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

Matias Etcheverry matias.etcheverry9@gmail.com
Amric Trudel amric.trudel@ens-paris-saclay.fr

April 11, 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 $(Y_n)_{n\geq 1}$ be a sequence of i.i.d. variables with finite variance σ^2. We define, an estimator of the mean $\mu = \mathbb{E}[Y_i]$

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

  We use the Bienaymé-Tchebychev inequality, knowing that $\mathbb{E}[\overline{Y}_n] = \mu$ and $\mathrm{Var}(\overline{Y}_n) = \frac{\sigma^2}{n}$. Let $\epsilon > 0$, we have:

$$\mathbb{P}(|\overline{Y}_n - \mu| \geq \epsilon) \leq \frac{2\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0$$

  Thus, $\overline{Y}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mu$ at the rate $\frac{1}{n}$.

- Let's prove the the L_2 convergence:

$$\begin{aligned}
\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)\right)\left(\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu)\right)\right] \\
&= \frac{1}{n^2}\left(2\sum_{i=1}^{n}\sum_{j>i}^{n}\mathbb{E}\left[(Y_i - \mu)\left(Y_j - \mu\right)\right] + \sum_{i=1}^{n}\underbrace{\mathbb{E}\left[(Y_i - \mu)^2\right]}_{\gamma(0)}\right) \\
&= \frac{1}{n^2}\left(2\sum_{i=1}^{n}\sum_{k=1}^{n-i}\gamma(k) + n\gamma(0)\right) \\
&= \frac{2}{n^2}\sum_{i=1}^{n}(n-i)\gamma(i) + \frac{1}{n}\gamma(0) \\
&\leq \frac{2}{n}\sum_{i=0}^{n}\gamma(k) \\
&\leq \frac{2}{n}\sum_{i=0}^{\infty}|\gamma(k)|
\end{aligned}$$

  Let $C > 0$ such that $\sum\limits_{k=0}^{\infty}|\gamma(k)| = C$. We have:

$$\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right] \leq \frac{2C}{n} \xrightarrow[n\to\infty]{} 0$$

We have $\overline{Y}_n \xrightarrow[n\to\infty]{L_2} \mu$. Thus $\boxed{\overline{Y}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mu}$, i.e. $\overline{Y}_n$ is consistent. In order to find the rate of consistency, we use the Bienaymé-Tchebychev inequality again. Let $\epsilon > 0$, we have:

$$\begin{aligned}\mathbb{P}(|\overline{Y}_n - \mu| \geq \epsilon) &\leq \frac{2\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right]}{\epsilon^2} \\ &\leq \frac{4C}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

Thus, $\overline{Y}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mu$ at the rate $\frac{1}{n}$.

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 $\mathbb{E}(Y_t Y_{t-k})$. 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

- We have

$$\begin{aligned}
\mathbb{E}[Y_t] &= \sum_{k=0}^{\infty} \mathbb{E}[\psi_k \varepsilon_{t-k}] \\
&= \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\varepsilon_{t-k}] \quad (\psi_k \text{ is deterministic}) \\
&= 0 \quad (\varepsilon_k \text{ is a zero mean white noise}).
\end{aligned}$$

  Moreover,

$$\begin{aligned}
\mathbb{E}[Y_t Y_{t+k}] &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t+k-j}\right)\right] \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \underbrace{\mathbb{E}\left[\varepsilon_{t-i}\varepsilon_{t+k-j}\right]}_{=0 \text{ if } t-i \neq t+k-j} \\
&= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{k+i} \\
&= \sigma_\varepsilon^2 \gamma(k)
\end{aligned}$$

  Thus, $\mathbb{E}[Y_t Y_{t+k}]$ depends only on k, and not on t, i.e. $\{Y_t\}_{t\geq 0}$ is a weakly stationary process.

- Let's first compute $\left|\sum_{j=0}^{N} \psi_j e^{-2i\pi f j}\right|^2$, for $N \in \mathbb{N}$:

$$\begin{aligned}
\left|\sum_{j=0}^{N} \psi_j e^{-2i\pi f j}\right|^2 &= \left(\sum_{j=0}^{N} \psi_j e^{-2i\pi f j}\right)\left(\sum_{l=0}^{N} \psi_l e^{2i\pi f l}\right) \\
&= \sum_{j=0}^{N}\sum_{l=0}^{N} \psi_j \psi_l e^{-2i\pi f (j-l)} \\
&= \sum_{\tau=-N+1}^{N-1} \sum_{n=0}^{N-\tau-1} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau} \quad \text{(using the same trick as in Assignment 1)}
\end{aligned}$$

We make $N \to \infty$, and we finally get:

$$\left|\phi\left(e^{-2\pi i f}\right)\right|^2 = \sum_{\tau=-\infty}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau}$$

Let's compute the power spectrum, let f:

$$\begin{aligned}
S(f) &= \sum_{\tau=-\infty}^{\tau=+\infty} \gamma(\tau) e^{-2i\pi f \tau} \quad \text{with } f_s = 1\text{ Hz} \\
&= \sigma_\epsilon^2 \sum_{\tau=-\infty}^{\tau=+\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau}
\end{aligned}$$

$$\text{i.e.} \quad \boxed{S(f) = \sigma_\epsilon^2 \left|\phi\left(e^{-2\pi i f}\right)\right|^2}$$

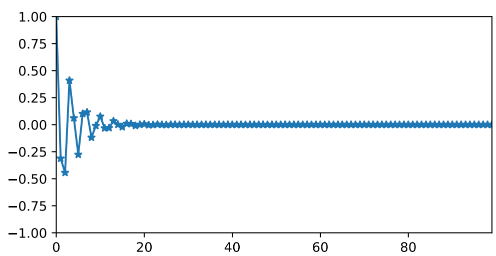
Question 3 *AR(2) process*

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

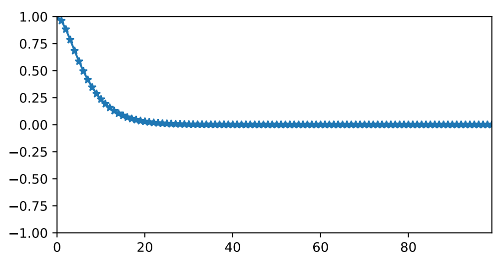
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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

- Let's compute the autocovariance. We suppose that $\mathbb{E}[\epsilon] = 0$. Let $\tau \geq 2$,:

$$\begin{aligned}\gamma(\tau) &= \mathbb{E}[Y_t Y_{t+\tau}] \\ &= \mathbb{E}[Y_t \phi_1 Y_{t+\tau-1} + Y_t \phi_2 Y_{t+\tau-2} + Y_t \epsilon_t] \\ &= \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2)\end{aligned}$$

  i.e. γ is solution of the characteristic polynomial. It should be noted that this characteristic polynomial is the reciprocal of the usually defined characteristic polynomial. Thus,

  1. if $r_1, r_2 \in \mathbb{R}$, then it exists a unique $\lambda, \mu \in \mathbb{R}$ such that for all τ:

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

  2. if $r_1, r_2 \in \mathbb{C}$, i.e. $r_1 = re^{i\theta}$ and $r_1 = re^{-i\theta}$ (with $r > 0$ and $\theta \in \mathbb{R}$) then it exits a unique $\lambda, \mu \in \mathbb{R}$ such that for all τ:

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

In either case, $\gamma(0)$ and $\gamma(1)$ are required to estimate λ and μ.

- Thanks to the answer above, we understand that $r_1, r_2 \in \mathbb{C}$ corresponds to an oscillating system, fading to 0. Thus, the correlogram on the left corresponds to complex roots.

  Similarly, $r_1, r_2 \in \mathbb{R}$ corresponds to a second order system, fading to 0. Thus, the correlogram on the right corresponds to real roots.

- The general idea is to transform the AR(2) process into a MA(∞) process. The idea is inspired from this thread. We introduce the lag operator L such that $LY_t = Y_{t-1}$. We thus have:

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = \epsilon_t$$

  And

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 \\ &= (1 - \frac{1}{r_1} z)(1 - \frac{1}{r_2} z) \end{aligned}$$

  Thus, with $z = L$

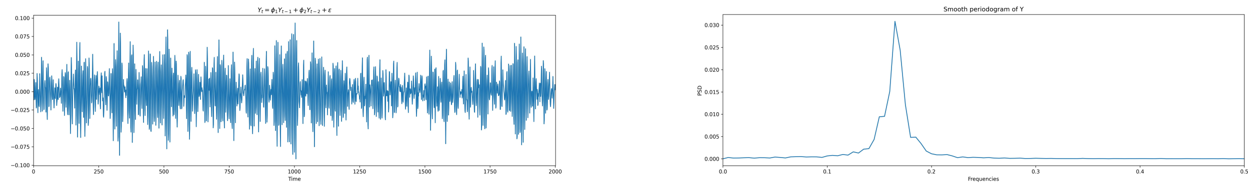
$$Y_t = \frac{1}{(1 - \frac{1}{r_1} L)(1 - \frac{1}{r_2} L)} \epsilon_t$$

  Mentally, the fraction term corresponds to an operator applied to the sequence ϵ_t. A full formula can be found for this operator, using a geometric serie, as $|r_i| > 1$:

$$\begin{aligned} Y_t &= \left(\sum_{i=0}^{\infty} \frac{1}{r_1^i} L^i \right) \left(\sum_{j=0}^{\infty} \frac{1}{r_2^j} L^j \right) \epsilon_t \\ &= \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{r_1^i r_2^j} L^{i+j} \right) \epsilon_t \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{r_1^i r_2^j} \epsilon_{t-(i+j)} \\ &= \sum_{k=0}^{\infty} \underbrace{\sum_{\substack{i,j=0 \\ i+j=k}}^{k} \frac{1}{r_1^i r_2^j}}_{\psi_k} \epsilon_{t-k} \\ &= \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \end{aligned}$$

  We recognize a MA(∞) where the coefficients ψ_k are sums of product of the 2 roots. we can keep the formula $S(f) = \sigma_\epsilon^2 \left|\phi\left(e^{-2\pi i f}\right)\right|^2$ with $\psi_k = \sum\limits_{\substack{i,j=0 \\ i+j=k}}^{k} \frac{1}{r_1^i r_2^j}$

- We have:



Signal Periodogram

Figure 2: AR(2) process

We observe a peak in the periodogram at $f \approx 0.17$ Hz. Indeed, having complex roots with $\theta = \frac{2\pi}{6}$ implies a rotating frequency f verifying $2\pi f = \frac{2\pi}{6}$ i.e. $f = \frac{1}{6} \approx 0.17$ Hz.

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 ($k = 0, \ldots, L-1$) by

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

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

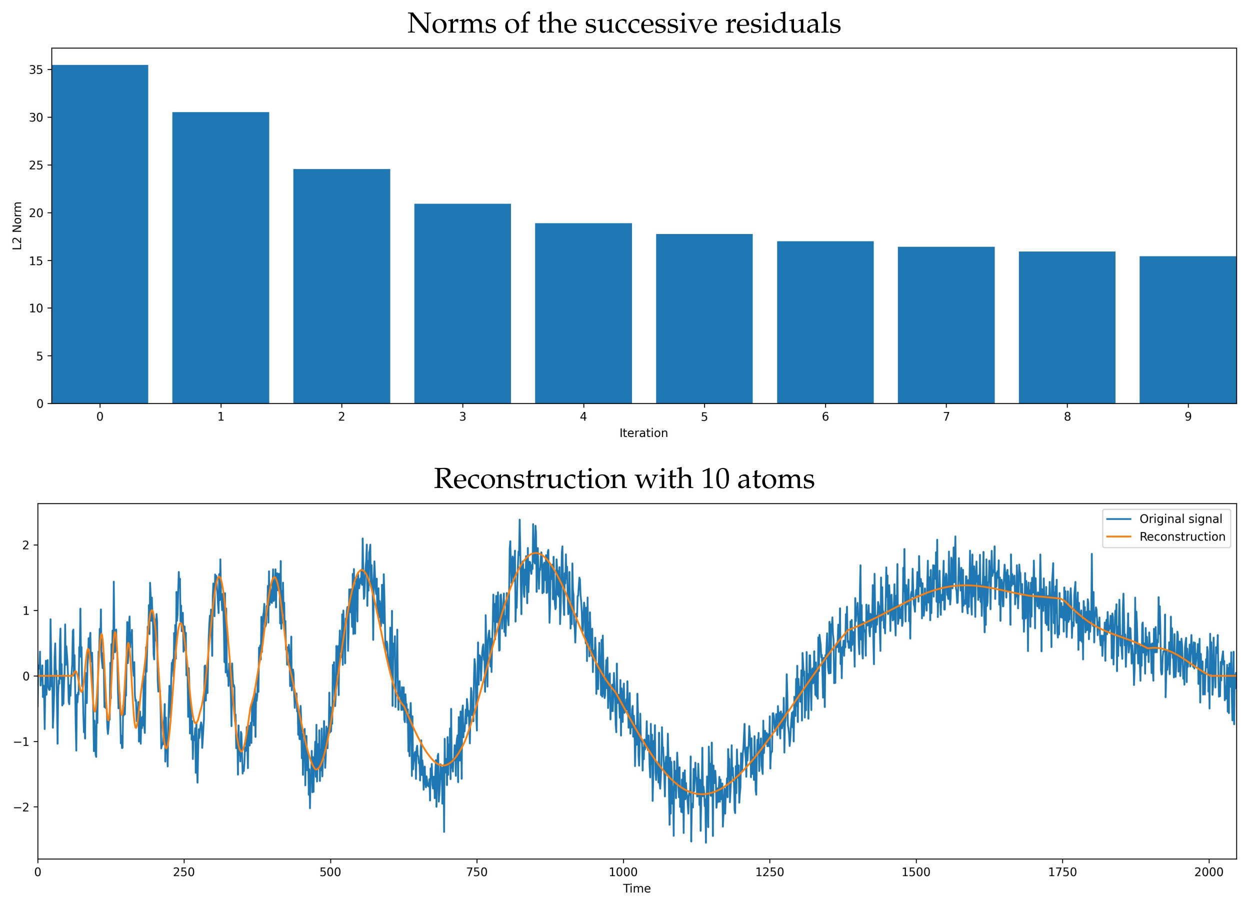


Figure 3: Question 4