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

Thibault de Surrel thibault.de-surrel@ensta-paris.fr
Alexandre Pradeilles alexandre.pradeilles@gmail.com

February 27, 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 $(Y_n)_{n\in\mathbb{N}}$ a sequence of i.i.d. random variables with finite variance. According to the weak law of large numbers, we now that the sample mean $\bar{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$ converges in probability to $\mathbb{E}[Y_1]$. Moreover, according to the central limit theorem, we have that

$$\sqrt{n}\frac{\bar{Y}_n - \mathbb{E}[Y_1]}{\sqrt{\mathrm{Var}(Y_1)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

  $\boxed{\text{The rate of convergence is therefore } 1/\sqrt{n}.}$

- As $\{Y_t\}_{t\geq 1}$ is a wide-sense stationary process, we can denote $\mu := \mathbb{E}[Y_1]$ and we have that $\mu = \mathbb{E}[Y_t]$ for all $t \geq 1$. In order to show that $\bar{Y}_n$ is consistent, it is sufficient to show that $\bar{Y}_n$ converges to μ in L^2, that is $\mathbb{E}[(\bar{Y}_n - \mu)^2] \xrightarrow[n\to\infty]{} 0$. As we have $\bar{Y}_n - \mu = \frac{1}{n}\sum_{i=1}^n (Y_i - \mu)$, we have

$$\begin{aligned}
\mathbb{E}\left[(\bar{Y}_n - \mu)^2\right] &= \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (Y_i - \mu)(Y_j - \mu)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left[(Y_i - \mu)(Y_j - \mu)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \gamma(|i-j|)
\end{aligned} \tag{1}$$

  because $\{Y_t\}_{t\geq 1}$ a wide-sense stationary process, so $\mathbb{E}[(Y_i - \mu)(Y_j - \mu)] = \gamma(|i-j|)$. Since we have $1 - n \leq i - n \leq i - j \leq i - 1 \leq n - 1$, and $\gamma(-k) = \gamma(k)$, we know that in the final sum of 1, the term in $\gamma(k)$ for $k \in \{0, ..., n-1\}$ will not appear more than n times. So we have the following bound :

$$\mathbb{E}\left[(\bar{Y}_n - \mu)^2\right] \leq \frac{1}{n}\sum_{k=0}^{n-1}\gamma(k)$$

  Therefore, as $\sum_k |\gamma(k)| < +\infty$, we have,

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] \leq \frac{1}{n}\sum_{k=0}^{n}|\gamma(k)| \xrightarrow[n\to 0]{} 0.$$

  Thus, $\bar{Y}_n$ converges to μ in L^2 so $\boxed{\bar{Y}_n \text{ is consistent.}}$

Finally, we have

$$\mathbb{E}[\bar{Y}_n - \mu]^2 \leq \mathbb{E}[(\bar{Y}_n - \mu)^2] \leq \frac{1}{n} \sum_{k=0}^{n} |\gamma(k)|$$

So,

$$\mathbb{E}[\bar{Y}_n - \mu] \leq \frac{1}{\sqrt{n}} \left(\sum_{k=0}^{n} |\gamma(k)| \right)^{1/2}.$$

$\boxed{\text{We then get that the convergence rate is } 1/\sqrt{n}\text{, the same as in the i.i.d. case.}}$

3 AR and MA processes

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

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

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

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 (2).

Answer 2

- We first have :

$$\mathbb{E}[Y_t] = \sum_{k=0}^{\infty} \psi_k \underbrace{\mathbb{E}[\varepsilon_{t-k}]}_{=0} = 0$$

  Then we have :

$$\mathbb{E}[Y_t Y_{t-k}] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \underbrace{\mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-k-j}]}_{=\sigma_\varepsilon^2 \delta_{t-i,t-k-j}} = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_l \psi_{l+|k|} < \infty$$

  As the expected value is 0 for all t and the covariance depends only on k, then this process is weakly stationary.

- We have :

$$\sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2 = \sigma_\varepsilon^2 (\sum_{l=0}^{\infty} \psi_l e^{-2\pi l i f})(\sum_{n=0}^{\infty} \psi_n e^{2\pi n i f}) = \sum_{k=-\infty}^{\infty} (\sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_l \psi_{l+|k|}) e^{-2\pi k i f}$$

  So finally :

$$\sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2 = \sum_{k=-\infty}^{\infty} \underbrace{\mathbb{E}[Y_t Y_{t-k}]}_{=\gamma(k)} e^{-2\pi k i f} = S(f)$$

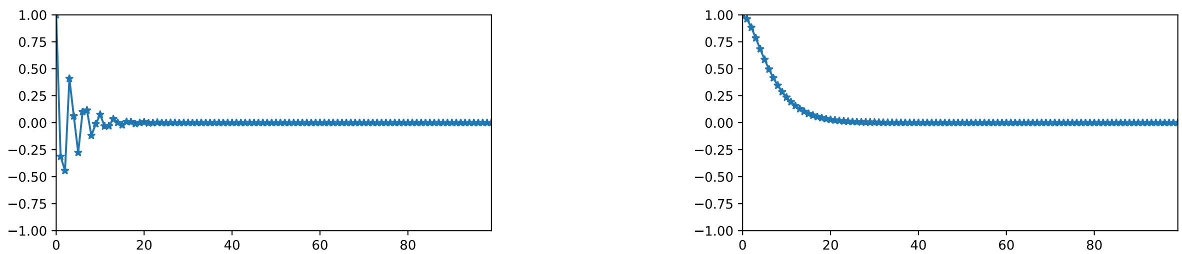
Question 3 *AR(2) process*

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

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{3}$$

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 take $G_1 = \frac{1}{r_1}, G_2 = \frac{1}{r_2}$
  Y_t can be written as a sum of the previous ϵ_t, then $\forall (t,k) > 0, \epsilon_{t+k}$ and Y_t are independent.
  Then, we obtain $\forall k, \gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2)$
  As $\gamma(k) = \gamma(-k)$ and $\phi_1 = G_1 + G_2$ and $\phi_2 = -G_1 G_2$ and $G_1 \neq G_2$:
  we have $\gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0) = \frac{(1-G_2)G_1^2 - (1-G_1)G_2^2}{(G_1-G_2)(1+G_1G_2)}\gamma(0)$
  And $\gamma(0) = \frac{(1-G_2)G_1 - (1-G_1)G_2}{(G_1-G_2)(1+G_1G_2)}\gamma(0)$
  So let's show by recurrence that

  $$\forall k > 0, \gamma(k) = \frac{(1-G_2)G_1^{k+1} - (1-G_1)G_2^{k+1}}{(G_1-G_2)(1+G_1G_2)}\gamma(0)$$

  We take k such as the propriety is true for k-1 and k-2. Then :

  $$\frac{\gamma(k)}{\gamma(0)} = \phi_1 \frac{(1-G_2)G_1^k - (1-G_1)G_2^k}{(G_1-G_2)(1+G_1G_2)} + \phi_2 \frac{(1-G_2)G_1^{k-1} - (1-G_1)G_2^{k-1}}{(G_1-G_2)(1+G_1G_2)}$$

But $\phi_1 = G_1 + G_2$ and $\phi_2 = -G_1G_2$. So :

$$\frac{\gamma(k)}{\gamma(0)} = \frac{(1-G_2)G_1^{k+1} + (1-G_2)G_1^kG_2 - (1-G_1)G_2^kG_1 - (1-G_1)G_2^{k+1}}{(G_1-G_2)(1+G_1G_2)} + \frac{-(1-G_2)G_1^kG_2 + (1-G_1)G_2^kG_1}{(G_1-G_2)(1+G_1G_2)}$$

i.e

$$\boxed{\frac{\gamma(k)}{\gamma(0)} = \frac{(1-G_2)G_1^{k+1} - (1-G_1)G_2^{k+1}}{(G_1-G_2)(1+G_1G_2)}}$$

which gives the result.

- According to the previous result, we can notice that in the case of real roots, then $\gamma(k)$ is monotone so it is the second case. In the first case the roots are then complex.
- Let us introduce the lag operator denote L defined as follows :

$$\forall t \geq 1,\ LY_t = Y_{t-1}.$$

We can then rewrite the random process $\{Y_t\}_{t\geq 0}$ as follows :

$$Y_t - \phi_1 L(Y_t) - \phi_2 L^2(Y_t) = \varepsilon_t \text{ i.e. } \phi(L)Y_t = \varepsilon_t$$

Using previous notations, we have that

$$\phi(z) = (z-r_1)(z-r_2) = r_1r_2(z/r_1 - 1)(z/r_2 - 1) = r_1r_2(1-G_1z)(1-G_2z)$$

so we have

$$Y_t = \frac{1}{\phi(L)}\varepsilon_t = \frac{1}{r_1r_2(1-G_1L)(1-G_2L)}\varepsilon_t$$

As $|r_1| > 1$ and $|r_2| > 1$, we have $|G_1| < 1$ and $|G_2| < 1$ so we can use the Taylor expansion of $x \mapsto \frac{1}{1-x}$:

$$\frac{1}{1-G_1L} = \sum_{n=0}^{\infty} G_1^n L^n, \qquad \frac{1}{1-G_2L} = \sum_{n=0}^{\infty} G_2^n L^n$$

Then, we have

$$Y_t = \frac{1}{r_1r_2}\left(\sum_{n=0}^{\infty} G_1^n L^n\right)\left(\sum_{n=0}^{\infty} G_2^n L^n\right)\varepsilon_t = \left(\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} G_1^{k+1}G_2^{n-k-1}\right)L^n\right)\varepsilon_t$$

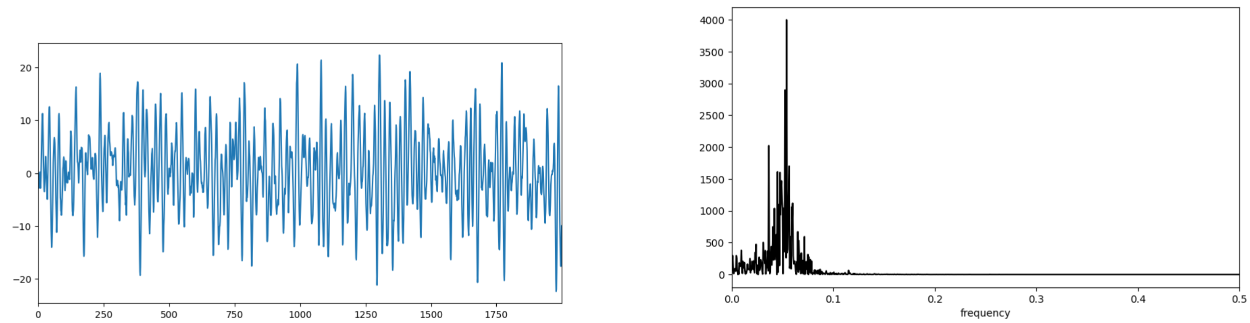
So, Y_t can be seen as a MA(∞) process, and as $Y_t = \frac{1}{\phi(L)}\varepsilon_t$, we can use question 2 and we get :

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

- $r1 = 0.575\sqrt{3} + i0.575$, $r2 = 0.575\sqrt{3} - i0.575$ verify the conditions. This leads to

$$\phi_2 = \frac{-1}{(1.05)^2}, \phi_1 = \frac{2 * \sqrt{3} * 0.575}{(1.05)^2}$$

Then we obtain :



Signal

Periodogram

Figure 2: AR(2) process

We notice a large spike at around 0.05 Hz. If we zoom on the signal, we can indeed see that the signal is really similar every 20 points which correspond to the frequency of 0.05 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 ($k = 0, \dots, L-1$) by

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{4}$$

where w_L is a modulating window given by

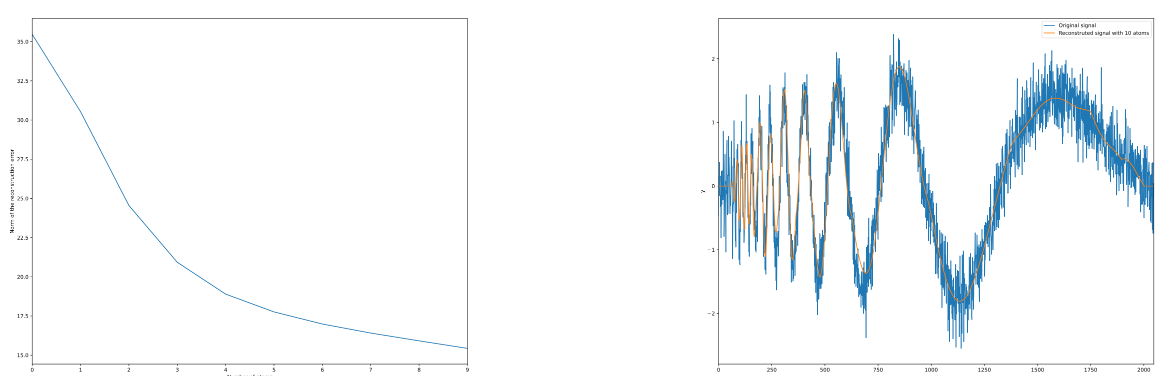
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{5}$$

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