Assignment 2 (ML for TS) - MVA 2023/2024

Bastien LE CHENADEC bastien.le-chenadec@eleves.enpc.fr
Sofiane EZZEHI sofiane.ezzehi@eleves.enpc.fr

December 5, 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 Tuesday 5th December 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: docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2cBHQM

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

Given $(X_i)_{i\in\mathbb{N}}$ i.i.d. random variables with finite variance σ^2 and mean μ, the sample mean, for $n \in \mathbb{N}$, is

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

To compute the rate of convergence, we use the Chebyshev inequality:

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2}$$

Since the random variables are i.i.d., we have $Var(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$. Therefore, we have:

$$\boxed{\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\varepsilon^2}}$$

Let's now show that $\bar{Y}_n$ is consistent and enjoys the same rate of convergence as the i.i.d. case. Using Chebyshev's inequality, we have,

$$\mathbb{P}(|\bar{Y}_n - \mu| \geq \varepsilon) \leq \frac{Var(\bar{Y}_n)}{\varepsilon^2} = \frac{\mathbb{E}[(\bar{Y}_n - \mu)^2]}{\varepsilon^2}$$

We now bound $\mathbb{E}[(\bar{Y}_n - \mu)^2]$:

$$\begin{aligned}
\mathbb{E}[(\bar{Y}_n - \mu)^2] &= \mathbb{E}[\bar{Y}_n^2] + \mu^2 - 2\mu\mathbb{E}[\bar{Y}_n] \\
&= \mathbb{E}[\bar{Y}_n^2] + \mu^2 - 2\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i] \\
&= \mathbb{E}[\bar{Y}_n^2] + \mu^2 - 2\mu^2 \\
&= \mathbb{E}[\bar{Y}_n^2] - \mu^2 \\
&\leq \mathbb{E}[\bar{Y}_n^2]
\end{aligned}$$

We now bound $\mathbb{E}[\bar{Y}_n^2]$,

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

We introduce the sets $A_k = \{(i,j) \in [\![1,n]\!]^2 \mid j-i=k\}$, for $k \in [\![1-n, n-1]\!]$. We easily show that

$Card(A_k) = n - |k|$ by enumerating, for each k, the elements of $\{(i, i+k) \mid i \in [\![1,n]\!], i+k \in [\![1,n]\!]\}$. Therefore, we can write,

$$\begin{aligned}\mathbb{E}[|\bar{Y}_n^2] &= \frac{1}{n^2}\sum_{k=1-n}^{n-1} (n-|k|)\gamma(k) \\ &\leq \frac{1}{n^2}\sum_{k=1-n}^{n-1} (n-|k|)|\gamma(k)| \\ &\leq \frac{1}{n^2}\sum_{k=1-n}^{n-1} n|\gamma(k)| \\ &= \frac{1}{n}\sum_{k\in\mathbb{Z}} |\gamma(k)|\end{aligned}$$

Thus, we obtain the following bound on $\mathbb{P}(|\bar{Y}_n - \mu| \geq \varepsilon)$,

$$\boxed{\mathbb{P}(|\bar{Y}_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{E}[\bar{Y}_n^2]}{\varepsilon^2} \leq \frac{1}{n\varepsilon^2}\sum_{k\in\mathbb{Z}} |\gamma(k)|}$$

We have therefore shown that $\bar{Y}_n$ is consistent 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 $\mathbb{E}(Y_t Y_{t-k})$. 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

Let's first compute $\mathbb{E}[Y_t]$. We introduce the random variables $S_n(t) = \sum_{k=0}^n \psi_k \varepsilon_{t-k}$. We have,

$$\begin{aligned}
\mathbb{E}[S_n(t)] &= \mathbb{E}[\sum_{k=0}^n \psi_k \varepsilon_{t-k}] \\
&= \sum_{k=0}^n \psi_k \mathbb{E}[\varepsilon_{t-k}] \\
&= 0 \qquad (\text{ since } \mathbb{E}[\varepsilon_{t-k}] = 0)
\end{aligned}$$

Therefore, $\mathbb{E}[S_n(t)]$ is a stationary series of random variables, all equal to 0. Therefore, $\mathbb{E}[Y_t] = 0$.

Let's now compute $\mathbb{E}[Y_t Y_{t-k}]$.

$$\begin{aligned}
\mathbb{E}[S_n(t)S_n(t-k)] &= \mathbb{E}[\sum_{i=0}^n \sum_{j=0}^n \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t-k-j}] \\
&= \sum_{i=0}^n \sum_{j=0}^n \psi_i \psi_j \mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-k-j}]
\end{aligned}$$

We now distinguish two cases:

- If $i \neq k + j$, then $\mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-k-j}] = 0$ since ε_{t-i} and ε_{t-k-j} are independent.
- If $i = k + j$, then $\mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-k-j}] = \mathbb{E}[\varepsilon_{t-i}^2] = \sigma_\varepsilon^2$.

Therefore, we have,

$$\begin{aligned}
\mathbb{E}[S_n(t)S_n(t-k)] &= \sum_{i=0}^n \sum_{j=0}^n \psi_i \psi_j \mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-k-j}] \\
&= \sigma_\varepsilon^2 \sum_{i=0}^n \psi_i \psi_{k+i}
\end{aligned}$$

Since $\sum_k \psi_k^2 < \infty$, this sum converges to $\sigma_\varepsilon^2 \sum_k \psi_k^2$ when $n \to \infty$.

Therefore, we have,

$$\boxed{\mathbb{E}[Y_t Y_{t-k}] = \sigma_\varepsilon^2 \sum_{i=0}^\infty \psi_i \psi_{k+i}}$$

We can name this term $\gamma(k)$.

We have thus proven that $\{Y_t\}_t$ is a weakly stationary process.

Let's now compute the power spectrum of $\{Y_t\}_t$. We recall the definition of the power spectrum:

$$S(f) = \sum_{k\in\mathbb{Z}} \gamma(k) e^{\frac{-2i\pi k f}{f_s}}$$

where f_s is the sampling frequency (here, $f_s = 1$).

We have,

$$\begin{aligned}
S(f) &= \sum_{k\in\mathbb{Z}} \gamma(k) e^{-2i\pi k f} \\
&= \sigma_\varepsilon^2 \sum_{k=-\infty}^{+\infty} \sum_{i=0}^{+\infty} \psi_i \psi_{k+i} e^{-2i\pi k f} \\
&= \sigma_\varepsilon^2 \sum_{j=-\infty}^{+\infty} \sum_{i=0}^{+\infty} \psi_i \psi_j e^{-2i\pi (j-i) f}
\end{aligned}$$

Since the ψ_i are defined for $i \geq 0$, we consider the terms where $i \leq 0$ as equal to 0.

We can therefore write,

$$\begin{aligned}
S(f) &= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} \psi_i \psi_j e^{-2i\pi (j-i) f} \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \psi_j e^{-2i\pi j f} \sum_{i=0}^{+\infty} \psi_i e^{2i\pi i f} \\
&= \sigma_\varepsilon^2 |\phi(e^{-2i\pi f})|^2
\end{aligned}$$

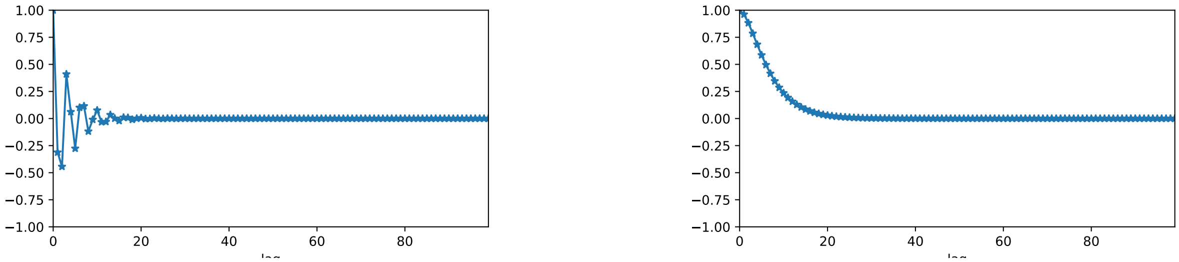
Question 3 *AR(2) process*

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

We multiply both sides of (2) by $Y_t, Y_{t-1}, \dots, Y_{t-k}$ and take the expectation. We obtain the $k+1$ equalities :

$$\begin{aligned}\gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \mathbb{E}[Y_t\varepsilon_t] \\ \gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(1) + \mathbb{E}[Y_{t-1}\varepsilon_t] \\ &\vdots \\ \gamma(k) &= \phi_1\gamma(k-1) + \phi_2\gamma(k-2) + \mathbb{E}[Y_{t-k}\varepsilon_t]\end{aligned}$$

By multiplying both sides of (2) by ε_t and taking the expectation we have that $\mathbb{E}[Y_{t-k}\varepsilon_t] = \sigma^2\delta(k)$. We get a linear recurring sequence of the second order. Let's start by solving for the initial conditions $\gamma(0)$ and $\gamma(1)$. We denote $V = \gamma(0)$ the variance of the AR(2) process. Using the first three equalities we have :

$$\begin{cases} V = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2 \\ \gamma(1) = \frac{\phi_1 V}{1-\phi_2} \\ \gamma(2) = V\left(\frac{\phi_1^2}{1-\phi_2} + \phi_2\right) \end{cases}$$

Solving for V and after simplification we have :

$$V = \frac{\sigma^2(1-\phi_2)}{(1+\phi_2)(1+\phi_1-\phi_2)(1-\phi_1-\phi_2)}$$

This does not depend on the rest of the sequence, so we are done for the initial conditions. Let's suppose that $r_1 \neq r_2$ (which is a more complicated case than $r_1 = r_2$[1]). We know that the linear recurring sequence of the second order $\gamma(k) = \phi_1\gamma(k-1) + \phi(2)\gamma(k-2)$ with characteristic equation $z^2 - \phi_1 z - \phi_2 = 0 \Leftrightarrow \phi(1/z) = 0$ has a solution of the form :

$$\gamma(k) = \alpha r_1^{-k} + \beta r_2^{-k}$$

We can find α and β by solving the system :

$$\begin{cases} \gamma(0) = \alpha + \beta = V \\ \gamma(1) = \alpha r_1^{-1} + \beta r_2^{-1} = \frac{\phi_1 V}{1-\phi_2} \end{cases}$$

[1] In the case $r_1 = r_2$, we proceed similarly with a solution of the form $\gamma(k) = \alpha r^{-k} + \beta n r^{-k}$.

Remember the relations between the roots and the coefficients of a polynomial of degree 2, in our case $r_1 + r_2 = -\phi_1/\phi_2$ and $r_1 r_2 = -1/\phi_2$, which allows us to write $\frac{\phi_1}{1-\phi_2} = \frac{r_1+r_2}{1+r_1 r_2}$. We can also express V in terms of r_1 and r_2. We can now solve the system. Skipping the (long) calculations, we find :

$$\begin{cases} \alpha = \frac{\sigma^2 r_1^3 r_2^2}{(r_1 r_2 - 1)(r_2 - r_1)(r_1^2 - 1)} \\ \beta = -\frac{\sigma^2 r_1^2 r_2^3}{(r_1 r_2 - 1)(r_2 - r_1)(r_2^2 - 1)} \end{cases}$$

Thus we have :

$$\gamma(k) = \frac{\sigma^2 r_1^2 r_2^2}{(r_1 r_2 - 1)(r_2 - r_1)}[(r_1^2 - 1)^{-1} r_1^{1-k} - (r_2^2 - 1)^{-1} r_2^{1-k}] \quad (3)$$

This result allows us to say that on the figure (1), the first AR(2) process has complex roots and the second one has real roots. Indeed, the first one has a sinusoidal shape, which is explained by the complex exponential $\exp(i\theta) = \cos(\theta) + i\sin(\theta)$, and the second one has a decaying exponential shape, which is explained by the real exponential.

We recall the power spectrum:

$$S(f) = \sum_{k \in \mathbb{Z}} \gamma(k) e^{\frac{-2i\pi k f}{f_s}}$$

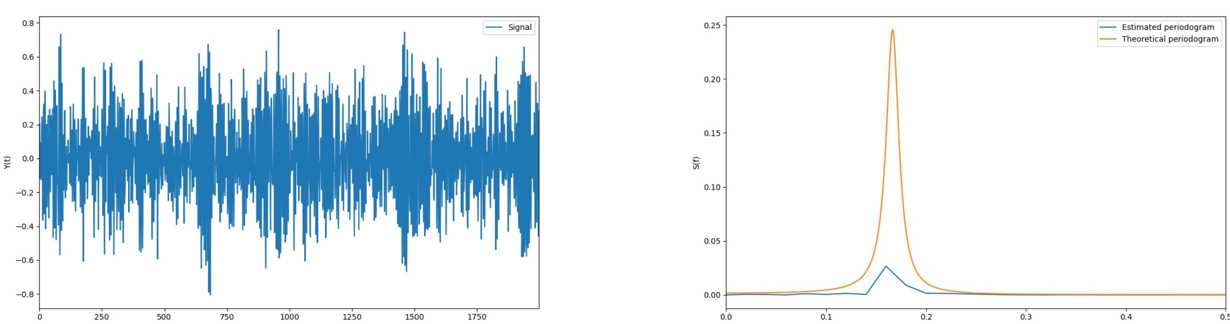
where f_s is the sampling frequency (here, $f_s = 1$).

For an AR(2) process, it can be shown that the power spectrum is given by :

$$S(f) = \frac{\sigma^2}{2\pi |\phi(\exp(-2i\pi f/f_s))|^2}$$

which can be developed to :

$$S(f) = \frac{\sigma^2}{2\pi(1 + \phi_1^2 + 2\phi_2 + \phi_2^2 + 2(\phi_1\phi_2 - \phi_1)\cos(2\pi f/f_s) - 4\phi_2 \cos^2(2\pi f/f_s))}$$



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

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

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

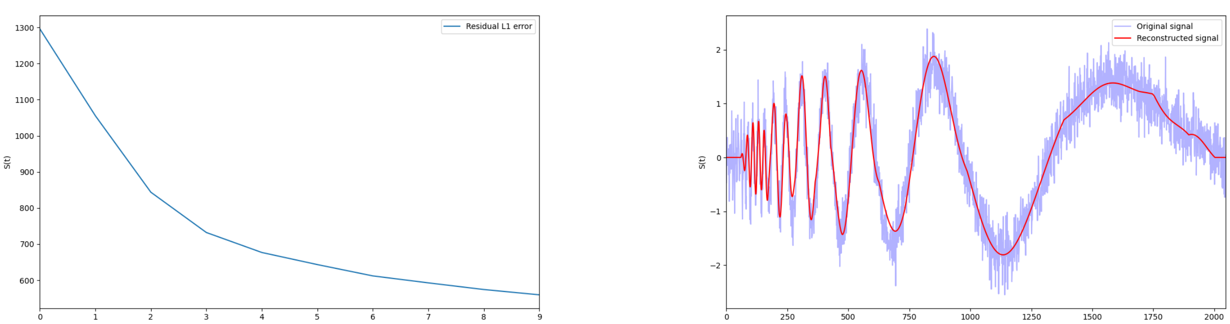
How to project orthogonally the signal ? We denote $P \in \mathbb{R}^{n,K}$ the matrix whose columns are the atoms selected by the algorithm so far, and $S \in \mathbb{R}^n$ the signal. The orthogonal projection is given by :

$$\arg\min_X \|PX - S\|_2^2$$

Taking the gradient with respect to X and setting it to zero yields $P^T(PX - S) = 0$ for which we have the solution :

$$X = (P^T P)^{-1} P^T S$$

If we run into issues where $P^T P$ is not invertible, we can always use the pseudo-inverse. This will be useful to compute the residual at each iteration of the algorithm.



Norms of the successive residuals　　Reconstruction with 10 atoms

Figure 3: Question 4