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

Boubacar Sow boubacar.sow@ens-paris-saclay.fr
Theïlo Terrisse theilo.terrisse@eleves.enpc.fr

November 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 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

3 AR and MA processes

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

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

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

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

Deriving $\mathbb{E}(Y_t)$ and $\mathbb{E}(Y_t Y_{t-k})$:

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

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

The mean function, $\mathbb{E}$ is constant and does not depent on time t. The covariance between Y_t and Y_{t+k} depends only on the lag k and not on time t. We conclude that the process is 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 $\phi(z) = \sum_j \psi_j z^j$

$$
\begin{aligned}
\mathbb{E}(Y_t Y_{t-\tau}) &= \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-\tau-j}\right] \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}[\varepsilon_{t-i} \varepsilon_{t-\tau-j}] \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_\varepsilon(i-j-\tau)
\end{aligned}
$$

Then, we can write the power spectrum as follows:

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

Now, let's make a change of variables. Let $j = i - \tau$. Then, the equation becomes:

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

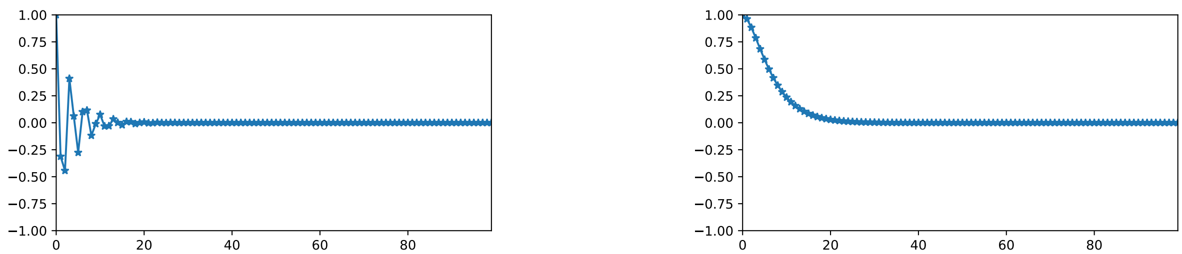
Question 3 *AR(2) process*

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

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

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

Express the autocovariance coefficients $\gamma(\tau)$ using the roots r_1 and r_2

$$\gamma(\tau) = E[Y_t Y_{t-\tau}]$$

$$Y_{t-\tau} = \phi_1 Y_{t-\tau+1} + \phi_2 Y_{t-\tau+2} + \varepsilon_{t-\tau}$$

$$\begin{aligned}\gamma(\tau) &= E[Y_t(\phi_1 Y_{t-\tau+1} + \phi_2 Y_{t-\tau+2} + \varepsilon_{t-\tau})] \\ &= \phi_1 E[Y_t Y_{t-\tau+1}] + \phi_2 E[Y_t Y_{t-\tau+2}] + E[Y_t \varepsilon_{t-\tau}] \\ \gamma(\tau) &= \begin{cases} \phi_1 \gamma(\tau-1) + \phi_2 \gamma(\tau-2), & \text{if } \tau \neq 0 \\ \phi_1 \gamma(\tau-1) + \phi_2 \gamma(\tau-2) + \sigma_\varepsilon^2, & \text{if } \tau = 0 \end{cases}\end{aligned}$$

Here, σ^2 is the variance of the error term ε_t.

The characteristic polynomial of this equation is: $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$. This characteristic polynomial is the reciprocal of the polynomial $z^2 - \phi_1 z - \phi_2$. If the roots of this polynomial are r_1 and r_2, then roots of our characteristic polynomial are $1/r_1$ and $1/r_2$

The final solution for the autocovariance function is thus:

$$\gamma(\tau) = c_1 \left(\frac{1}{r_1}\right)^\tau + c_2 \left(\frac{1}{r_2}\right)^\tau + \sigma_\varepsilon^2 \delta_0(\tau)$$

A more detailed solution can be found here, at the section 1.3

When $r_1, r_2 \in \mathbb{C}$, ie. $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$, where r is the magnitude of the roots and θ is the phase.

$$\gamma(\tau) = E[Y_t Y_{t-\tau}] = \frac{1}{r^\tau}(c_1 \cos(\tau\theta) + c_2 \sin(\tau\theta))$$

This equation represents a damped oscillatory behavior, which is characteristic of AR processes with complex roots.

The correlogram on the left corresponds to an AR(2) process with complex roots, while the correlogram on the right corresponds to an AR(2) process with real roots.

Express the power spectrum $S(f)$ (assume the sampling frequency is 1 Hz) using $\phi(\cdot)$.

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

We know that:

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

Applying the lag operator L to the AR(2) process gives:

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

This equation can be rewritten as:

$$Y_t = \frac{\varepsilon_t}{\phi(L)} = \frac{\varepsilon_t}{1 - \frac{1}{r_1} L - \frac{1}{r_2} L^2} = \frac{\varepsilon_t}{(1 - \frac{1}{r_1} L)(1 - \frac{1}{r_2} L)}$$

Each term in the product $(1 - \frac{1}{r_1} z)^{-1} (1 - \frac{1}{r_2} z)^{-1}$ can be expanded as a power series in z:

$$\begin{aligned}
(1 - \frac{1}{r_1} z)^{-1} &= 1 + \frac{1}{r_1} z + \frac{1}{r_1^2} z^2 + \frac{1}{r_1^3} z^3 + \cdots = \sum_{i=0}^{\infty} \frac{1}{r_1^i} z^i, \\
(1 - \frac{1}{r_2} z)^{-1} &= 1 + \frac{1}{r_2} z + \frac{1}{r_2^2} z^2 + \frac{1}{r_2^3} z^3 + \cdots = \sum_{j=0}^{\infty} \frac{1}{r_2^j} z^j.
\end{aligned}$$

The product of these two power series gives the power series expansion of $\psi^{-1}(z)$

$$\begin{aligned}
\phi^{-1}(z) &= \left(\sum_{i=0}^{\infty} \frac{1}{r_1^i} z^i\right) \left(\sum_{j=0}^{\infty} \frac{1}{r_2^j} z^j\right) \\
&= \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} \frac{1}{r_1^i r_2^{k-i}}\right) z^k.
\end{aligned}$$

This shows that the AR(2) process can be represented as an MA(∞) process with MA coefficients given by the terms in the power series expansion of $\psi^{-1}(z)$

$$Y_t = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} \frac{1}{r_1^i r_2^{k-i}} \right) \varepsilon_{t-k}.$$

$$S(f)_{\mathrm{AR}(2)} = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^{-2} = \frac{\sigma_\varepsilon^2}{|\phi(e^{-2\pi i f})|^2}.$$

This equation gives the power spectrum of the AR(2) process, which is inversely proportional to the square of the magnitude of the Fourier transform of the MA coefficients.

Choose ϕ_1 and ϕ_2 such that the characteristic polynomial has two complex conjugate of norm $r = 1.05$ and phase $\theta = 2\pi/6$

$$1 - \phi_1 z - \phi_2 z^2 = (1 - \frac{1}{r_1} z)(1 - \frac{1}{r_2} z)$$

Expanding the right-hand side, we get:

$$1 - \phi_1 z - \phi_2 z^2 = 1 - (\frac{1}{r_1} + \frac{1}{r_2})z + \frac{1}{r_1 r_2} z^2$$

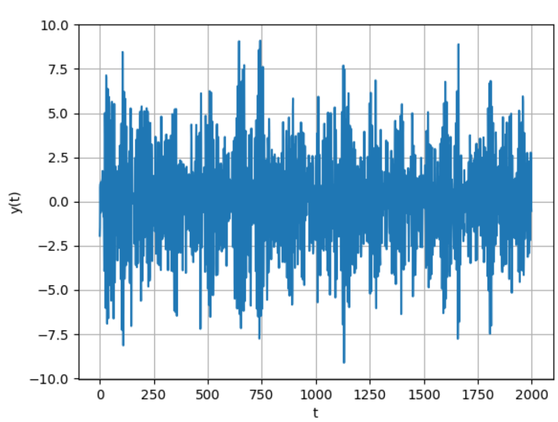
Comparing coefficients, we can see that $\phi_1 = (\frac{1}{r_1} + \frac{1}{r_2})$ and $\phi_2 = \frac{-1}{r_1 r_2}$.

Given that $r_1 = r\exp(i\theta)$ and $r_2 = r\exp(-i\theta)$, we have:

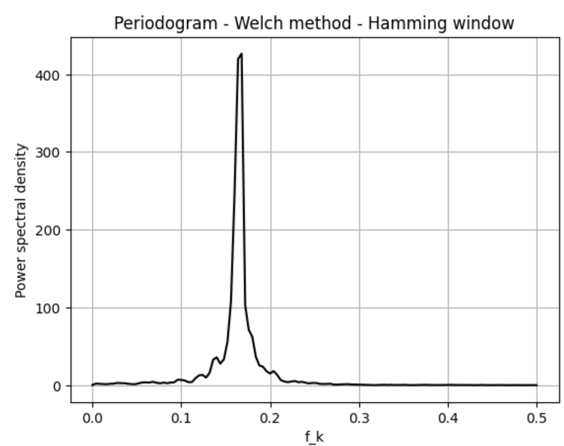
$$\phi_1 = \frac{2cos(\theta)}{r}$$

$$\phi_2 = -\frac{1}{r^2}$$

.



Signal



Periodogram

Figure 2: AR(2) process

A peak is observed at a frequency of $f = 0.17$. This peak signifies a strong periodic component in the time series at this frequency. The frequency corresponds to the phase of the roots of the AR(2) process, which was set to $2\pi/6$. This frequency is the reciprocal of the period, hence the peak at $f = 0.17$.

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

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

where w_L is a modulating window given by

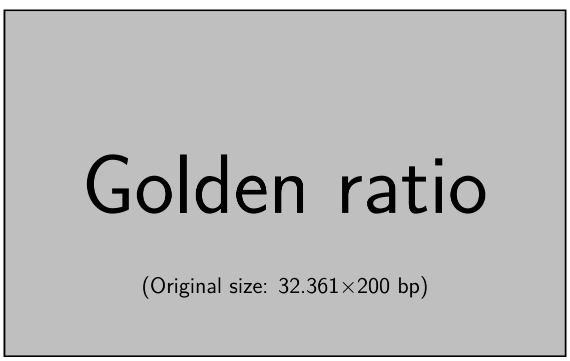
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right].$$

Question 4 *Sparse coding with OMP*

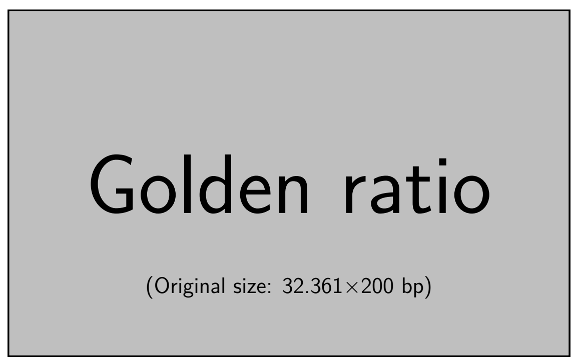
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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