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

Benjamin Lapostolle benjamin.lapostolle@polytechnique.edu

December 15, 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

For the convergence rate of the sample mean of i.i.d. random variables with finite variance, we have to use the Bienaymé-Tchebychev inequality.

Let's call: $\forall n \geq 1, Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

We have:

- $\forall n \geq 1, E(Y_n) = E(X)$
- $\forall n \geq 1, V(Y_n) = \frac{V(X)}{n}$

As such, and according to the Bienaymé-Tchebychev inequality:

$$\begin{aligned} \forall n \geq 1, \mathbf{P}(|Y_n - E(Y_n)| \geq \alpha) &\leq \frac{V(Y_n)}{\alpha^2} \\ \forall n \geq 1, \mathbf{P}(|Y_n - E(X)| \geq \alpha) &\leq \frac{V(X)}{n\alpha^2} \end{aligned} \tag{1}$$

As such, its convergence rate is in $\mathcal{O}(\frac{1}{n})$.

For the non-i.i.d. case, we will use the same process.

Let's call: $\forall n \geq 1, \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. We have: $\mathbb{E}(\bar{Y}_n) = \mu$

$$\begin{aligned} \forall n \geq 1, E(|\bar{Y}_n - \mu|^2) &= \frac{1}{n^2}\sum_{i,j=1}^{n} E((Y_i - \mu)(Y_j - \mu)) \\ E(|\bar{Y}_n - \mu|^2) &= \frac{1}{n^2}\sum_{i=1}^{n} E((Y_i - \mu)^2) + 2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} E((Y_i - \mu)(Y_j - \mu)) \\ E(|\bar{Y}_n - \mu|^2) &\leqslant \frac{1}{n^2}\sum_{i=1}^{n}\gamma(0) + 2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}|\gamma(j-i)| \\ E(|\bar{Y}_n - \mu|^2) &\leqslant \frac{1}{n^2}(n\gamma(0)) + \frac{2}{n^2}(n\sum_{k=1}^{n}|\gamma(k)| \\ E(|\bar{Y}_n - \mu|^2) &\leqslant \frac{1}{n}\sum_{k=0}^{n}|\gamma(k)| \end{aligned} \tag{2}$$

As $\sum_k |\gamma(k)|$ converges, $V(\bar{Y}_n)$ is bounded and converges to 0 with a speed in $\mathcal{O}(\frac{1}{n})$. By using once more the Bienaymé-Tchebychev inequality, we retrieve the same convergence speed for $\bar{Y}_n$.

3 AR and MA processes

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

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

$$Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k} \tag{3}$$

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 $\mathbb{E}(Y_tY_{t-k})$. Is this process weakly stationary?
- Show that the power spectrum of $\{Y_t\}_t$ is $S(f) = \sigma_\varepsilon^2|\phi(e^{-2\pi if})|^2$ 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 (3).

Answer 2

We consider $Y_t^N = \sum_{k=0}^N \psi_k\varepsilon_{t-k}$ for $N \in \mathbb{N}^*$ and a given t. As the ϵ_t is white noise, it has 0 mean. Therefore, $\forall N \in \mathbb{N}^*, \mathbb{E}(Y_t^N) = 0$. As convergence in $\mathbb{L}_2$ implies convergence in $\mathbb{L}_1$ (Jensen inequality). Therefore, we can conclude that $\lim_{N\to\infty}\mathbb{E}(Y_t^N) = \mathbb{E}(Y_t) = 0$.

For $E(Y_tY_{t-k})$, we will work directly with the limit in $\mathbb{L}_2$. Rigorously demonstrating the result employs the same argument as previously presented (also we can intervert expectation and infinite sum as they converge in $\mathbb{L}_2$).

$$\begin{aligned}
\forall k, \forall t \geq 0, E(Y_tY_{t-k}) &= E\left((\sum_{i=1}^{\infty}\psi_i\epsilon_{t-i})(\sum_{j=1}^{\infty}\psi_j\epsilon_{t-k-j}) \right) \\
E(Y_tY_{t-k}) &= E\left(\sum_{m=0}^{\infty}\sum_{n=0}^{m}\psi_n\epsilon_{t-n}\psi_{m-n}\epsilon_{t-k-m+n} \right) \\
E(Y_tY_{t-k}) &= \sum_{m=0}^{\infty}\sum_{n=0}^{m}E(\psi_n\epsilon_{t-n}\psi_{m-n}\epsilon_{t-k-m+n}) \\
E(Y_tY_{t-k}) &= \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}E(\psi_n\epsilon_{t-n}\psi_{m-n}\epsilon_{t-k-m+n}) \\
E(Y_tY_{t-k}) &= \sum_{n=0}^{\infty}E(\psi_n\psi_{n-k})E(\epsilon_{t-n}^2) \\
E(Y_tY_{t-k}) &= \sum_{n=k}^{\infty}\psi_n\psi_{n-k}\sigma_\epsilon^2
\end{aligned} \tag{4}$$

The process is thus weakly stationary, as the final sum converges according to Cauchy-Schwartz inequality.

For the power spectrum:

$$
\begin{aligned}
S(f) &= \sum_{\tau=-\infty}^{\infty} E(Y_n Y_{n-\tau}) \exp \frac{2i\pi f\tau}{f_s} \\
S(f) &= \sum_{\tau=-\infty}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n-\tau} \sigma_\epsilon^2 \exp \frac{2i\pi f\tau}{f_s} \\
S(f) &= \sum_{\tau=-\infty}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} \sigma_\epsilon^2 \exp \frac{-2i\pi f\tau}{f_s} \\
S(f) &= \sigma_\epsilon^2 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_m \exp \frac{2i\pi f(n-m)}{f_s} \\
S(f) &= \sigma_\epsilon^2 \sum_{m=0}^{\infty} \psi_m \exp \frac{-2i\pi fm}{f_s} \sum_{n=0}^{\infty} \psi_n \exp \frac{2i\pi fn}{f_s} \\
S(f) &= \sigma_\epsilon^2 \sum_{m=0}^{\infty} \psi_m \exp \frac{-2i\pi fm}{f_s} \overline{\sum_{m=0}^{\infty} \psi_m \exp \frac{-2i\pi fm}{f_s}} \\
S(f) &= \sigma_\epsilon^2 \phi(\exp \frac{-2i\pi f}{f_s}) \overline{\phi(\exp \frac{-2i\pi f}{f_s})} \\
S(f) &= \sigma_\epsilon^2 |\phi(\exp \frac{-2i\pi f}{f_s})|^2
\end{aligned}
\tag{5}
$$

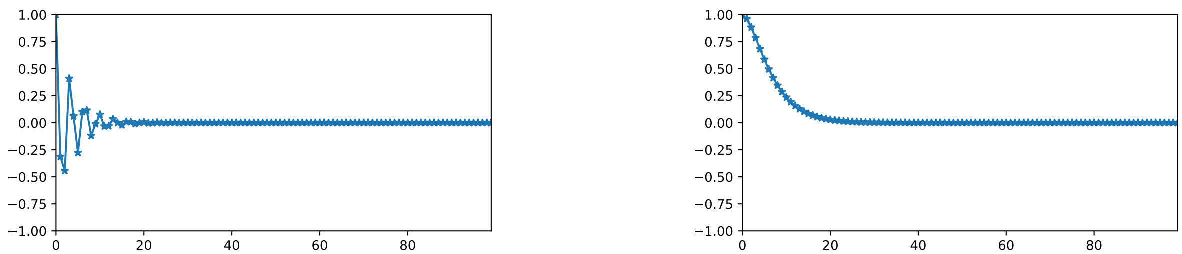
Question 3 *AR(2) process*

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

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{6}$$

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

Autocovariance $\gamma(\tau)$:

For a give τ, we have:

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

To solve a series following such a relation, we need to introduce the characteristic polynomial of the problem, denoted as Φ. We assume that r_1 and r_2 are distinct.

Hence, the solution is:

- If $r_1, r_2 \in \mathbb{R}$, then there exist unique $\lambda, \mu \in \mathbb{R}$ such that for all $\forall \tau \in \mathbb{N}$:

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

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

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

μ and λ depend on the value of $\gamma(0)$ and $\gamma(1)$.

Analysis of the figures

From the results we obtained, we can say that the graph on the right corresponds to the case with complex roots because we observe oscillation corresponding to the cosines. The graph on the right corresponds to the real roots, the autocovariance converges to 0.

Power Spectrum

We want to transform the AR(2) process into a MA(∞) process to use the previous result. To do so, we need to introduce the lag operator L such that $LY_t = Y_{t-1}$. Wehave:

$$\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\
\varepsilon_t &= Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} \\
\varepsilon_t &= Y_t - \phi_1 L Y_t - \phi_2 L^2 Y_t \\
\varepsilon_t &= (1 - \phi_1 L - \phi_2 L^2) Y_t
\end{aligned}$$

Because $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = (1 - \frac{1}{r_1}z)(1 - \frac{1}{r_2}z)$, We can write:

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

As $|r_1| > 1, |r_2| > 1$, we can write

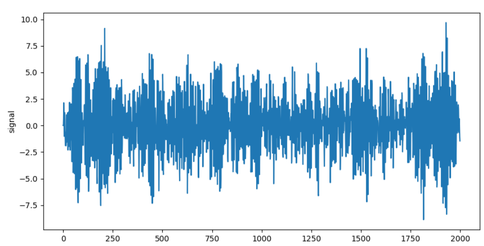
$$\frac{1}{(1 - \frac{1}{r_k}L)} = \sum_{i=0}^{\infty} \frac{L^i}{r_k^i}$$

Therefore, we can compute:

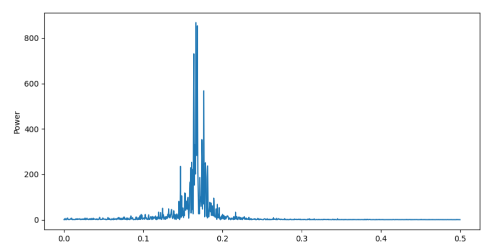
$$
\begin{aligned}
Y_t &= \sum_{i=0}^{\infty} \frac{L^i}{r_1^i} \sum_{j=0}^{\infty} \frac{L^j}{r_2^j} \epsilon_t \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{r_1^i} \frac{1}{r_j^2} L^{i+j} \epsilon_t \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{r_1^i} \frac{1}{r_2^j} \epsilon_{t-(i+j)} \\
&= \sum_{k=0}^{\infty} \left(\sum_{\substack{i,j=0 \\ i+j=k}} \frac{1}{r_1^i} \frac{1}{r_2^j} \right) \psi_k \epsilon_{t-k} = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}
\end{aligned}
\tag{7}
$$

With $\psi_k = \sum_{\substack{i,j=0 \\ i+j=k}} \frac{1}{r_1^i} \frac{1}{r_2^j}$. Because we have written our process in this form, we can apply the result of the previous question and conclude that, Φ designing the same function as before:

$$
S(f) = \sigma_\epsilon^2 |\phi(e^{-2\pi i f})|^2 \text{ with } \psi_k = \sum_{\substack{i,j=0 \\ i+j=k}}^{\infty} \frac{1}{r_1^i r_2^j}.
$$



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

$$\forall u = 0, \ldots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k+\frac{1}{2})] \tag{8}$$

where w_L is a modulating window given by

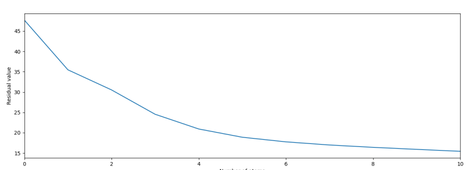
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u+\frac{1}{2}\right)\right]. \tag{9}$$

Question 4 *Sparse coding with OMP*

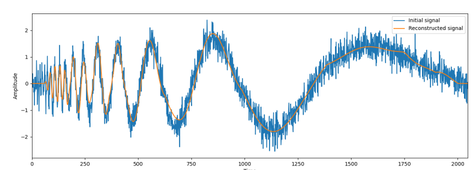
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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