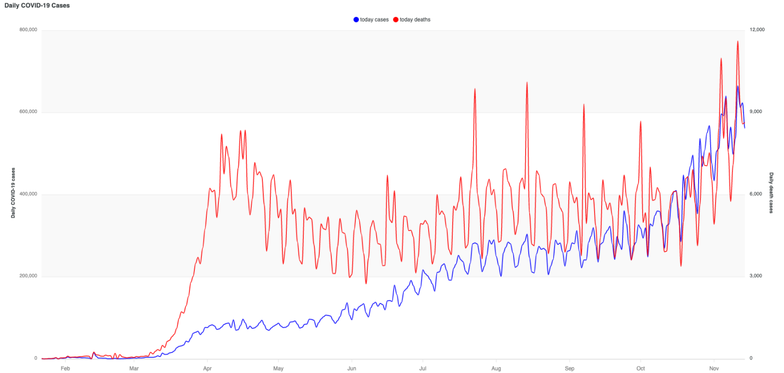


Introduction to Time Series

Dhafer Malouche

Dauphine, Tunis, Feb 2021

Outline

What's a time-series?

Examples, Visualizations

Stationary time series, White Noise, MA and AR time series

Estimation of the autocovariance function

Exercises

What's a time-series?

Types of Data

- Cross-sectionnal data
- Time Series data
- Panel/Longitudinal Data

Cross-sectional data

- This kind of data is obtained by collecting several observations from a sample of individuals extracted from a given population.
- Time doesn't play any role in the analysis.
- Examples:
  - SAT scores of high school students in a particular year.
  - Gross domestic product (GDP) of a set of countries for a specific year.
  - Data from a Survey on the opinion of a sample of people.

Cross-sectional data

- Formally
  - $\{w_1, \ldots, w_n\}$ represents the sample of the observed individuls
  - $\{x_1, \ldots, x_n\}$ represents the observed data, for all $i = 1, \ldots, n$

  $$x_i = X(w_i)$$

  where X is a random vector representing the set of variables of interest.

- The analysis of cross-sectional data aims to estimate the probability distribution of the random vector. This task can be performed by either visualizing or estimating statistical parameters such as central tendency, dispersion, and many other statistics.

Time series data

- A time series data is an observation of an individual for a span of regular times.
- Time series can be heights of ocean tides measured at fixed instant of the day. It can also be the daily closing value of the Dow Jones Industrial Average.
- Formally, a time series data is a sequence $(x_t)_{t \in T}$ where $x_t = X_t(w)$ is the observation of a sequence of random vectors $(X_t)_{t \in T}$ and T can be a subset of $\mathbb{Z}$

Panel/Longitudinal Data

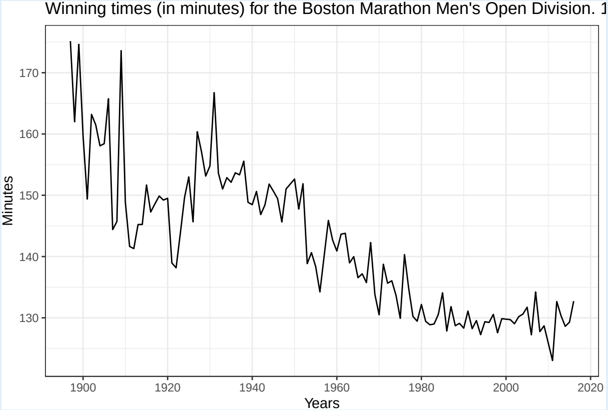
- Panel/Longitudinal data is a time-series data taken from multiple individuals. It's a data obtained when we observe multiple entities over multiple points in time.
- Formally a panel data will be sequence of vectors $(x_t^1, \ldots, x_t^n)_{t \in T}$ where each $i = 1, \ldots, n$ $(x_t^i)_{t \in T}$ is a time-series data.

Examples, Visualizations

Example: CAC France: 1991 to 1990

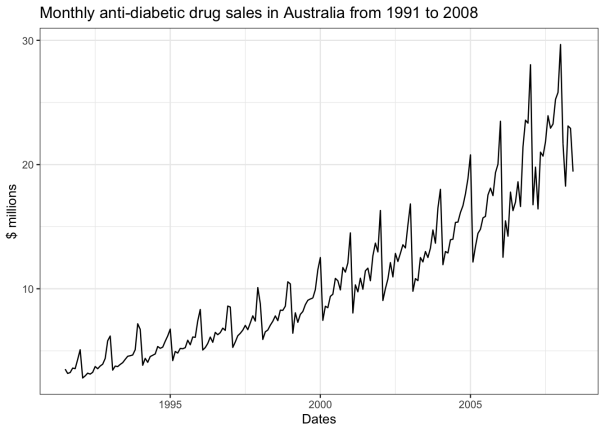
Source: `EuStockMarkets` from `datasets` `R` package (French index, stock market).

Example: Winning times for the Boston Marathon.



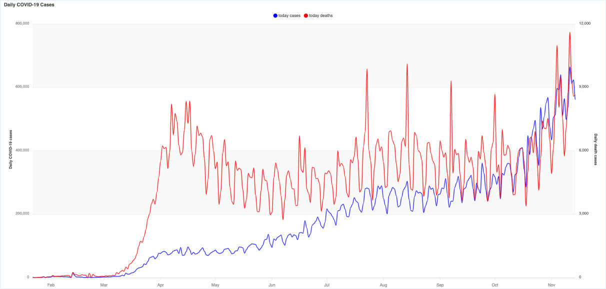
Source: `marathon` data from `fpp2` package.

Example: Anti-diabetic drug sales.



Source: `a10` data from `fpp2` package.

Example: World COVID-19 Cases/Deaths.



Source: John Hopkins University, https://covid19data.website

Visualizing Time Series

1. Import your data file into R/Python
   - R: read_csv , from readr library.
   - Python: pd.read_csv, from pandas library

2. Converting the column containing the date-times to a Date-time Object.

3. Using specific commands for visualization:
   - R: ggplot from ggplot2 library, autoplot from forecast library
   - Python: pyplot from matplotlib library

WDI data with `Python` and `R`

Exercise 1 (Project)

Use `WDI` package (for `R`) and/or `WDdata` package (for `Python`) to extract time series data using interactive visualisations.

1. In `R`: you can use `dygraph`, `rbokeh`, `highcharter`.
2. In `Python`: you can use `dyplot`, `bokeh`.

dygraph package

Stationary time series, White Noise, MA and AR time series

Stochastic process/Time Series

- $\mathrm{X} = (X_t)_{t \in \mathbb{Z}}$ is called a stochastic process or time series, if for all $t \in \mathbb{Z}$, X_t is a random variable:

$$\begin{array}{rcccc} \mathrm{X} & : & \Omega \times T & \to & \mathbb{R} \\ & & (w, t) & \mapsto & X_t(w) \end{array}$$

- We're interested in the Probability distribution of X: $\forall\, n \in \mathbb{N}^*$, $\forall\, (t_1, \ldots, t_n) \in \mathbb{Z}^n$

$$F_{\mathrm{X}}^{t_1, \ldots, t_n}(x_1, \ldots, x_n) = \mathbb{P}\left(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_b\right), \;\; \forall\, (x_1, \ldots, x_n) \in \mathbb{R}^n$$

- $\mathrm{X} = (X_t)_{t \in \mathbb{Z}}$ is a Gaussian stochastic process if $\forall\, n \in \mathbb{N}$, $\forall\, (t_1, \ldots, t_n) \in \mathbb{Z}^n$, $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian random vector.

Mean, autocovariance functions

- The **mean** function, μ_X, is a function from $\mathbb{Z}$ to $\mathbb{R}$ such that $\forall\, t \in \mathbb{Z}$,

  $$\mu_X(t) = \mathbb{E}(X_t)$$

  where $\mathbb{E}(X_t)$ is the expectation of X_t

- The **autocovariance** function (ACF) is defined as the second moment product, i.e., it's a function γ_X from $\mathbb{Z} \times \mathbb{Z}$ into $\mathbb{R}$ such that

  $$\gamma_X(s,t) = \text{cov}(X_t, X_s) = \mathbb{E}\left[(X_t - \mu_X(t))\,(X_s - \mu_X(s))\right]$$

Mean, autocovariance functions

- The **mean** function, μ_X, is a function from $\mathbb{Z}$ to $\mathbb{R}$ such that $\forall\, t \in \mathbb{Z}$,
$$\mu_X(t) = \mathbb{E}(X_t)$$
where $\mathbb{E}(X_t)$ is the expectation of X_t
- The **autocorrelation** function is defined as a function, $\rho_X(s,t)$, from $\mathbb{Z} \times \mathbb{Z}$ into $\mathbb{R}$ such that
$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

Stationary processes, strictly stationary

Let $\mathrm{X} = (X_t)_{t\in\mathbb{Z}}$ be a stochastic process,

A stochastic process is called **strictly stationary** if $\forall\ n \in \mathbb{N}^*$,
$\forall\ n-$uplets $(t_1, \ldots, t_n)$, and $\forall\ h \in \mathbb{Z}$,

the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+h}, \ldots, X_{t_n+h})$ have the same probability distribution, i.e,

$$F_{\mathrm{X}}^{t_1,\ldots,t_n}(x_1, \ldots, x_n) = F_{\mathrm{X}}^{t_1+h,\ldots,t_n+h}(x_1, \ldots, x_n)\ \ \forall\,(x_1, \ldots, x_n) \in \mathbb{R}^n$$

$\Longleftrightarrow$ *The distribution of a number of random variables of the stochastic process is the same as we shift them along the time index.*

Stationary processes, weakly stationary

Let $\mathsf{X} = (X_t)_{t\in\mathbb{Z}}$ be a stochastic process,

A stochastic process is called **weakly stationary** or **stationary** if the following conditions are satisfied

i. $\forall\, t \in \mathbb{Z},\ \mathbb{E}\left[|X_t|^2\right] < \infty$

ii. $\forall\, t,\ h \in \mathbb{Z},\ \mu_X(t) = \mu_X(t+h)$

ii. $\forall\, t,\ s \in \mathbb{Z},\ \gamma_X(t,s) = \gamma_X(t-s,0)$

Examples

- *Example 1:* Let Y be a random variable, and define a stochastic process $(X_t)_{t\in\mathbb{Z}}$, by $X_t = Y$ for all t and $(X_t)_{t\in\mathbb{Z}}$ are mutually independent. Then $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary.

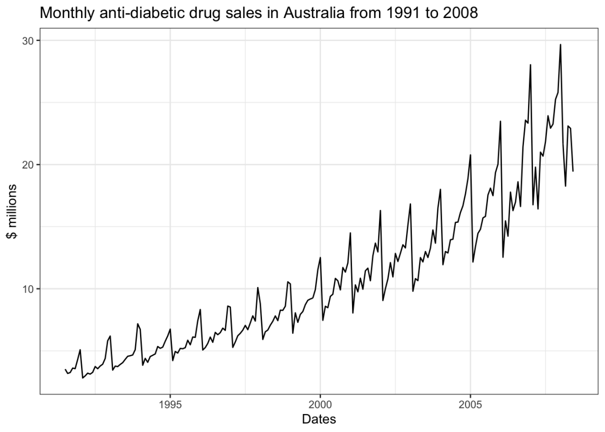
- *Example 2:* Let Y be a random variable with uniform distribution on $[0, 2\pi]$, and define a stochastic process $(X_t)_{t\in\mathbb{Z}}$, by $X_t = \cos(t + Y)$.

- *Example 3:* Let $(X_t)_{t\in\mathbb{Z}}$ be sequence of independent variables such that of t is even $X_t \sim \mathrm{Exp}(1)$ and if t is odd $X_t \sim \mathcal{N}(1, 1)$.

Exercise 2

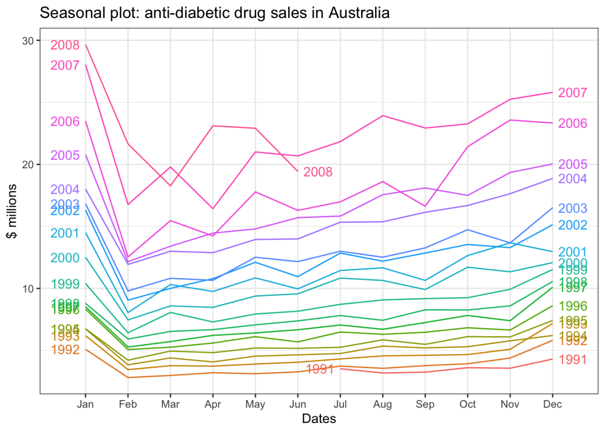
Prove that the processes in Examples 2 & 3 are a stationary, but they are not a strictly stationary.

Example of **non-stationary** process: Anti-diabetic drug sales



Source: `a10` data from `fpp2` package.

Example of **non-stationary** process: Anti-diabetic drug sales, Seasonal plot.



Source: `a10` data from `fpp2` package.

Exercise

Consider the data `a10` in the `R` package `forecaset`

1. Visualize the time series `a10` using `autoplot`
2. Use `ggseasonplot` to draw the time series in every year in the same plot.
3. `a10` can be considered as a stationary time series.

Exercise

Exercise 3

Prove, using simulations, that the mean function of the time series defined in Example 2 is constant. Compare this result if we change the probability distribution of Y with $\mathcal{N}(1, 1)$.

Theorem

Theorem 1

If X is strictly stationary and has finite second moment, then X is stationary.

However, if X is a weakly stationary Gaussian stochastic process, then X is strongly stationary.

ACF of a Stationary processes

$X = (X_t)_{t \in \mathbb{Z}}$ a stationary processes

- The function mean is constant: $\mu = \mathbb{E}(X_t), \forall\, t \in \mathbb{Z}$
- The autocovariance function(ACF) is

$$\gamma_X(h) = \gamma_X(h, 0) = \gamma_X(t + h, t), \;\; \forall\, t,\; h \in \mathbb{Z}$$

- The autocorrelation function is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

ACF of a Stationary processes, Properties

The ACF γ of a stationary process satisfies the following

i. $\gamma(0) \geq 0$

ii. $|\gamma(h)| \leq \gamma(0)$

iii. $\gamma(-h) = \gamma(h)$

iv. γ is a positive semidefinite function, i.e.,
$\forall\, n$, $\forall\,(t_1, \ldots, t_n) \in \mathbb{Z}^n$, $\forall\,(v_1, \ldots, v_n) \in \mathbb{R}^n$:

$$\sum_{j=1}^{n} \sum_{k=1}^{n} v_j v_k \gamma(t_j - t_k) \geq 0$$

Furthermore, any function γ that satisfies (iii) and (iv) is an ACF of some stationary time series.
(iv.) $\Rightarrow$ (i) and (ii).

Exercise

Exercise 4, Proof of ii. and iv.

1. Let $(t_1, \ldots, t_n) \in \mathbb{Z}^n$ and $V = [X_{t_1} - \mathbb{E}(X_{t_1}), \ldots, X_{t_n} - \mathbb{E}(X_{t_n})]$
    1.1 Prove that $\sum_{j=1}^{n} \sum_{k=1}^{n} v_j v_k \gamma(t_j - t_k) = \mathrm{Var}(v^T V)$
    1.2 Deduce then iv.
2. Prove ii. using Cauchy-Schwarz inequality.

ACF of Gaussian process

Theorem 2

In the class of stationary, zero mean, Gaussian processes there is a one-to-one correspondence between the family of the finite dimensional distributions and autocovariance function

$\Longleftrightarrow$ *Gaussian processes are entirely determined by its autocovariance and mean functions*

Example: Periodic processes

- Let A and B two independent random vectors, $\mathbb{E}(A) = \mathbb{E}(B) = 0$ and $\text{Var}(A) = \text{Var}(B) = \sigma^2$, let $\theta \in [-\pi, \pi]$
- The following
$$X_t = A\cos(\theta t) + B\sin(\theta t)$$

Exercise 5

Prove that (X_t) is a stationary process and $\gamma_X(h) = \sigma^2 \cos(\theta h)$.

White Noise

Let $Z = (Z_t)_{t\in\mathbb{Z}}$ be a (weak)-stationary stochastic process.
$(Z_t)_{t\in\mathbb{Z}}$ is a

- *weak white noise* if for every $(s,t) \in \mathbb{Z}\times\mathbb{Z}$, $\gamma_Z(s,t) = 0$ if $s \neq t$. Since Z is stationary then $\gamma_Z(t) = \sigma^2 1_{t=0}$ et the mean function is constant and equal $\mu_Z \in \mathbb{R}$. We denote then $Z \sim \mathrm{WN}(\mu_Z, \sigma^2)$
- *strong white noise* if $Z = (Z_t)_{t\in\mathbb{Z}}$ is a sequence of independent variables.
- If Z is a Gaussian process and a WN, we denote $Z \sim \mathrm{GWN}(\mu_Z, \sigma^2)$

Exercise

Exercise 6

Let X be a random variable with Gaussian distribution $\mathcal{N}(0,1)$, and let

$$Y = X1_{U=1} - X1_{U=0}$$

where U is a random variable with Bernoulli distribution with parameter $1/2$, independent from X.

1. Show that X and Y have the same probability distribution.
2. Show that $\mathrm{Cov}(X, Y) = 0$ and X and Y are not independent.
3. Can you come up with an example of a weak white noise that cannot be considered a strong white noise?

MA and AR processes

- $(X_t)_{t \in \mathbb{Z}}$ is a **moving-average** model of order q, MA(q), if $(X_t)_{t \in \mathbb{Z}}$ can be defined as follows.

$$X_t = Z_t + \sum_{i=1}^{q} \theta_i Z_{t-i}$$

where $(Z_t)_{t \in \mathbb{Z}}$ is a white noise (WN).

- $(X_t)_{t \in \mathbb{Z}}$ is an **autoregressive** model of order p, AR(p), if $(X_t)_{t \in \mathbb{Z}}$ satisfies the following equation.

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + Z_t$$

where $(Z_t)_{t \in \mathbb{Z}}$ is a WN.

Exercise

Exercise 7

Let $(X_t)_{t\in\mathbb{Z}}$ be a **moving-average** model of order 1, MA(1):

$$X_t = Z_t + \theta Z_{t-1}$$

Prove that

1. $\mu_X(t) = 0$, $\forall\, t$
2. $\gamma_X(s,t) = \begin{cases} 0 & \text{if } |s-t| > 1 \\ \theta\sigma^2 & \text{if } |s-t| = 1 \\ (\theta^2+1)\sigma^2 & \text{if } |s-t| = 0 \end{cases}$

Simulating MA with `Python`

Exercise 8

Write a python code simulating the following time series:

1. MA(1) process: $X_t = Z_t + 0.3Z_{t-1}$

2. $X_t = -0.3X_{t-1} + Z_t \iff X_t + 0.3X_{t-1} = Z_t$

hint: use python libraries: `numpy`, `ArmaProcess`

Random Walk with Drift

The stochastic process $(X_t)_{t\in\mathbb{N}}$ defined by

$$X_t = \lambda + X_{t-1} + Z_t, \quad X_0 = 0$$

where $\lambda \in \mathbb{R}$ and $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$ is is called random walk with drift .

- The mean function $\mu_X(t) = t\lambda$ (then it's not a stationary process), since

$$X_t = t\lambda + \sum_{k=1}^{t} Z_k = \sum_{k=1}^{t} (Z_k + \lambda)$$

- Since X_t and $X_{t+h} - X_t$ are independent, then

$$\gamma_X(t, t+h) = \mathrm{Cov}(X_t, X_{t+h} - X_t + X_t) = \mathrm{Var}(X_t) = t\sigma^2$$

- **Conclusion**: a random walk is a non-stationary AR(1).

Estimation of the autocovariance function

Estimators

- Let $x_1, \ldots, x_T$ be a realisation (observations) of a time series $\mathrm{X} = (X_t)_{t\in\mathbb{Z}}$.
- The mean and autocovariance (ACF) functions can be estimated by

$$\overline{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t, \quad \textit{sample mean.}$$

and $\forall, k = 1, \ldots, T-1$

$$\widehat{\gamma}_X(k) = \frac{1}{T}\sum_{t=k+1}^{T} (x_t - \overline{x}_T)(x_{t-k} - \overline{x}_T), \quad \textit{sample ACF.}$$

Estimating ACF with Python

Exercise 9

1. Consider the csv file `avoir_devises.csv`. By using `seaborn` library, draw the time series using `sns.lineplot` and draw the ACF function.

2. Do the same as in 1/ with the data `flights`.

Estimation in the case of i.i.d

Theorem 3

Let $x_1, \ldots, x_T$ be a realisation of a time series $\mathrm{X} = (X_t)_{t\in\mathbb{Z}}$.
Assume that $(X)_{t\in\mathbb{Z}}$ is an i.i.d process where $\mathbb{E}(X_t) = \mu$ and $\mathrm{Var}(X_t) = \sigma^2$, then

1. $\mathbb{E}(\overline{x}_T) = \mu$
2. $\overline{x}_T \xrightarrow[T\to+\infty]{} \mu$ almost surely
3. $\mathbb{E}\left((\overline{x}_T - \mu)^2\right) \xrightarrow[T\to+\infty]{} \dfrac{\sigma^2}{T}$
4. $\dfrac{\overline{x}_T - \mathbb{E}(\overline{x}_T)}{\sqrt{\mathrm{Var}(x_T)}} \xrightarrow[T\to+\infty]{\text{in Law}} \mathcal{N}(0,1)$

Estimation in the case of stationary processes

Theorem 4

Let $x_1, \ldots, x_T$ be a realisation of a time series $\mathrm{X} = (X_t)_{t\in\mathbb{Z}}$.
Assume that $(X)_{t\in\mathbb{Z}}$ is a zero-mean covariance stationary process such that

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty,$$

then $\mathbb{E}(\overline{x}_T) = 0$, $\overline{x}_T \xrightarrow[T\to+\infty]{} 0$ almost surely and

$$T \times \mathbb{E}\left(\overline{x}_T^2\right) \xrightarrow[T\to+\infty]{} \sum_{h=-\infty}^{\infty} \gamma_X(h)$$

Ergodicity

Definition

A stationary stochastic process is called **ergodic**, with respect to a given *population moment*, if the corresponding sample moment for a single finite realisation of length T converges in quadratic mean to the population moment as T increases to ∞.

Remark

Covariance-ergodicity implies mean-ergodicity, but not the reverse.

Ergodicity (intuitive definition)

Definition

- A stochastic process $\mathsf{X} = (X_t)_{t \in \mathbb{Z}}$ is ergodic if any two collections of random variables partitioned far apart in the sequence are essentially independent.
- The stochastic process $\mathsf{X} = (X_t)_{t \in \mathbb{Z}}$ is ergodic if X_t and X_{t-j} are essentially independent if j is large enough.

Ergodicity under Gaussianity

Theorem 5

A stationary process $X = (X_t)_{t\in\mathbb{Z}}$, with mean function μ_X and ACF γ_X. Assume that X is a Gaussian process.

1. If
$$\sum_{k=0}^{\infty} |\gamma_X(T)| < \infty$$
then X is ergodic for all moments.

2. $\lim_{T\to\infty} \gamma_X(T) = 0$ iff X is ergodic for all moments.

Summary (1/2)

- Time series (TS) is a data $(x_t)_{t \in T}$ indexed in time order and supposed to be generated from a sequence of random vectors $(X_t)_{t \in T}$
- TS analysis are methods that can be used to extract meaningful characteristics of the data. It comprises methods to predict future values based on previously observed values.
- Statistics from a TS: mean function, autocovariance, autocorrelation functions
- Two types of stationary TS

$$\begin{array}{ccc} \text{Strictly stationary} & \Rightarrow & \text{Weakly stationary} \\ \mathrm{JPD}(t+h) = \mathrm{JPD}(t) & & \mathrm{Stat}(t+h) = \mathrm{Stat}(t) \end{array}$$

Summary (2/2)

- Estimation of the mean, autocovariance and autocorrelation functions, Ergodicity condition.
- Definition and Simulation of MA, AR and Random walks
- White Noise: weak and strong (strong $\Rightarrow$ weak)
  - Strong WN: mutually independence
  - Weak WN: mutually non-correlated

Exercises

Exercise

Exercise 10

Let $X = (X_t)_{t\in\mathbb{Z}}$ and $Y = (Y_t)_{t\in\mathbb{Z}}$ be two stationary processes and non-correlated, i.e., $\mathrm{cov}(X_t, Y_s) = 0$ for all s, t.
Show that the process $Z = (Z_t)_{t\in\mathbb{Z}}$, defined by

$$Z_t = X_t + Y_t$$

for all $t \in \mathbb{Z}$, is a stationary process and compute γ_Z in function of γ_X and γ_Y

Exercise

Exercise 11

Let $X = (X_t)_{t \in \mathbb{N}}$ be a random walk with drift μ:

$$X_t = \mu + X_{t-1} + Z_t$$

for all $t \geq 1$, and $X_0 = 0$, where $(Z_t)_{t \in \mathbb{N}}$ is a strong white noise.

1. Compute the ACF γ_X of X. Is X a stationary process?
2. Let's consider $\Delta X = (\Delta X_t)_{t \in \mathbb{N}}$, such that $\Delta X_t = X_t - X_{t-1}$. Is ΔX a stationary process?

Exercise

Exercise 12

Find the stationary processes among the following processes:

1. $X_t = Z_t$ if t is even and $X_t = Z_t + 1$ if t is odd, and $(Z_t)_{t\in\mathbb{Z}}$ is stationary;
2. $X_t = Z_1 + \cdots + Z_t$ where $(Z_t)_{t\in\mathbb{Z}}$ is a WN;
3. $X_t = Z_t + \theta Z_{t-1}$, where $(Z_t)_{t\in\mathbb{Z}}$ is a WN and $\theta \in \mathbb{R}$ is a constant;
4. $X_t = Z_t Z_{t-1}$ where $(Z_t)_{t\in\mathbb{Z}}$ is a WN;
5. $Y_t = (-1)^t Z_t$ and $X_t = Y_t + Z_t$ where $(Z_t)_{t\in\mathbb{Z}}$ is a SWN

Exercise

Exercise 13

1. Show that the autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$ of a stationary process is even and of positive type (in fact the equivalence is true but admitted here).

2. Show that the function γ defined by

$$\gamma(h) = 1_{h=0} + \rho 1_{|h|=1}$$

iff $|\rho| \leq 1/2$. Give one example of a stationary process having such a function autocovariance;

Exercise

Exercise 14

Are the following functions autocovariance functions of a process stationary:

1. $\gamma(h) = 1$ si $h = 0$ et $\gamma(h) = 1/h$ si $h \neq 0$;
2. $\gamma(h) = 1 + \cos(h\pi/2)$;
3. $\gamma(h) = (-1)^{|h|}$.