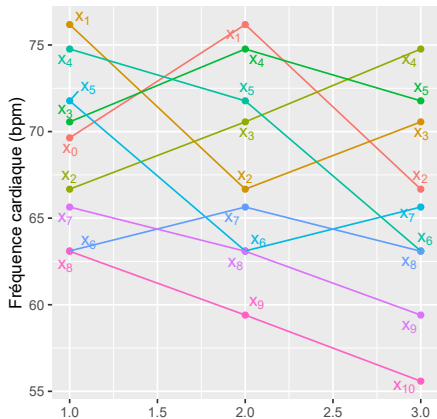
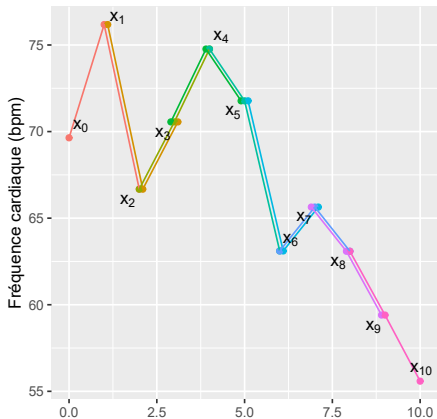
- Série temporelle : on ne dispose que d'une seule trajectoire $\{X_t\}_{t \in \mathbb{N}}$.
- Comment inférer le comportement statistique de X_t ?



Stationnarité au sens strict

On dit que la série (X_t) est **stationnaire au sens strict** si le comportement probabiliste de toute collection $(X_{t_1}, \dots, X_{t_k})$ est invariant par translation.

→ Pour tout h , le comportement probabiliste de $(X_{t_1}, \dots, X_{t_k})$ est identique à $(X_{t_1+h}, \dots, X_{t_k+h})$.



Stationnarité au sens faible (ou du 2^{ème} ordre)



On dit que la série (X_t) est **stationnaire au sens faible** si ses deux premiers moments sont invariants par translation :

- $\mu_t = \mathbb{E}[X_t]$ ne dépend pas de t ;
- $\gamma(s, t) = \text{Cov}(X_s, X_t)$ ne dépend que de $|s - t|$.

Il s'agit d'une propriété plus faible que la stationnarité stricte :

stationnarité stricte \implies stationnarité faible,

stationnarité stricte $\not\Leftarrow$ stationnarité faible,

stationnarité stricte $\xleftarrow{X_t \text{ gaussien}}$ stationnarité faible.

Enjeu de la propriété de stationnarité

Pour une série stationnaire, il est possible d'identifier et d'estimer les caractéristiques statistiques de la série à partir d'une unique trajectoire.



- **Espérance.** $\mathbb{E}[X_t] = \mu$;
- **Fonction d'autocovariance.**

$$\begin{aligned}\gamma(t+h, t) &= \text{Cov}(X_{t+h}, X_t) \\ &= \text{Cov}(X_h, X_0) \\ &= \gamma(h, 0) \\ &=: \gamma(h) ;\end{aligned}$$

- **Variance.** $\text{Var}(X_t) = \gamma(t, t) = \gamma(0, 0) = \gamma(0)$;
- **Fonction d'autocorrélation.** $\rho(h) := \rho(t+h, t) = \dots = \gamma(h)/\gamma(0)$.
- *Exemples.* Les séries 1, 2, 4 et 5 sont-elles stationnaires ?
- *Exemple.* Sous l'hypothèse de stationnarité, calculer l'espérance et la variance pour la série 3.

- 4 Estimateurs de la corrélation
 - Définitions
 - Exemples

- **Fonction d'autocovariance empirique.** Il s'agit de l'estimateur défini par :



$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n).$$

- **Fonction d'autocorrélation (ACF) empirique.** Il s'agit de l'estimateur précédent normalisé :

$$\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0).$$


Propriétés asymptotiques de l'ACF

Pour un bruit blanc Gaussien (W_t), si n est "grand", la loi de l'ACF empirique peut être approchée par :

$$\hat{\rho}_w(h) \sim \mathcal{N}(0, \sigma_n^2),$$

où $\sigma_n = \frac{1}{\sqrt{n}}$.

Test d'hypothèse sur l'autocorrélation

Motivation : Permet de déterminer si l'autocorrélation $\rho(\cdot)$ à un décalage h fixé est différent de 0. 

- Hypothèse nulle \mathcal{H}_0 : $\rho(h) = 0$.
- Statistique de test :

$$Z = \frac{\widehat{\rho}(h) - \rho(h)}{\sigma_n} \underset{\mathcal{H}_0}{=} \frac{\widehat{\rho}(h)}{\sigma_n}.$$

- Loi *asymptotique* sous \mathcal{H}_0 :

$$Z \sim \mathcal{N}(0, 1).$$

- Région de rejet au risque α : $|Z| > u_{1-\alpha/2}$, où $u_{1-\alpha/2}$ désigne le quantile de niveau $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$.

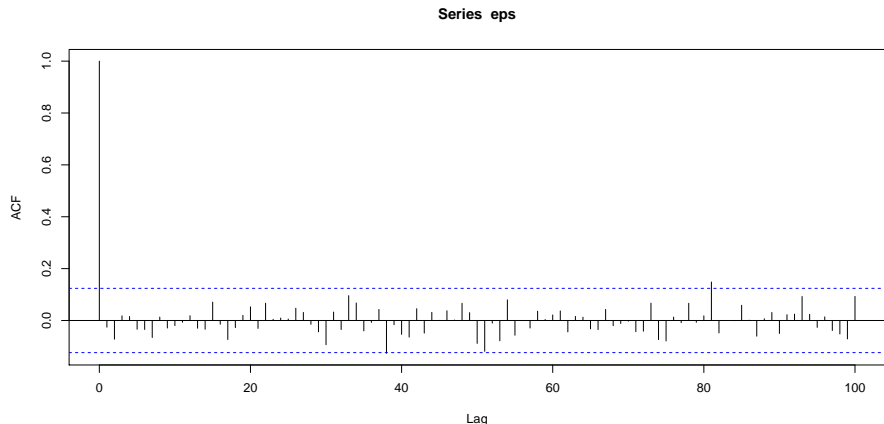
Rejet de \mathcal{H}_0

L'existence d'un pic d'autocorrélation $\rho(h)$ non nul représente une association statistique entre deux observations de la série séparées de h unités de temps.

ACF 4.1 : Bruit blanc Gaussien

Collection de variables aléatoires normales *indépendantes* et *identiquement distribuées* :

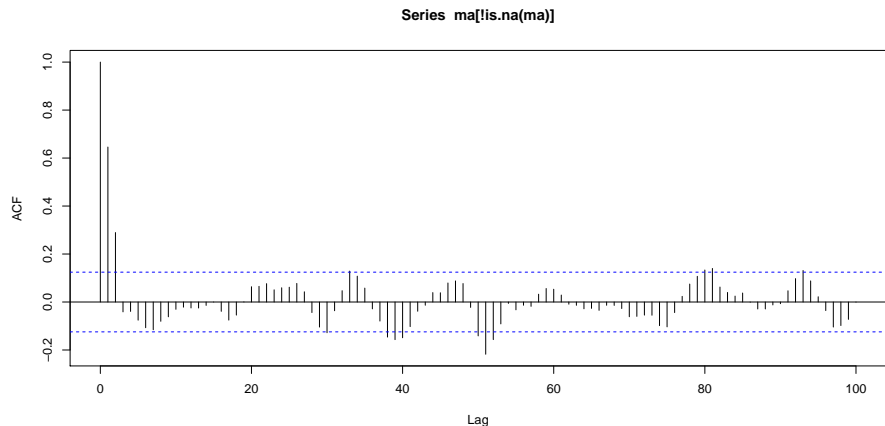
$$W_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$$



ACF 4.2 : Moyenne mobile

Combinaison linéaire d'un bruit blanc Gaussien :

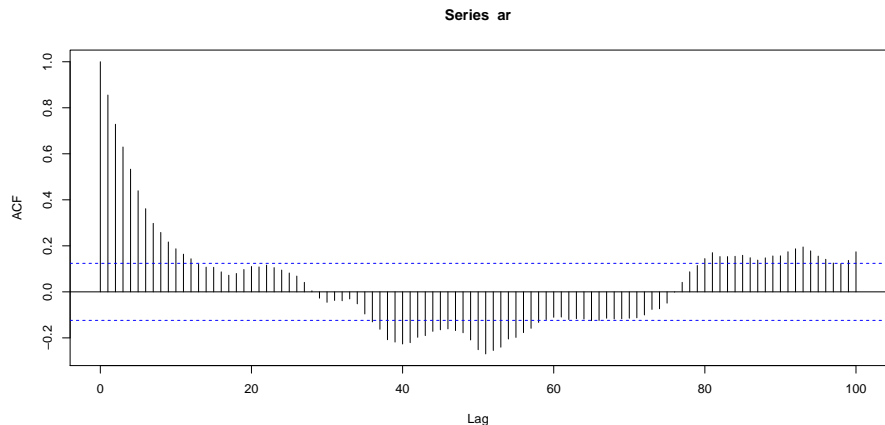
$$X_t = \frac{1}{3} (W_{t-1} + W_t + W_{t+1})$$



ACF 4.3 : Processus autorégressif

Régression sur les valeurs passées de la série :

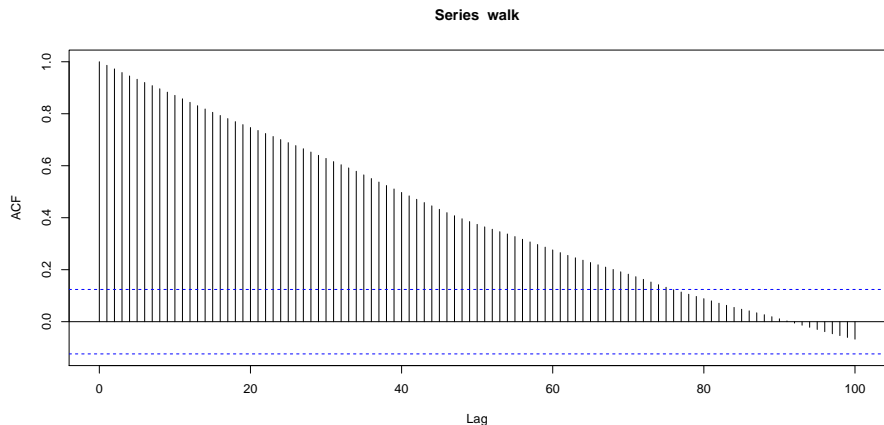
$$X_t = 0.9 X_{t-1} + W_t$$



ACF 4.4 : Marche aléatoire avec dérive

Succession de pas aléatoire avec dérive δ :

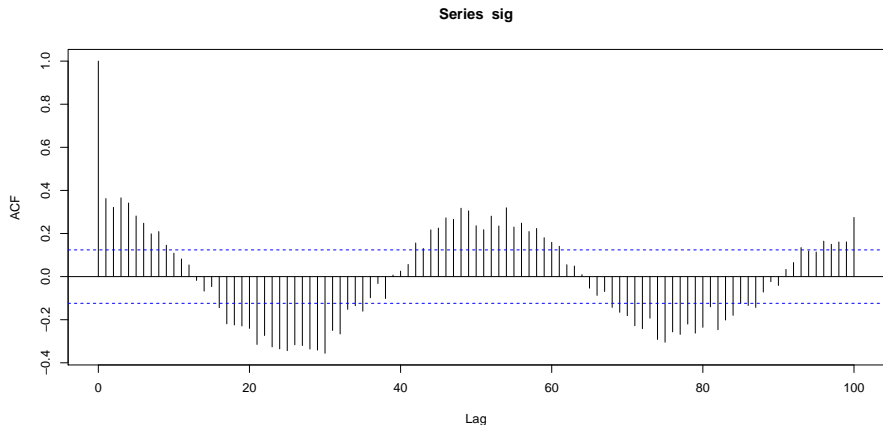
$$X_t = \delta + X_{t-1} + W_t$$



ACF 4.5 : Signal bruité

Signal périodique bruité :

$$X_t = 2 \cos\left(\frac{2\pi t}{52}\right) + W_t$$



- 5 Données non stationnaires
 - Prise en compte d'une tendance par régression
 - Prise en compte d'une tendance par différenciation
 - Stabilisation de la variance



Problématique

En pratique, la plupart des jeux de données ne sont pas issus de processus stationnaires.

Méthode. On va chercher à *stationnariser* (rendre stationnaire) la série de données afin de pouvoir analyser sa structure de dépendance.

- Prise en compte d'une tendance par régression ;
- Prise en compte d'une tendance par différentiation.
 - Tendance linéaire ;
 - Tendance stochastique ;
 - Tendance saisonnière.


On considère une série du type $X_t = \mu_t + \nu_t$, où U_t est stationnaire.

Par exemple :

- Régression linéaire avec erreurs stationnaires : $X_t = \beta_0 + \beta_1 t + U_t$;
- Signal bruité : $X_t = 2 \cos\left(\frac{2\pi t}{52}\right) + W_t$

Démarche de modélisation :

1. Modéliser μ_t , et effectuer la régression linéaire $X_t = \mu_t + \varepsilon_t$;
2. Etudier la structure de dépendance des résidus $\hat{U}_t = X_t - \hat{\mu}_t$;
3. Utiliser la structure identifiée pour estimer conjointement les paramètres de la régression et de la structure de dépendance.

- **Principe.** On retire la tendance en étudiant les différences de la série. 
- On note B l'**opérateur retard** défini par :

$$B X_t = X_{t-1},$$

$$B^2 X_t = B(BX_t) = B(X_{t-1}) = X_{t-2},$$

$$B^k X_t = X_{t-k}.$$

- On note $\nabla := 1 - B$ l'opérateur de différence d'ordre 1 :

$$\nabla X_t = (1 - B) X_t = X_t - BX_t = X_t - X_{t-1},$$

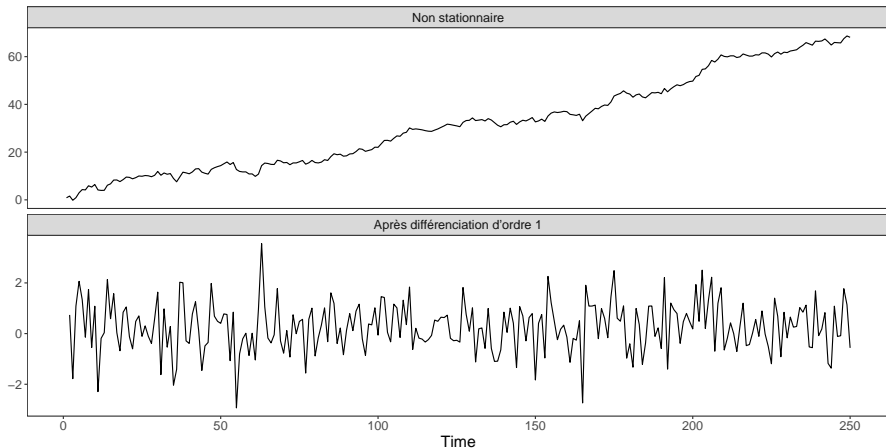
$$\nabla^2 X_t = (1 - B)^2 X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}.$$

- *Exemple.* Etudier la stationnarité de la série 4 par différenciation.
- *Exemple.* Etudier la stationnarité de la régression linéaire avec erreurs stationnaires par différenciation.

Différentiation de la marche aléatoire

Succession de pas aléatoire avec dérive δ :

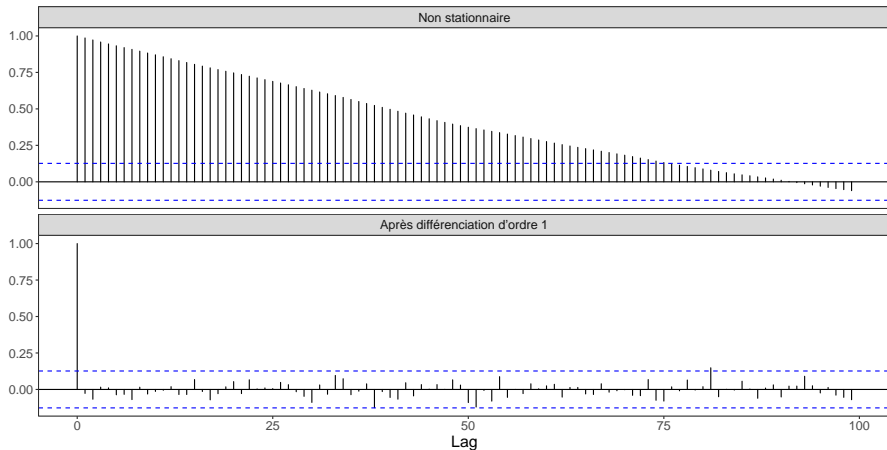
$$X_t = \delta + X_{t-1} + W_t$$



Stationnarité : Marche aléatoire avec dérive

Succession de pas aléatoire avec dérive δ :

$$X_t = \delta + X_{t-1} + W_t$$



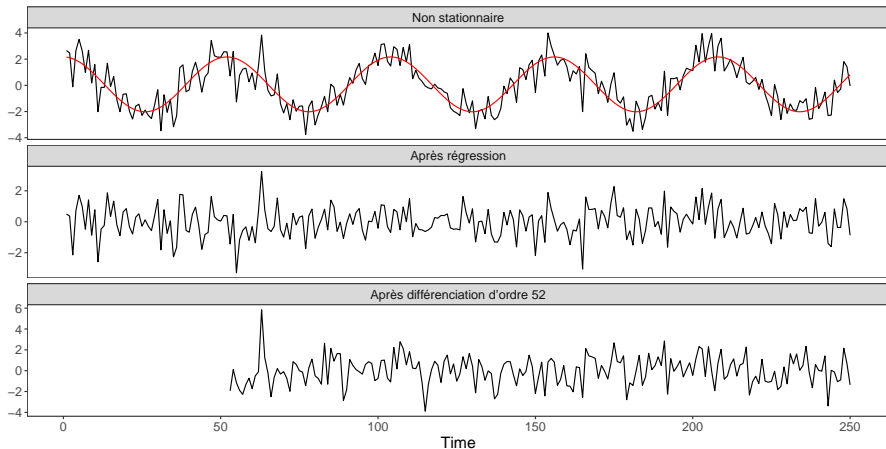
- Pour une série saisonnière, on pourra étudier une différenciation de plus grand ordre.
- On note $\nabla_S = 1 - B^S$ l'opérateur de différence d'ordre S :

$$\nabla_S X_t = (1 - B^S)X_t = X_t - X_{t-S}.$$

- *Exemple.* Etudier la stationnarité de la série 5 par différenciation.

Différentiation ou régression du signal bruité

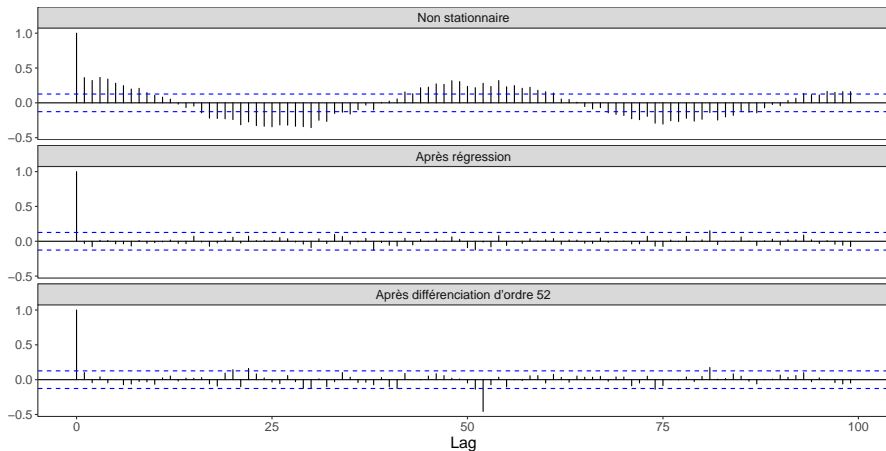
$$X_t = 2 \cos\left(\frac{2\pi t}{52}\right) + W_t$$



Stationnarité : Signal bruité

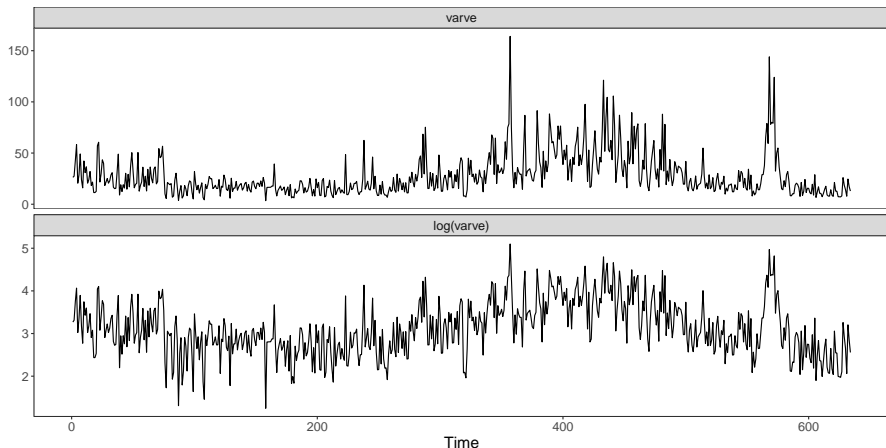
Signal périodique bruité :

$$X_t = 2 \cos\left(\frac{2\pi t}{52}\right) + W_t$$



Stabilisation de la variance

- Pour des séries dont la variabilité augmente avec l'espérance, on peut stabiliser la variance en transformant la série par un logarithme.
- *Exemple.* $Y_t = \log X_t$ avec X_t la série des couches sédimentaires.



6 Modèles ARIMA

- Modèles MA
- Modèles AR
- Autocorrélation partielle
- Modèles ARMA
- Estimation
- Prédiction

Modèle moyenne mobile (*moving average*)

- **Définition.** On appelle processus moyenne mobile d'ordre q , noté MA(q), le processus X_t défini par :



$$\begin{aligned} X_t &= W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q} \\ &= W_t + \sum_{j=1}^q \theta_j W_{t-j}, \end{aligned}$$

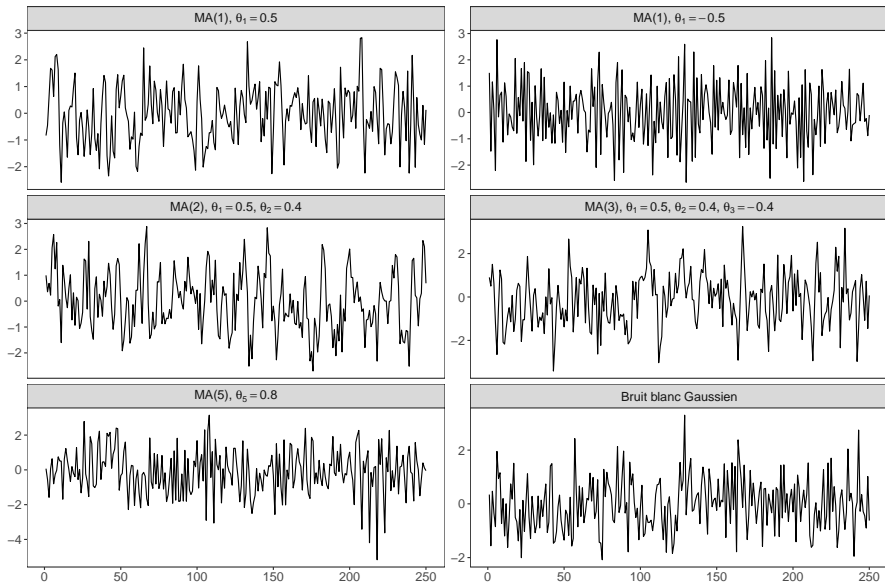
où $\theta_q \neq 0$ et $W_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

- *Remarque.*

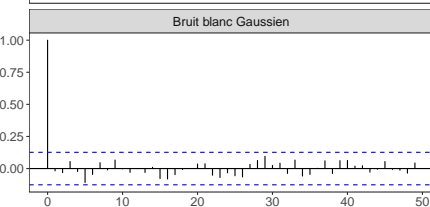
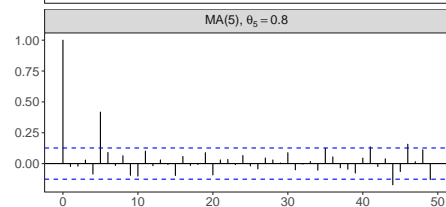
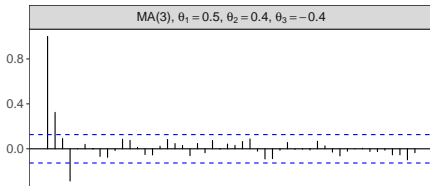
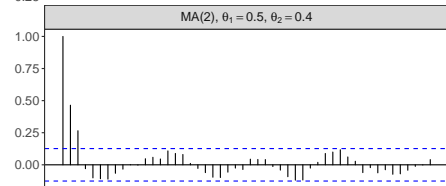
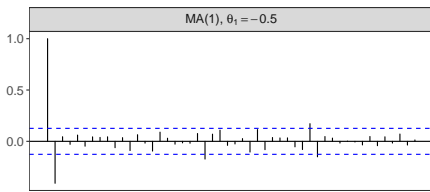
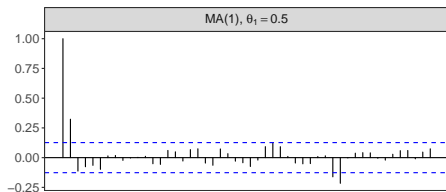
$$\begin{aligned} X_t &= (1 + \theta_1 B + \dots + \theta_q B^q) W_t \\ X_t &= \Theta(B) W_t \end{aligned}$$

- **Définition.** On appelle Θ le polynôme moyenne mobile.
- *Exemple.* MA(1) : $X_t = W_t + \theta W_{t-1}$. Que valent $\mathbb{E}[X_t]$ et $\rho(h)$?

Trajectoires de processus MA



ACF de processus MA



- **Définition.** On appelle processus autorégressif d'ordre p , noté $AR(p)$, le processus stationnaire X_t vérifiant :

$$\begin{aligned}X_t &= W_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \\ &= W_t + \sum_{i=1}^p \phi_i X_{t-i},\end{aligned}$$

où $\phi_p \neq 0$ et $W_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

- *Remarque.* Si $\mathbb{E}[X_t] = \mu$, on considérera $Y_t = X_t - \mu$.
- *Remarque.*

$$\begin{aligned}X_t &= (\phi_1 B + \dots + \phi_p B^p) X_t + W_t \\ (1 - \phi_1 B - \dots - \phi_p B^p) X_t &= W_t \\ \Phi(B) X_t &= W_t\end{aligned}$$

- **Définition.** On appelle Φ le polynôme autorégressif.

Écriture d'un AR(1) comme processus linéaire

1. Soit X_t un AR(1). Prouver que :

$$X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j W_{t-j}.$$

2. Si $|\phi| < 1$ et X_t stationnaire, on peut donc représenter un AR(1) par un MA(∞) :

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}.$$

3. En déduire :

$$\mathbb{E}[X_t] = 0,$$

$$\gamma(h) = \frac{\sigma^2 \phi^h}{1 - \phi^2},$$

$$\rho(h) = \phi^h.$$

- **Définition.** Lorsque, pour un $AR(p)$ de la forme $\Phi(B)X_t = W_t$, l'écriture

$$X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j} = \Psi(B)W_t$$

existe, on dira que le processus X_t est causal.

- *Remarque.* Dans ce cas, on a

$$\Phi^{-1}(B)\Phi(B)X_t = \Phi^{-1}(B)W_t,$$

soit

$$\Psi(B) = \Phi^{-1}(B).$$

- **Définition.** Lorsque, pour un MA(q) de la forme $X_t = \Theta(B)W_t$, l'écriture

$$W_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = \Pi(B)X_t$$

existe, on dira que le processus X_t est inversible.

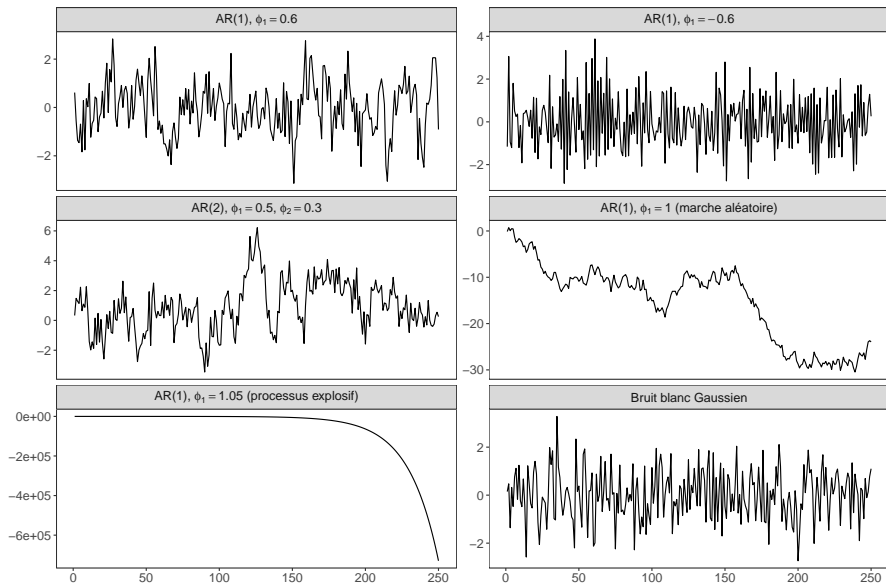
- *Remarque.* Dans ce cas, on a

$$\Theta^{-1}(B)X_t = \Theta^{-1}(B)\Theta(B)W_t,$$

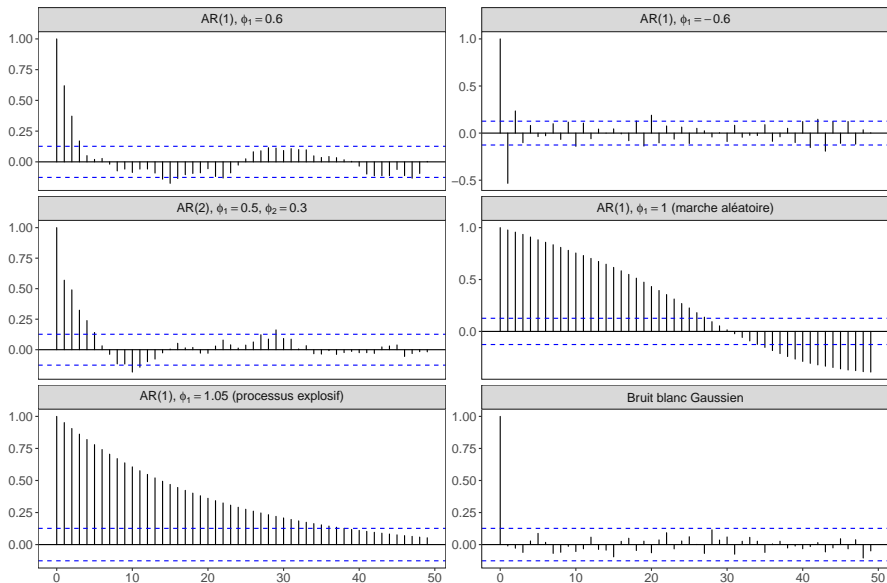
soit

$$\Pi(B) = \Theta^{-1}(B).$$

Trajectoires de processus AR



ACF de processus AR



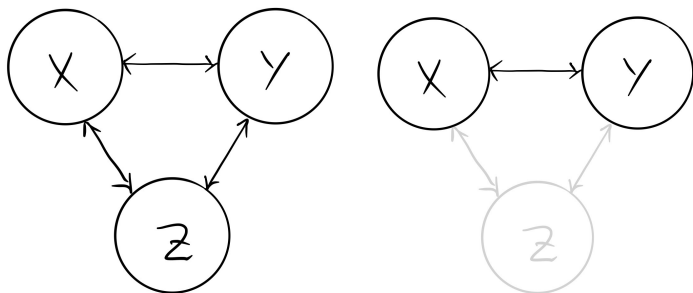
Corrélation partielle

- *Idée.* Contrôler les facteurs de confusion dans le calcul de corrélation.
- **Définition.** Le coefficient de corrélation partielle entre deux variables aléatoires X et Y sachant une troisième variable Z est défini par :

$$\text{Corr}(X, Y|Z) = \text{Corr}(E, F),$$

où E et F sont les résidus des régressions de X et Y sur Z :

$$X = a_1 + b_1Z + E, \quad Y = a_2 + b_2Z + F.$$





- **Définition.** La fonction d'autocorrélation partielle (**PACF**) d'une série stationnaire X_t , notée ϕ_{hh} , est définie par :

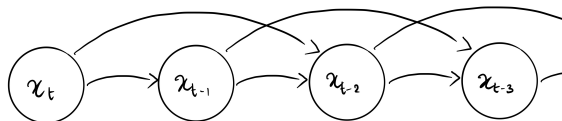
$$\phi_{11} = \text{Corr}(X_t, X_{t-1}) = \rho(1)$$

$$\phi_{hh} = \text{Corr}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h+1}), \quad h \geq 2.$$

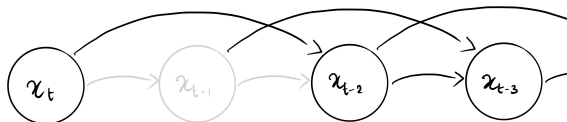
- *Idée.* On retire la dépendance linéaire des termes intermédiaires de la série.

Autocorrélation partielle : processus AR

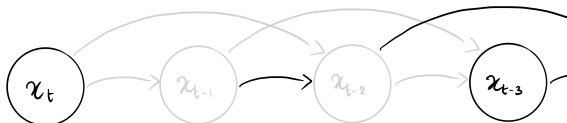
- On considère un processus AR(2) : $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$.



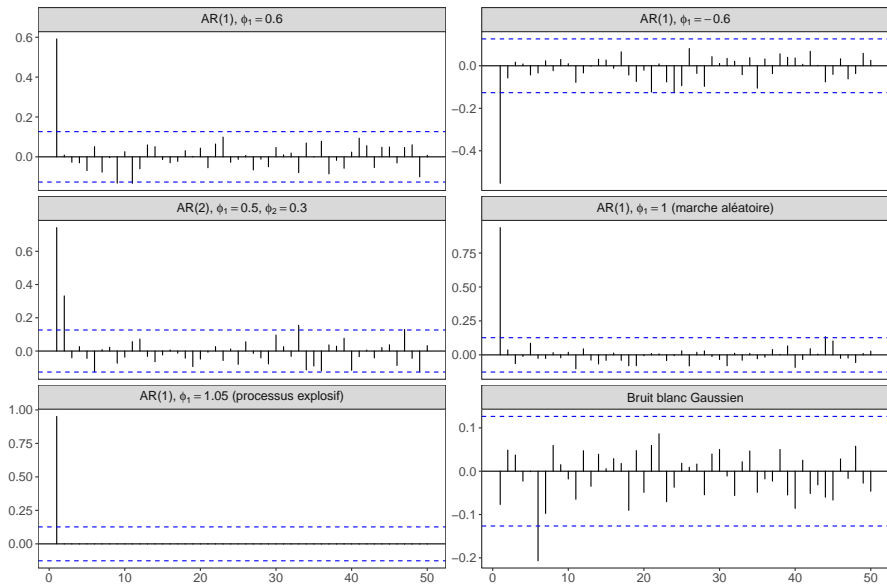
- PACF d'ordre 2 : $\phi_{22} = \text{Corr}(X_t, X_{t-2} | X_{t-1})$.



- PACF d'ordre 3 : $\phi_{33} = \text{Corr}(X_t, X_{t-3} | X_{t-1}, X_{t-2})$.

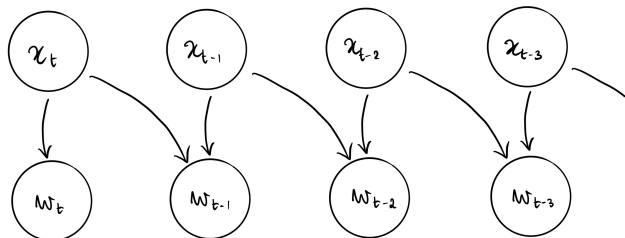


PACF de processus AR

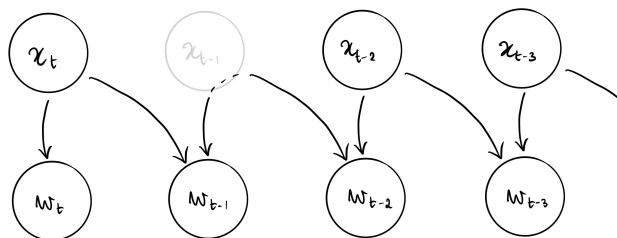


Autocorrélation partielle : processus MA

- On considère un processus MA(1) : $X_t = W_t + \theta_1 W_{t-1}$.

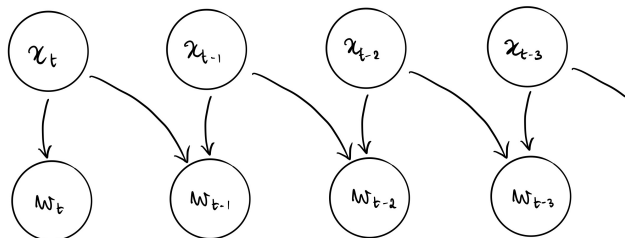


- PACF d'ordre 2 : $\phi_{22} = \text{Corr}(X_t, X_{t-2} | X_{t-1})$.

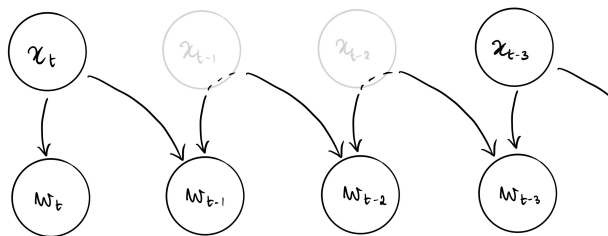


Autocorrélation partielle : processus MA (cont'd)

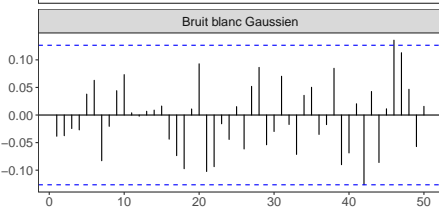
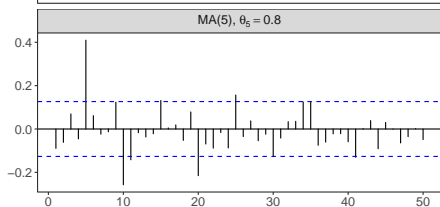
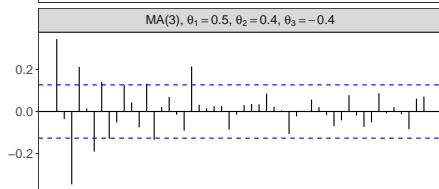
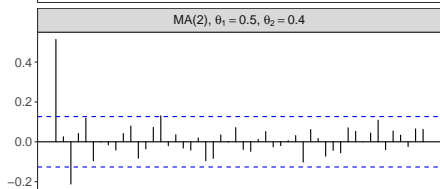
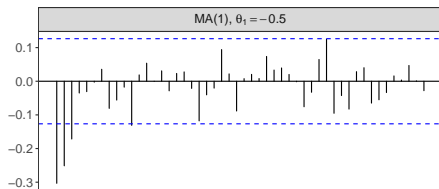
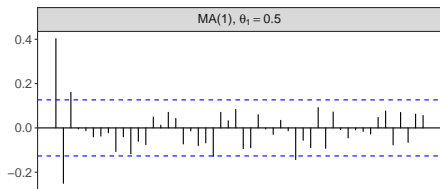
- On considère un processus MA(1) : $X_t = W_t + \theta_1 W_{t-1}$.



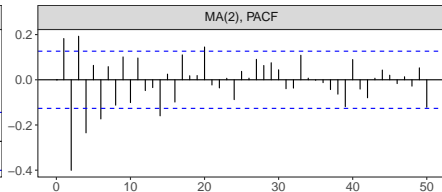
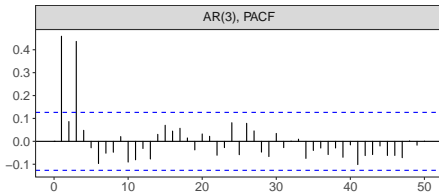
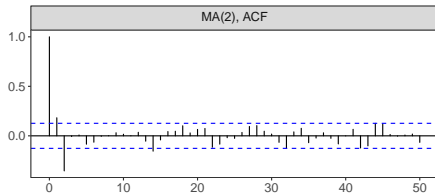
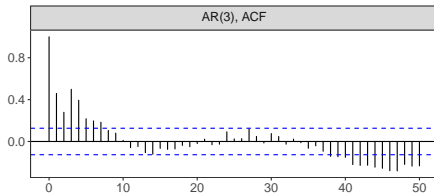
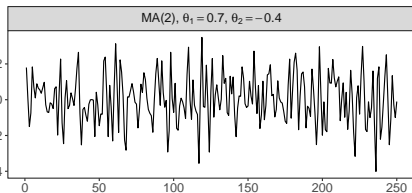
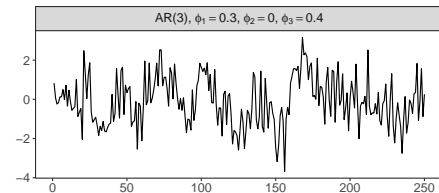
- PACF d'ordre 3 : $\phi_{33} = \text{Corr}(X_t, X_{t-3} | X_{t-1}, X_{t-2})$.



PACF de processus MA



Modélisation : identification des ordres du processus



- **Définition.** On appelle processus autorégressif moyenne mobile d'ordre (p, q) , noté $\text{ARMA}(p, q)$, le processus stationnaire X_t vérifiant :

$$X_t = W_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j W_{t-j},$$

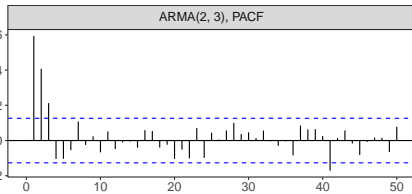
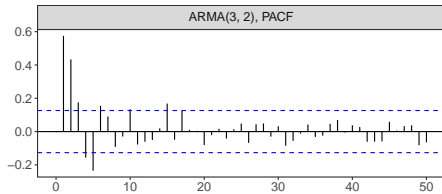
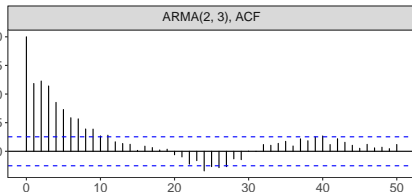
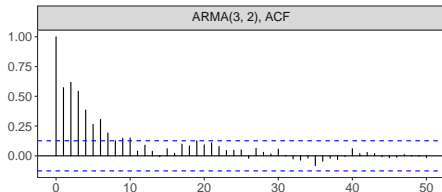
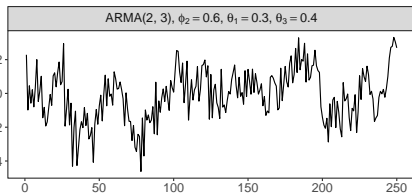
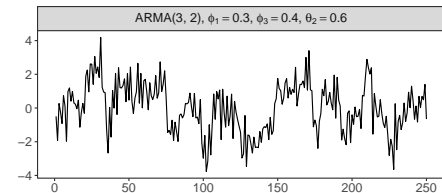
où $\phi_p \neq 0$, $\theta_q \neq 0$ et $W_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

- *Remarque.* En utilisant les polynômes AR Φ et MA Θ , on a :

$$\Phi(B)X_t = \Theta(B)W_t.$$

- **Théorème de représentation de Wold.** Toute série stationnaire au second ordre peut être représentée par un processus $\text{ARMA}(p, q)$.

Identification des ordres du processus ?



Un sophisme

- Soit $X_t = W_t$ un bruit blanc Gaussien.
- On a donc $0.5 X_{t-1} = 0.5 W_{t-1}$.
- Puis $X_t - 0.5 X_{t-1} = W_t - 0.5 W_{t-1}$.
- Donc X_t est un processus ARMA(1, 1).

Redondance paramétrique. La surparamétrisation cache le fait que le processus est un bruit blanc Gaussien :

$$(1 - 0.5B)X_t = (1 - 0.5B)W_t.$$

Pour choisir entre deux modèles, on utilise l'**AIC** ou le **BIC** (plus parcimonieux).

- Basés sur la fonction de vraisemblance ;
- Pénalisent le nombre de paramètres du modèle ;
- Compromis entre la qualité de l'ajustement et la complexité du modèle.

$$\text{AIC} = 2 \#\{\text{paramètres}\} - 2 \log(\text{vraisemblance}),$$

$$\text{BIC} = \log(n) \#\{\text{paramètres}\} - 2 \log(\text{vraisemblance}).$$

- **Définition.** On appelle processus autorégressif intégré moyenne mobile d'ordre (p, d, q) , noté $\text{ARIMA}(p, d, q)$, un processus X_t tel que $\nabla^d X_t = (1 - B)^d X_t$ est un modèle $\text{ARMA}(p, q)$:

$$\Phi(B)\nabla^d X_t = \Theta(B)W_t.$$

- Aucune nouveauté ici : il s'agit principalement d'un jeu d'écriture.
- *Remarque.* L'écriture ARIMA est surtout utile pour \mathbb{R} pour effectuer des prédictions.

- On procède par maximum de vraisemblance : on cherche à maximiser $\mathcal{L}(\mu, \Phi, \Theta, \sigma^2; X_{1:n}) = f_{\mu, \Phi, \Theta, \sigma^2}(X_1, \dots, X_n)$.
- Cette fonction de densité jointe se décompose en conditionnant par rapport aux points du passé :

$$f_{\mu, \Phi, \Theta, \sigma^2}(X_1, \dots, X_n) = f(X_1) f(X_2 | X_1) \cdots f(X_n | X_{1:(n-1)}).$$

- Chaque fonction intermédiaire $f(X_j | X_{1:j})$ est la fonction de densité d'une loi normale univariée.

- **Propriété.** Sous de “bonnes conditions”, pour un processus ARMA **causal** et **inversible**, le maximum de vraisemblance donne des estimateurs optimaux de σ^2 et $\beta = (\mu, \Phi, \Theta)$, et

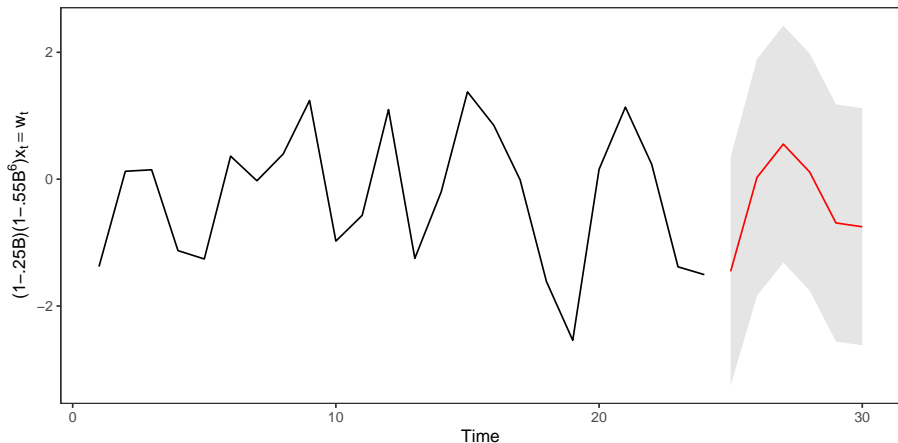
$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}(\beta)).$$

- *Remarque.* Dans le cas d'un modèle surparamétrisé, la variance des estimateurs augmente : $\text{Var}(\hat{\beta}) \nearrow$.

Prédiction

- On cherche à prédire X_{n+m} à partir des observations (X_1, \dots, X_n) .
- On appelle **prédicteur au pas de temps m** la variable aléatoire définie par :

$$X_{n+m}^n = \mathbb{E}[X_{n+m} \mid X_{1:n}].$$



- *Exemple.* Pour un $AR(p)$:

$$\hat{X}_{n+1} : X_{n+1}^n = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1},$$

$$\hat{X}_{n+2} : X_{n+2}^n = \phi_1 X_{n+1}^n + \phi_2 X_n + \dots + \phi_p X_{n-p+2}.$$

- *Exemple.* Pour un $ARMA(p, q)$, il faut d'abord l'écrire sous forme de $MA(\infty)$:

$$X_{n+m} = \sum_{j=0}^{\infty} \psi_j W_{n+m-j}.$$

Alors, on a

$$X_{n+m}^n = \sum_{j=m}^{m+n} \psi_j W_{n+m-j}.$$

- 7 Modèles saisonniers SARIMA
 - ARIMA saisonnier pur
 - Modèles SARIMA

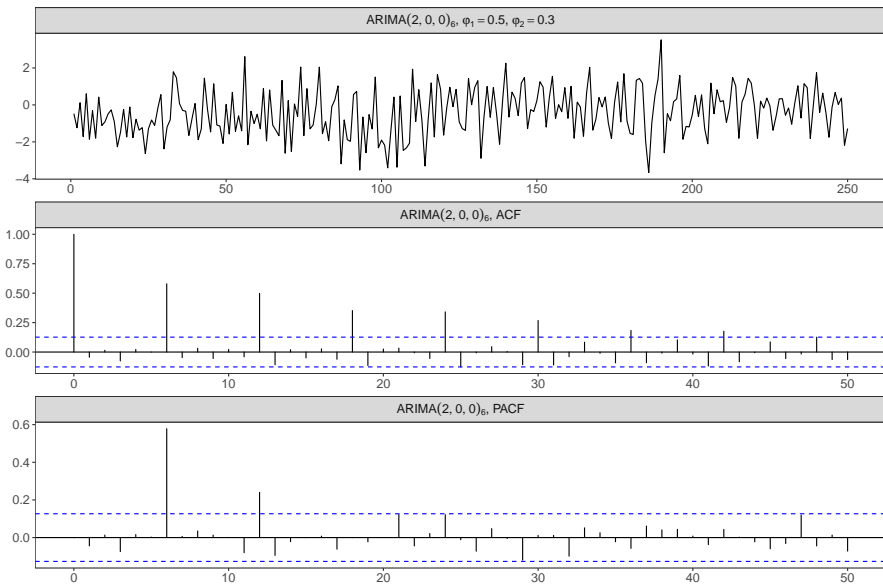
- *Idée.* On souhaite modéliser une structure de dépendance saisonnière.
- **Principe.** On va appliquer un modèle ARIMA, dont les ordres sont des multiples de la saisonnalité S .
- On note $\nabla_S := 1 - B^S$ l'opérateur de différence d'ordre S .
- **Définition.** On appelle processus ARIMA(P, D, Q) $_S$ saisonnier le processus stationnaire X_t vérifiant :

$$\Phi_P(B^S)\nabla_S^D X_t = \Theta_Q(B^S)W_t,$$

où

$$\begin{aligned}\Phi_P(B^S) &= 1 - \varphi_1 B^S - \dots - \varphi_P B^{PS}, \\ \Theta_Q(B^S) &= 1 + \vartheta_1 B^S + \dots + \vartheta_Q B^{QS}.\end{aligned}$$

Trajectoire et (P)ACF d'ARIMA saisonnier

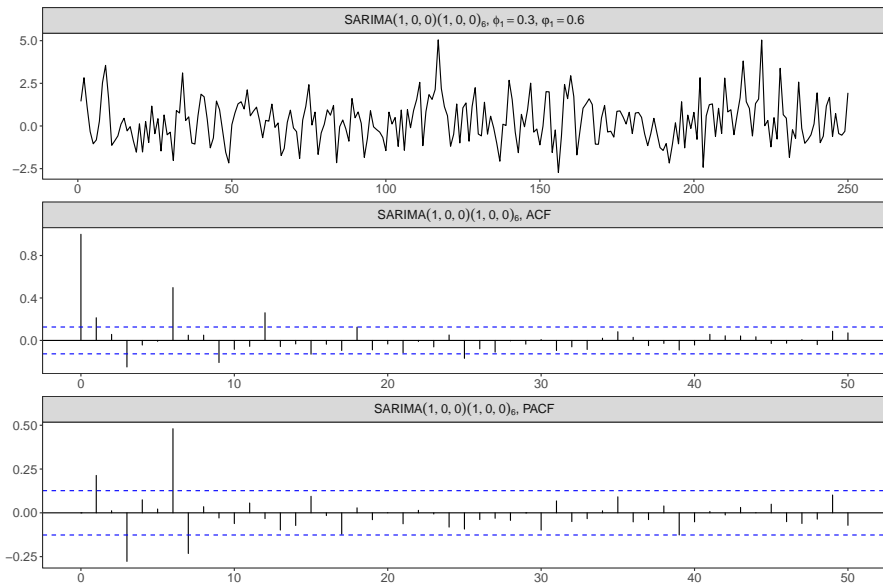


- *En pratique*, il existe généralement deux structures de dépendances :
 - Une à court terme, généralement ≤ 3 pas de temps ;
 - Une à long terme, due à la saisonnalité S de la série.
- **Idée.** On va combiner deux modèles ARIMA, le premier pour expliquer la dépendance à court terme, le deuxième pour la dépendance à long terme.
- **Définition.** On appelle processus SARIMA(p, d, q) \times (P, D, Q) $_S$ le processus stationnaire X_t vérifiant :

$$\Phi_P(B^S)\Phi(B)\nabla_S^D\nabla^d X_t = \Theta_Q(B^S)\Theta(B)W_t,$$

où Φ et Θ sont les polynômes AR et MA de court terme, et Φ_P et Θ_Q sont ceux modélisant la saisonnalité.

Trajectoire et (P)ACF d'ARIMA saisonnier

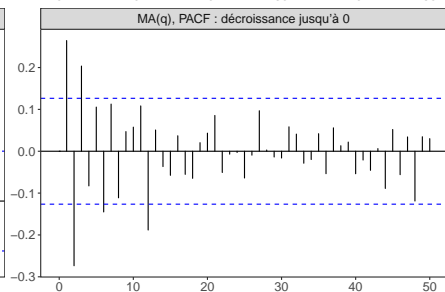
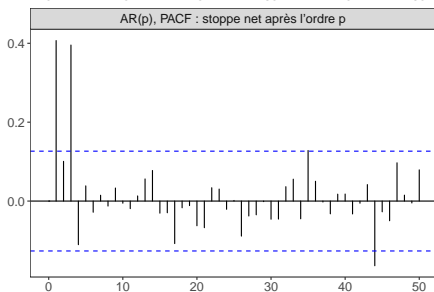
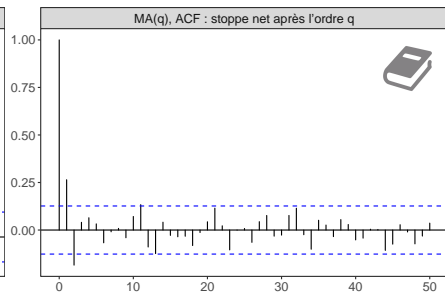
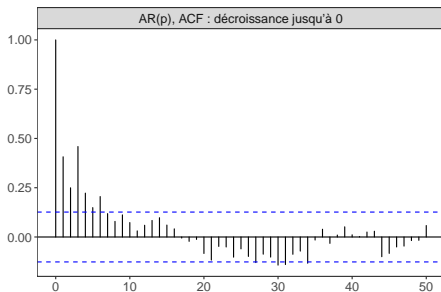


8 Procédure de Box et Jenkins



- 0a. Représenter les données.
- 0b. Transformer les données **si nécessaire**.
 - Régression linéaire ou différentiation ;
 - Stabilisation de la variance par logarithme.
1. **Identification** les ordres p et q du modèle ARMA.
 - ACF (acf) et PACF (pacf) empiriques.
 - Identifier les structures saisonnières avant celles de court terme.
2. **Estimation** des paramètres par maximum de vraisemblance.
 - `arima(series, order=c(p,d,q), seasonal=list(order=c(P,D,Q), period=S))`.
3. **Diagnostic** des résidus.
 - ACF des résidus : existe-t-il des pics de covariance non pris en compte ?
 - Test de Ljung-Box : l'ensemble des pics de covariance est-il conforme avec un bruit blanc ?
 - Graphiques et tests usuels : la normalité des résidus est-elle vérifiée ?
4. Choix du modèle.
 - Utilisation des critères d'information parcimonieux.

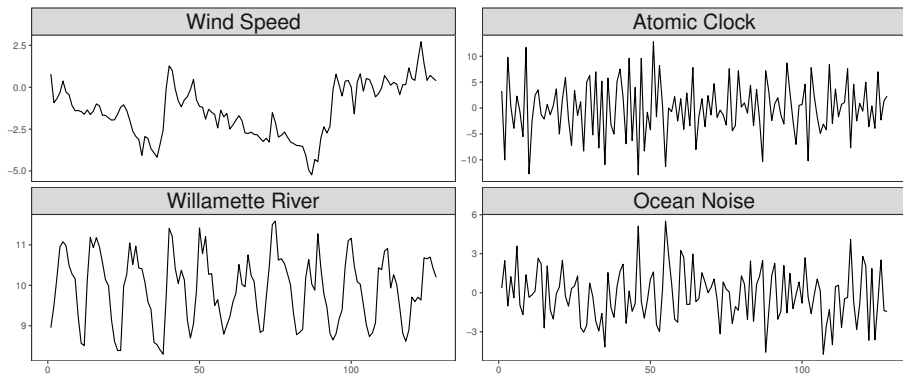
Identification des ordres du processus



- 9 Pour aller plus loin : introduction à l'analyse spectrale

Introduction to spectral analysis

We consider four examples of time series X_1, X_2, \dots, X_{128} . How would you describe them?

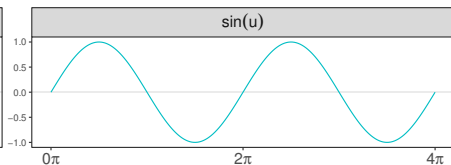
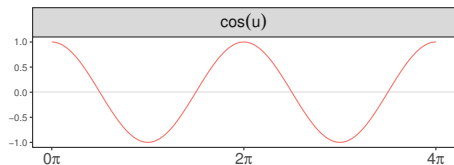


Spectral analysis describes X_t 's by comparing them to sines and cosines.

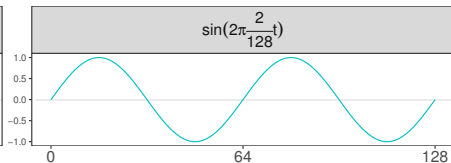
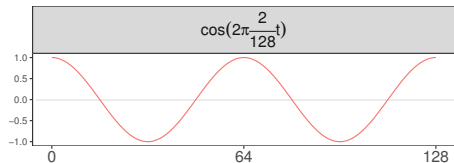
Sines and cosines?

The functions \sin and \cos are 2π -periodic: for $u \in \mathbb{R}$ and $k \in \mathbb{Z}$,

$$\cos(u + 2k\pi) = \cos(u)$$



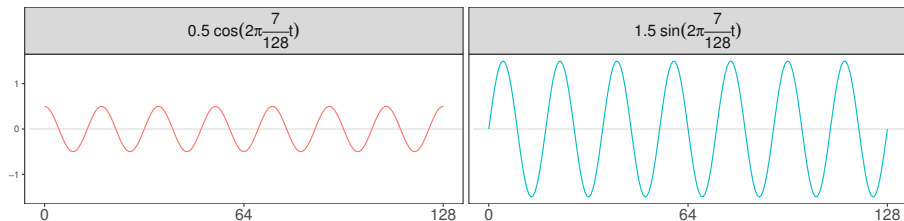
Let $u = 2\pi \frac{2}{128}t$ for $t = 1, 2, \dots, 128$.



$\frac{2}{128}$ can be interpreted as 2 cycles over the time span of 128.

Sines and cosines?

Similarly, with $u = 2\pi\frac{7}{128}t$ for $t = 1, 2, \dots, 128$.



- $\cos(2\pi\frac{k}{n}t)$ and $\sin(2\pi\frac{k}{n}t)$ have k cycles per n time steps.
- The quantity $f = \frac{k}{n}$ is called the *frequency* of the sine or cosine.
 - It is the amount (or rather fraction) of cycles per time step.
 - If f is small (large), the sine is said to have low (high) frequency.
- The *period* $T = \frac{1}{f}$ is the time steps needed for a full cycle.
- The *amplitude* is the maximum range of variation and is equal to 1 for the functions \sin and \cos .

Sines and cosines?

Summing up sines and cosines of different amplitudes and frequencies create time series that resemble actual data.

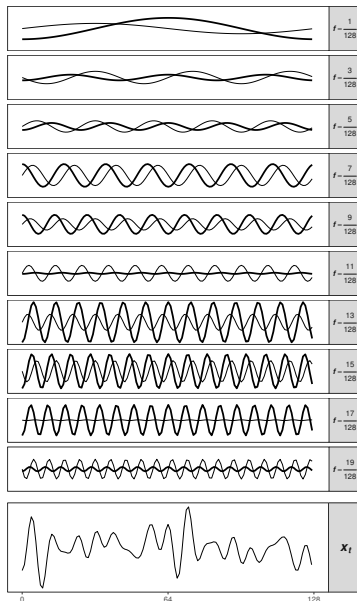
Goal of spectral analysis

Given a time series X_t , figure out its *Fourier representation*, i.e. its decomposition into sines and cosines:

$$X_t = \sum_k a_k \cos\left(2\pi \frac{k}{n} t\right) + b_k \sin\left(2\pi \frac{k}{n} t\right)$$

Actually easy to compute a_k and b_k :

$$a_k \propto \text{Cov} \left\{ X_t, \cos\left(2\pi \frac{k}{n} t\right) \right\}$$



The spectrum and periodogram

$\sigma_k^2 = a_k^2 + b_k^2$, the squared amplitude of the sine-cosine pair, highlights the importance of the frequency $\frac{k}{n}$ in the decomposition of X_t .

- $(\sigma_k^2)_{k=1,\dots,n}$ is called the *spectrum* of the time series X_t .
- If σ_k^2 is large, there are strong patterns of frequency $\frac{k}{n}$.
- The sample spectrum, noted I_k , is called the *periodogram* of X_t .

Decomposition of variance

Since the sines and cosines of different frequencies are uncorrelated, then

$$\text{Var } X_t = \sum_k (a_k^2 + b_k^2) = \sum_k \sigma_k^2$$

The spectrum (σ_k^2) is the decomposition of the variance of the time series X_t into its different frequencies $\frac{k}{n}$.

Examples

Recall the four examples: here are their periodograms.

