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

Farah Ellouze farah.ellouze@ens-paris-saclay.fr
Halvard Bariller halvard.bariller@ens-paris-saclay.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 5$^{\text{th}}$ 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

- We consider a sequence $(\theta_i)_i$ of random variables i.i.d., with mean μ and variance $\sigma^2 < \infty$. We define the sample mean as $\bar{\theta}_n = \frac{1}{n}\sum_{i=1}^n \theta_i$. According to the Bienaymé-Tchebychev inequality, we have:

$$\mathbb{P}(|\bar{\theta}_n - \mathbb{E}[\bar{\theta}_n]| > \epsilon) \leq \frac{\mathbb{V}(\bar{\theta}_n)}{\epsilon^2} \iff \mathbb{P}(|\bar{\theta}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0.$$

  We therefore have a convergence in $\mathcal{O}(\frac{1}{n})$.

- We consider a wide-sense stationary process $\{Y_t\}_{t\geq 1}$ with mean μ and with absolutely summable autocovariances, i.e. $\sum_k |\gamma(k)| < \infty$. We proceed to show the convergence in norm L^2:

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

If a sequence of random variables converges in L^2 towards a limit X, it converges in probability as well towards the same limit. We thus have the convergence in probability of $\bar{Y}_n$ towards μ, and therefore the consistency of the estimator.

We can use the Bienaymé-Tchebychev inequality to exhibit a convergence in $\mathcal{O}\left(\frac{1}{n}\right)$:

$$\begin{aligned}
\mathbb{P}(|\bar{Y}_n - \mu| > \epsilon) &\leq \frac{\mathbb{V}(\bar{Y}_n)}{\epsilon^2} = \frac{\mathbb{E}[(\bar{Y}_n - \mu)^2]}{\epsilon^2} \\
&\leq \frac{2}{n\epsilon^2} \left(\sum_{j=0}^{n-1} |\gamma(j)| \right).
\end{aligned}$$

We thus have shown that the wide-sense stationary process $\{Y_t\}_t$ possesses the same properties in terms of consistency and convergence rate than a sequence of i.i.d. random variables.

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

a)

- Using the fact that $\{\epsilon_t\}_t$ is a zero mean white noise, we have:

$$\mathbb{E}(Y_t) = \mathbb{E}(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}) = \sum_{k=0}^{\infty} \psi_k \mathbb{E}(\varepsilon_{t-k}) = 0$$

- By independence of the $\{\epsilon_t\}_t$, we have:

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

We can express the last term as a function of k alone: the autocovariance does not depend of the order t.

The mean and autocovariance being independent of the time, we thus have shown that the process is weakly stationary.

b) We start from $\sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2$:

$$\begin{aligned}
\sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2 &= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \psi_j e^{-2i\pi j f} \sum_{l=0}^{+\infty} \psi_j e^{2i\pi l f} \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} \psi_j \psi_l e^{-2\pi i (j-l) f} \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \sum_{\tau=-\infty}^{j} \psi_j \psi_{j-\tau} e^{-2\pi i \tau f} \text{ (with } \tau = j - l) \\
&= \sigma_\varepsilon^2 \sum_{\tau=-\infty}^{+\infty} \sum_{j=0}^{+\infty} \psi_j \psi_{j+\tau} e^{-2\pi i \tau f} \quad \text{(c.f. HW1 Ex.3 Q.5)} \\
&= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i \tau f} \\
&= S(f) \text{ with } f_s = 1
\end{aligned}$$

Thus

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

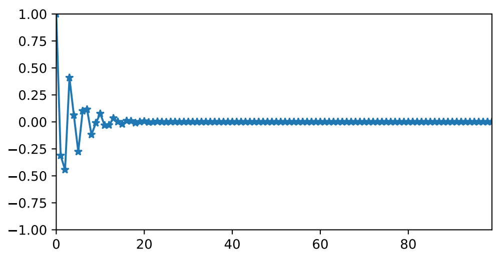
Question 3 *AR(2) process*

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

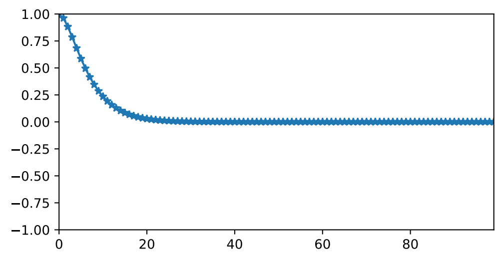
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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

a)

- We start by expressing $\gamma(\tau)$ as a linear recursive sequence of order 2 :

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

The characteristic polynomial of this sequence writes as : $r^2 - \phi_1 r - \phi_2 = 0$
Which is equivalent for $r \neq 0$ to : $1 - \phi_1 \frac{1}{r} - \phi_2 \frac{1}{r^2} = 0$
And, its roots are $\frac{1}{r_1}$ and $\frac{1}{r_2}$, $r_1 \neq 0, r_2 \neq 0$
We can then distinguish two cases depending on r_1, r_2 :
- If $r_1, r_2 \in \mathbb{R}$: there exists $\lambda, \mu \in \mathbb{R}$ such that

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

- If $r_1, r_2 \in \mathbb{C} : r_1 = re^{i\alpha}$ and $r_2 = re^{-i\alpha}$ with $r \in \mathbb{R}_+^*, \alpha \in \mathbb{R}$
and there exists $\lambda, \mu \in \mathbb{R}$ such that

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

b)

- In the first figure, the curve has an oscillatory behaviour before reaching 0 thus the first AR(2) process has complex roots.
  In the second figure, the autocorrelations exhibit a more straightforward pattern without oscillations and thus the curve belong to the AR(2) process with real roots.

c)

- Let L be the backshift operator, we express Y_t as follows :

$$Y_t = \phi_1 L Y_t + \phi_2 L^2 Y_t + \varepsilon_t$$

Thus

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

We can recognize the characteristic polynomial $\phi(z)$ with roots r_1 and r_2 such that $|r_i| > 1$.
$\phi(z)$ decomposes as $\phi(z) = (1 - z_1 z)(1 - z_2 z)$ where z_i are the inverse of the roots.
Thus,

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

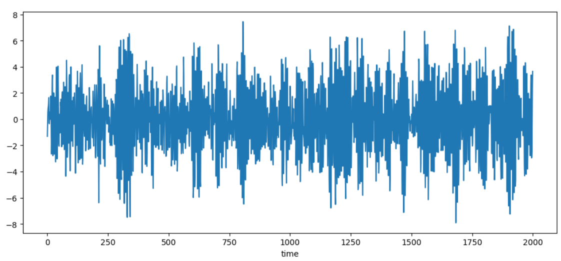
We recognize two geometric series $\sum_{i=0}^{+\infty} (\frac{L}{r_1})^i$ and $\sum_{i=0}^{+\infty} (\frac{L}{r_2})^i$.
We can rewrite Y_t as :

$$\begin{aligned}
Y_t &= \sum_{i=0}^{+\infty} (\frac{L}{r_1})^i \sum_{j=0}^{+\infty} (\frac{L}{r_2})^j \varepsilon_t \\
&= \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} (\frac{L^{i+j}}{r_1^i r_2^j}) \varepsilon_t \\
&= \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \frac{1}{r_1^i r_2^j} \varepsilon_{t-i-j} \text{(L is a backshift operator)} \\
&= \sum_{k=0}^{+\infty} \sum_{i+j=k} \frac{1}{r_1^i r_2^j} \varepsilon_{t-k} \\
&= \sum_{k=0}^{+\infty} \psi_k \varepsilon_{t-k}
\end{aligned}$$

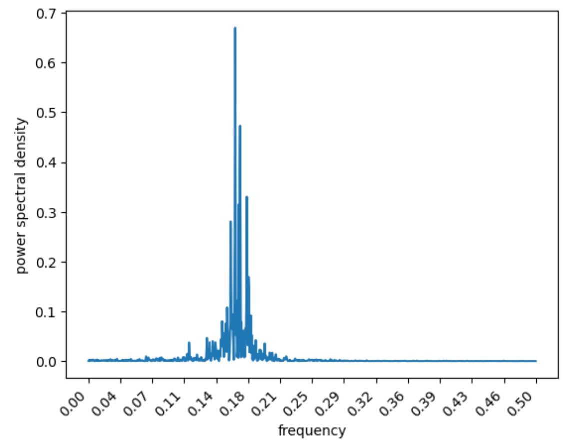
Y_t is hence written as an $M(\infty)$ process and the power spectrum $S(f)$ can be expressed using Question 2 as follows :

$$S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2 \text{ where } \phi(z) = \sum_j \psi_j z^j \text{ and } \psi_k = \sum_{i+j=k} \frac{1}{r_1^i r_2^j}$$

d)



Signal



Periodogram

Figure 2: AR(2) process

- The frequency of the peak corresponds to the frequency of the complex roots i.e $2\pi f = \frac{2\pi}{6}$ thus $f = \frac{1}{6}Hz = 0.1666Hz$. This is confirmed on the figure as the periodogram shows a peak around the frequency 0.16Hz.

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

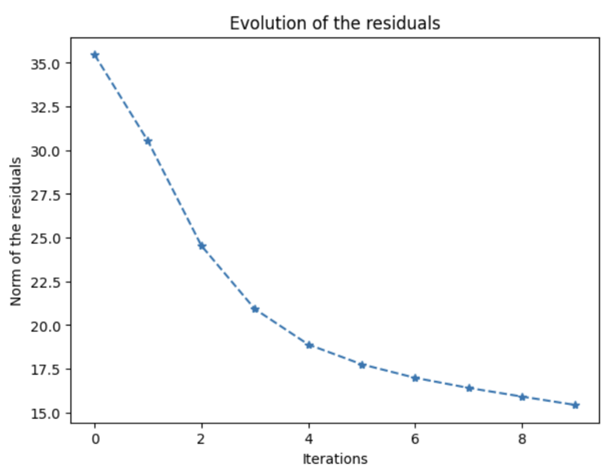


Figure 3: Norms of the successive residuals

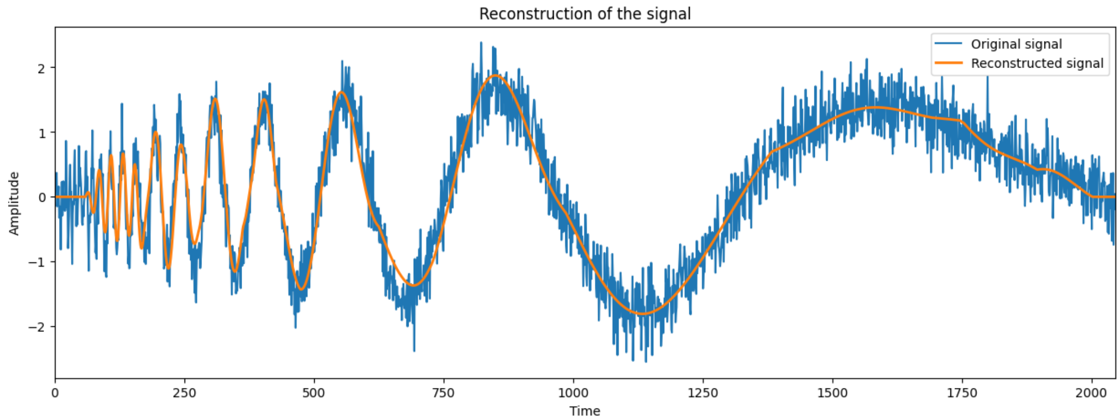


Figure 4: Reconstruction with 10 atoms