Assignment 2 (ML for TS) - MVA 2022/2023

Clémence Grislain clemence.grislain@eleves.enpc.fr
Roman Plaud roman.plaud@eleves.enpc.fr

May 30, 2023

1 Introduction

Objective. The goal is to better understand the properties of AR and MA processes, and do signal denoising with sparse coding.

Warning and advice.

- Use code from the tutorials as well as from other sources. Do not code yourself well-known procedures (e.g. cross validation or k-means), use an existing implementation.
- The associated notebook contains some hints and several helper functions.
- Be concise. Answers are not expected to be longer than a few sentences (omitting calculations).

Instructions.

- Fill in your names and emails at the top of the document.
- Hand in your report (one per pair of students) by Monday 27$^{\text{th}}$ February 11:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname1.pdf` and
  `FirstnameLastname2_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: .

2 General questions

A time series $\{y_t\}_t$ is a single realisation of a random process $\{Y_t\}_t$ defined on the probability space $(\Omega, \mathcal{F}, P)$, i.e. $y_t = Y_t(w)$ for a given $w \in \Omega$. In classical statistics, several independent realisations are often needed to obtain a "good" estimate (meaning consistent) of the parameters of the process. However, thanks to a stationarity hypothesis and a "short-memory" hypothesis, it is still possible to make "good" estimates. The following question illustrates this fact.

Question 1

An estimator $\hat{\theta}_n$ is consistent if it converges in probability when the number n of samples grows to ∞ to the true value $\theta \in \mathbb{R}$ of a parameter, i.e. $\hat{\theta}_n \xrightarrow{\mathcal{D}} \theta$.

- Recall the rate of convergence of the sample mean for i.i.d. random variables with finite variance.
- Let $\{Y_t\}_{t\geq 1}$ a wide-sense stationary process such that $\sum_k |\gamma(k)| < +\infty$. Show that the sample mean $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$ is consistent and enjoys the same rate of convergence as the i.i.d. case. (Hint: bound $\mathbb{E}[(\bar{Y}_n - \mu)^2]$ with the $\gamma(k)$ and recall that convergence in L_2 implies convergence in probability.)

Answer 1

- Let's $\{X_i\}_{\mathbb{N}}$ be iid random variable with finite variance μ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{\infty} X_i$.

  From the Central limit theorem, we have that

  $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{n\to\infty} \mathcal{N}(0, \sigma^2)$$

  Thus, the rate converge of $\bar{X}_n$ is $\sqrt{n}$.

- $\{Y_t\}_{\mathbb{N}}$ is wide sense stationnary:
  - $\forall t \quad \mathbb{E}(Y_t) = \mu = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} Y_i$
  - $\forall (k,t) \quad \mathbb{E}(Y_t Y_{t+k}) = \gamma_Y(k) = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} Y_i Y_{i+k}$

We have:

$$\mathbb{E}((\bar{Y}_t - \mu)^2) = \mathbb{E}((\frac{1}{t}\sum_{i=1}^{t} Y_i - \mu)^2)$$

$$= \mathbb{E}((\frac{1}{t}\sum_{i=1}^{t} Y_i)^2 + \mu^2 - 2\frac{1}{t}\sum_{i=1}^{t} Y_i \mu)$$

$$= \mathbb{E}((\frac{1}{t}\sum_{i=1}^{t} Y_i)^2) + \mu^2 - 2\frac{1}{t}\sum_{i=1}^{t} \mathbb{E}(Y_i)\mu$$

$$= \mathbb{E}((\frac{1}{t}\sum_{i=1}^{t} Y_i)^2) + \mu^2 - 2\frac{1}{t}\sum_{i=1}^{t} \mu^2$$

$$= \frac{1}{t^2}(\sum_{i=1}^{t} \mathbb{E}(Y_i^2) + 2\sum_{i=1}^{t}\sum_{j=1}^{t} \mathbb{E}(Y_i Y_j)) + \mu^2 - 2\mu^2$$

$$= \frac{1}{t^2}(t\gamma_Y(0) + 2\sum_{i=1}^{t}\sum_{j=i+1}^{t} \gamma_Y(j-i)) - \mu^2$$

with $k = j - i$

$$= \frac{\gamma_Y(0)}{t} + \frac{2}{t^2}\sum_{k=1}^{t}\sum_{j=k+1}^{t+k} \gamma_Y(k) - \mu^2$$

$$= \frac{\gamma_Y(0)}{t} + \frac{2}{t}\sum_{k=1}^{t}\gamma_Y(k) - \mu^2$$

$$= \leq \frac{\gamma_Y(0)}{t} + \frac{2}{t}\sum_{k=1}^{t}|\gamma_Y(k)|$$

$$\leq \frac{\gamma_Y(0)}{t} + \frac{2}{t}\underbrace{\sum_{k=1}^{\infty}|\gamma_Y(k)|}_{l<\infty} = \frac{1}{t}(\gamma_Y(0) + 2l)$$

Hence we have, $\mathbb{E}((\bar{Y}_t - \mu)^2) \xrightarrow{t\to\infty} 0$ in a rate convergence $\frac{1}{t}$

$\implies \bar{Y}_t \xrightarrow{\mathcal{L}_2} \mu$ in a rate convergence $\frac{1}{\sqrt{t}}$

$\implies \bar{Y}_t \xrightarrow{\mathbb{P}} \mu$ in a rate convergence $\frac{1}{\sqrt{t}}$, $\bar{Y}_t$ is a consistent estimator and enjoys the same rate of convergence as the i.i.d. case.

3 AR and MA processes

Question 2 *Infinite order moving average MA(∞)*

Let $\{Y_t\}_{t\geq 0}$ be a random process defined by

$$Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k} \tag{1}$$

where $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ ($\psi = 1$) are square summable, i.e. $\sum_k \psi_k^2 < \infty$ and $\{\varepsilon_t\}_t$ is a zero mean white noise of variance σ_ε^2. (Here, the infinite sum of random variables is the limit in L_2 of the partial sums.)

- Derive $\mathbb{E}(Y_t)$ and $\mathbb{E}(Y_tY_{t-k})$. Is this process weakly stationary?
- Show that the power spectrum of $\{Y_t\}_t$ is $S(f) = \sigma_\varepsilon^2|\phi(e^{-2\pi i f})|^2$ where $\phi(z) = \sum_j \psi_j z^j$. (Assume a sampling frequency of 1 Hz.)

The process $\{Y_t\}_t$ is a moving average of infinite order. Wold's theorem states that any weakly stationary process can be written as the sum of the deterministic process and a stochastic process which has the form (1).

Answer 2

-
  - $\mathbb{E}(Y_t) = \mathbb{E}(\sum_{i=0}^{\infty}\psi_i\epsilon_{t-i}) = \sum_{i=0}^{\infty}\psi_i\mathbb{E}(\epsilon_{t-i}) = 0$
  - $\mathbb{E}(Y_tY_{t-k}) = \mathbb{E}\left((\sum_{i=0}^{\infty}\psi_i\epsilon_{t-i}) \times (\sum_{j=0}^{\infty}\psi_j\epsilon_{t-k-j})\right) = \mathbb{E}\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\epsilon_{t-i}\epsilon_{t-k-j}\right) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\mathbb{E}\left(\epsilon_{t-i}\epsilon_{t-k-j}\right)$

    With $\mathbb{E}\left(\epsilon_{t-i}\epsilon_{t-k-j}\right) = \begin{cases}\sigma_\epsilon^2 & \text{if } i = k+j \\ 0 & \text{otherwise}\end{cases} \implies \mathbb{E}(Y_tY_{t-k}) = \sum_{i=0}^{\infty}\psi_i\psi_{i-k}\sigma_\epsilon^2$

Note: The expected value and the infinte sums were inverted thanks to the Fubini theorem as the square-summability condition on the coefficients ψ_k guarantees that the series converges in the mean square sense.

$\mathbb{E}(Y_t)$ is independent of t, the auto-covariance $\mathbb{E}(Y_t Y_{t-k})$ only depends k and $Var(Y_t) = \mathbb{E}(Y_t^2) = \sigma_\epsilon^2 \underbrace{\sum_{i=1}^{\infty} \psi^2}_{const<\infty} = \sigma_\epsilon^2 \times const$ is a constant. Thus, $\{Y_t\}_{\mathbb{N}}$ is weakly stationary.

- With $f_s = 1$,

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-2i\pi \times f \times k}$$

$$= \sigma_\epsilon^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j-k} e^{-2i\pi \times f \times k}$$

$$= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \sum_{k=-\infty}^{\infty} \psi_{j-k} e^{-2i\pi \times f \times k}$$

With l=j-k

$$= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \sum_{l=-\infty}^{\infty} \psi_l e^{-2i\pi \times f \times (j-l)}$$

$$= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j e^{-2i\pi \times f \times j} \sum_{l=-\infty}^{\infty} \psi_l e^{+2i\pi \times f \times l}$$

With $\psi_l = 0$ if $l < 0$

$$= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j e^{-2i\pi \times f \times j} \sum_{l=0}^{\infty} \psi_l e^{+2i\pi \times f \times l}$$

$$= \sigma_\epsilon^2 |\sum_{l=0}^{\infty} \psi_l e^{-2i\pi \times f \times l}|^2 = \sigma_\epsilon^2 |\phi(e^{-2i\pi \times f})|^2 \quad \square$$

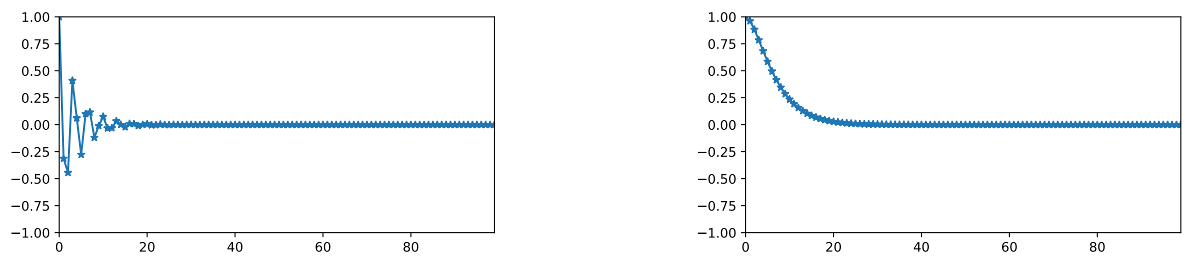
Question 3 *AR(2) process*

Let $\{Y_t\}_{t\geq 1}$ be an AR(2) process, i.e.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

with $\phi_1, \phi_2 \in \mathbb{R}$. The associated characteristic polynomial is $\phi(z) := 1 - \phi_1 z - \phi_2 z^2$. Assume that ϕ has two distinct roots (possibly complex) r_1 and r_2 such that $|r_i| > 1$. Properties on the roots of this polynomial drive the behaviour of this process.

- Express the autocovariance coefficients $\gamma(\tau)$ using the roots r_1 and r_2.
- Figure 1 shows the correlograms of two different AR(2) processes. Can you tell which one has complex roots and which one has real roots?
- Express the power spectrum $S(f)$ (assume the sampling frequency is 1 Hz) using $\phi(\cdot)$.
- Choose ϕ_1 and ϕ_2 such that the characteristic polynomial has two complex conjugate roots of norm $r = 1.05$ and phase $\theta = 2\pi/6$. Simulate the process $\{Y_t\}_t$ (with $n = 2000$) and display the signal and the periodogram (use a smooth estimator) on Figure 2. What do you observe?



Correlogram of the first AR(2)   Correlogram of the second AR(2)

Figure 1: Two AR(2) processes

Answer 3

- With independance property of Y_τ and ε_t, we can show that

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2)$$

  . γ is then the solution of a second order sequence whose characteriscal polynomial roots are $\frac{1}{r_1}$ and $\frac{1}{r_2}$ so that.

  - If $r_1, r_2 \in \mathbb{R}^2$, $\exists (\lambda, \mu) \in \mathbb{R}^2$ such that :

$$\gamma(\tau) = \frac{\lambda}{r_1^\tau} + \frac{\mu}{r_2^\tau}$$

  - If $r_1 = r_2 \in \mathbb{R}$, $\exists (\lambda, \mu) \in \mathbb{R}^2$ such that :

$$\gamma(\tau) = \frac{\lambda}{r_1^\tau} + \frac{\mu\tau}{r_2^\tau}$$

– If $r_1, r_2 \in \mathbb{C}^2, r_1 = r\exp^{i\theta}, r_1 = r\exp^{-i\theta}$ $\exists(\lambda, \mu) \in \mathbb{R}^2$ such that :

$$\gamma(\tau) = \frac{1}{r^\tau}(\lambda\cos(\tau\theta) + \mu\sin(\tau\theta))$$

- Observing the general form of γ thanks to previous question we can say that the oscillating correlogram correspond to the complex roots whereas the non-increasing one correspond to the real roots.

- To express the power spectrum of $S(f)$, we will show that an AR(2) process can be expressed as a MA(∞) process with characteristic polynomial $\frac{1}{\phi(.)}$.

  Let's L be the Lag operator such that $L^k X_t = X_{t-k}$. We can note that a MA(∞) process $Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}$ can be expressed thanks to the Lag operator as

$$Y_t = \sum_{k=0}^{\infty}\psi_k L^k\varepsilon_t = \phi(L)\varepsilon_t$$

  where $\phi(z) = \sum_j \psi_j z^j$ is the characteristic polynomial, and the power spectrum (with $f_s = 1$) is $S(f) = \sigma_\epsilon^2|\phi(e^{-2i\pi\times f})|^2$.

  AR(2) process $Y_t = \phi_1 Y_{t-1} +_2 Y_{t-2}$ can be expressed using the Lag operator $Y_t = (\phi_1 L + \phi_2 L^2)Y_t + \epsilon_t \implies \phi(L)Y_t = \epsilon_t$

$$\implies Y_t = \frac{1}{\phi(L)}\epsilon_t$$

  where $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$.

  The polynomial $\phi(z)$ can be factorized thanks to its roots r_1 and r_2: $\phi(z) = (1 - \frac{1}{r_1}z)(1 - \frac{1}{r_2}z)$

$$\implies Y_t = \frac{1}{(1 - \frac{1}{r_1}z)(1 - \frac{1}{r_2}z)}\epsilon_t$$

$$\implies Y_t = \left(\sum_{i=0}^{\infty}\frac{1}{r_1^i}L^i\right)\left(\sum_{j=0}^{\infty}\frac{1}{r_2^j}L^j\right)\epsilon_t$$

  as $|\frac{1}{r_1}| < 1$ and $|\frac{1}{r_2}| < 1$

$$\implies Y_t = \left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{1}{r_1^i}\frac{1}{r_2^j}L^{i+j}\right)\epsilon_t$$

  with $k = i + j$

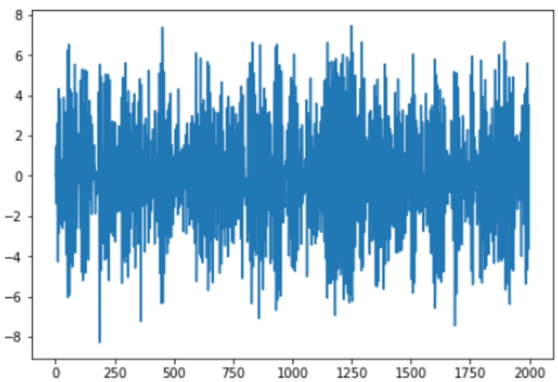
$$\implies Y_t = \left(\sum_{i=0}^{\infty}\sum_{k=i}^{\infty}\frac{1}{r_1^i}\frac{1}{r_2^{k-i}}L^k\right)\epsilon_t$$

$$\implies Y_t = \left(\sum_{k=0}^{\infty}\underbrace{\left(\sum_{i=0}^{k}\frac{1}{r_1^i}\frac{1}{r_2^{k-i}}\right)}_{a_k}L^k\right)\epsilon_t$$

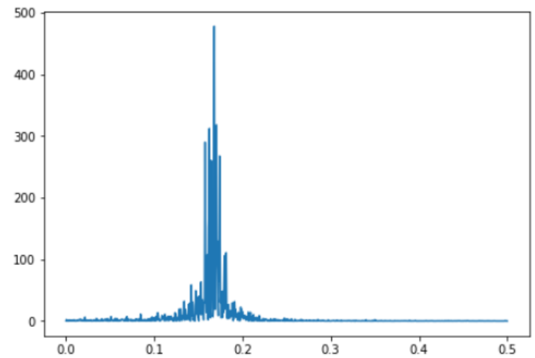
$$\implies Y_t = \left(\sum_{k=0}^{\infty} a_k L^k \right) \epsilon_t$$

An AR(2) process Y_t can be expressed as a MA(∞) process with characteristic polynomial $\frac{1}{\phi(.)}$ thus, its power spectrum with $fs = 1$ is:

$$S(f) = \sigma_\epsilon^2 \left| \frac{1}{\phi(e^{-2i\pi \times f})} \right|^2$$



Signal



Periodogram

Figure 2: AR(2) process

4 Sparse coding

The modulated discrete cosine transform (MDCT) is a signal transformation often used in sound processing applications (for instance to encode a MP3 file). A MDCT atom $\phi_{L,k}$ is defined for a length 2L and a frequency localisation k ($k = 0, \dots, L-1$) by

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{3}$$

where w_L is a modulating window given by

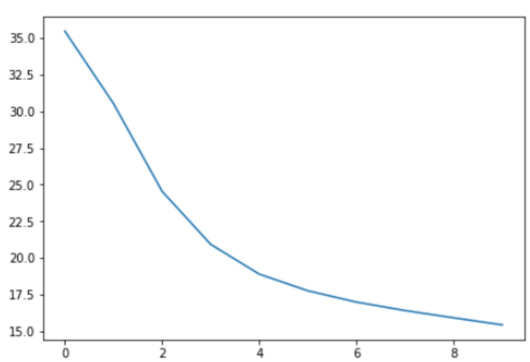
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

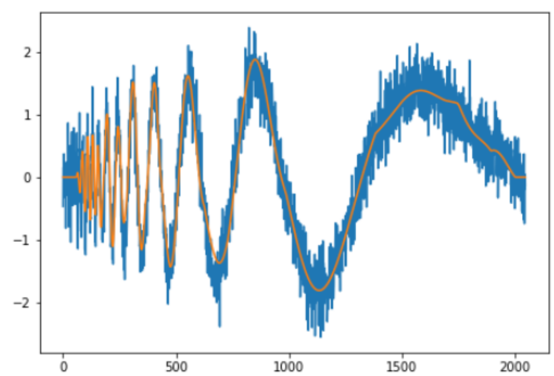
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- For the sparse coding, implement the Orthogonal Matching Pursuit (OMP). (Use convolutions to compute the correlations coefficients.)
- Display the norm of the successive residuals and the reconstructed signal with 10 atoms.

Answer 4



Norms of the successive residuals



Reconstruction with 10 atoms

Figure 3: Question 4