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

Rania Bennani rania.bennani@ens-paris-saclay.fr
Raphael Razafindralambo raphael.razafin@gmail.com

January 14, 2024

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

- Let $X_1, ...X_n$ be iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

  Then by Bienaymé-Chebychev's inequality, we have:

  $$P(|\overline{X_n} - \mu| \geq \epsilon) \leq \frac{Var(X_i)}{n\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0.$$

  Thus, the sample mean $\overline{X}$ converges in probability to μ when n $\to \infty$

- Let $\{Y_t\}_{t\geq 1}$ a wide-sense stationary process such that $\sum_k |\gamma(k)| < +\infty$.

  By using Bienaymé-Chebychev's inequality, we have:

  $$P(|\overline{Y_n} - \mu| \geq \epsilon) \leq \frac{Var(\overline{Y_n})}{\epsilon^2}$$

  The variance of $\{Y_t\}_{t\geq 1}$ can be expressed as: $Var(\overline{Y_n}) = \frac{1}{n^2}\left[n\gamma(0) + 2\sum_{k=1}^{n-1}(n-k)\gamma(k)\right]$

  By hypothesis, $\sum_k |\gamma(k)| < +\infty$. Then, $Var(\overline{Y_n}) \xrightarrow{n\to\infty} 0$.

  Finally, $P(|\overline{Y_n} - \mu| \geq \epsilon) \xrightarrow{n\to\infty} 0$.

  Which shows that $\overline{Y_n} \xrightarrow{n\to\infty} \mu$, hence, the sample mean 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

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

  ε_t is a zero mean white noise: $\mathbb{E}(\varepsilon_t) = 0$. Hence,

  $$\mathbb{E}(Y_t) = 0.$$

  Furthemore, we have:

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

  If $k \neq 0$, $\varepsilon_{t-i}\varepsilon_{t-k-j} = 0$ and $\mathbb{E}(Y_t Y_{t-k}) = 0$.

If $k = 0$:

$$\begin{aligned}
\mathbb{E}(Y_t Y_{t-k}) &= \mathbb{E}\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) \\
&= \mathbb{E}\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \varepsilon_{t-i} \psi_j \varepsilon_{t-j}\right) \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}(\varepsilon_{t-i} \varepsilon_{t-j}) \\
&= \sum_{i=0}^{\infty} \psi_i^2 \mathbb{E}(\varepsilon_{t-i}^2) \\
&= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2.
\end{aligned}$$

Finally, $\mathbb{E}(Y_t Y_{t-k}) = \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 & \text{if } k = 0, \\ 0 & \text{if } k \neq 0. \end{cases}$

We have $\begin{cases} \mathbb{E}(Y_t) = 0 \text{ constant} \\ \mathbb{E}(Y_t Y_{t-k}) \text{ is time-invariant} \end{cases}$ .Thus, this process is weakly stationnary.

- The power spectrum $S(f)$ is the Fourier transformation of $\mathbb{E}(Y_t Y_{t-k}) = \gamma(t, t-k)$

$$S(f) = \mathcal{F}[\gamma(t, t-k)] \tag{2}$$

$$= \sum_{k=-\infty}^{\infty} \gamma(t, t-k) e^{-2\pi i f k} \tag{3}$$

$$= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \psi_k^2 e^{-2\pi i f k} \tag{4}$$

We have $\phi(z) = \sum_j \psi_j z^j$, thus:

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

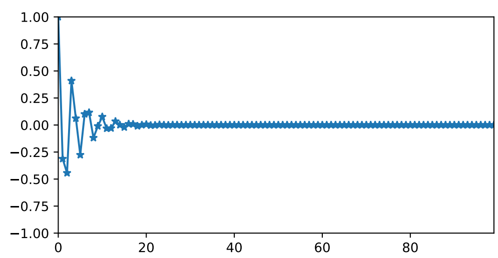
Question 3 *AR(2) process*

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

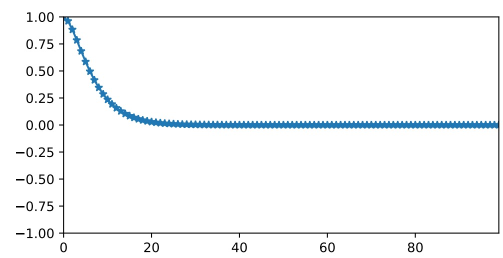
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{5}$$

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?



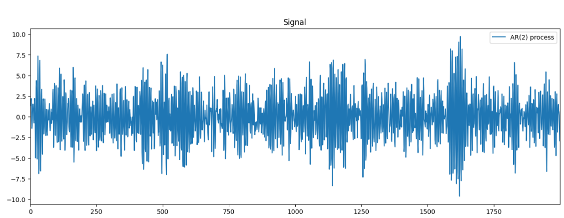
Correlogram of the first AR(2)



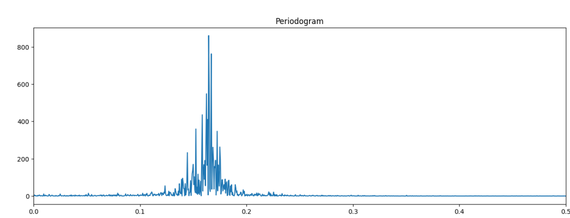
Correlogram of the second AR(2)

Figure 1: Two AR(2) processes

Answer 3



Signal



Periodogram

Figure 2: AR(2) process

- First, notice that $(Y_t)_{t\geq 1}$ is stationary with null expectation. Indeed, Y_t is stationary as an AR(2) process and

$$\mathbb{E}[Y_t] = \phi_1 \mathbb{E}[Y_{t-1}] + \phi_2 \mathbb{E}[Y_{t-2}] + 0 \implies \mu_Y(1 - \phi_1 - \phi_2) = 0 \implies \mu_Y = 0.$$

since 1 is not a root of ϕ. Immediately it follows that $\gamma(\tau) = \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2)$:

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

It's a second order linear homogeneous recurrence relation with characteristic equation $\psi(z) = z^2 - \phi_1 z - \phi_2 = 0$. Since $z^2\psi(\frac{1}{z}) = \phi(r)$ and r_1, r_2 are roots of ϕ, then roots of ψ are given by $\xi_1 = \frac{1}{r_1}$ and $\xi_2 = \frac{1}{r_2}$. Those roots are distincts, so we distinguish two cases.

1. $r_1, r_2 \in \mathbb{R}$. Then there exists λ, μ such that

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

   To give an explicit form to γ, we solve the following system:

$$\gamma(0) = \lambda + \mu \tag{6}$$

$$\gamma(1) = \frac{\lambda}{r_1} + \frac{\lambda}{r_2} \tag{7}$$

   where $r_1, r_2, \gamma(0), \gamma(1)$ are fixed (or known). The couple of solutions is unique and is

$$\begin{aligned}\lambda &= \frac{1}{\frac{1}{r_1} - \frac{1}{r_2}}(\gamma(1) - \frac{1}{r_2}\gamma(0)) \\ \mu &= \gamma(0) - \frac{1}{\frac{1}{r_1} - \frac{1}{r_2}}(\gamma(1) - \frac{1}{r_2}\gamma(0))\end{aligned}$$

2. $r_1, r_2 \in \mathbb{C}$. Then, there exists λ, μ, α such that

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

   where $\begin{cases} r_1 = re^{i\alpha} \\ r_2 = re^{-i\alpha} \end{cases}, \alpha \notin \{2k\pi\}$.

   By expressing this equation with $\tau = 0$ and $\tau = 1$ we retrieve again a system with two equations and two unknowns, and so

$$\begin{aligned}\lambda &= \lambda\cos(0) + \mu\sin(0) = \gamma(0) \\ \mu &= \frac{\gamma(1) - \gamma(0) r\cos(\alpha)}{r\sin(\alpha)}\end{aligned}$$

   .

- The first AR(2) process is the one with complex roots since γ is a combinaison of sinusoidal functions. The second AR(2) process is the one with real ones since γ is decreasing with respect to τ (r_1 and r_2 are larger than one in absolute value).

- Any AR(2) process can be expressed as a MA(∞) process. Thus, we can use the result of Question 2. Let's show that there exists $\tilde{\psi}_0, \ldots, \tilde{\psi}_n$ such that

$$Y_t = \lim_{n \to +\infty} \sum_{k=0}^{n} \tilde{\psi}_k \epsilon_{t-k}.$$

Let's denote B the backshift operator ($BY_t = Y_{t-1}$ for all t). Then we can rewrite from (5):

$$\begin{aligned} &(1 - \phi_1 B - \phi_2 B^2) Y_t = \epsilon_t \\ \iff &(1 - \xi_1 B)(1 - \xi_2 B) Y_t = \epsilon_t \end{aligned}$$

by definition of ξ_1, ξ_2. Thus,

$$\begin{aligned} Y_t &= \frac{1}{(1 - \xi_1 B)(1 - \xi_2 B)} \epsilon_t \\ &= \left(\frac{a}{(1 - \xi_1 B)} + \frac{b}{(1 - \xi_2 B)} \right) \epsilon_t \end{aligned}$$

by the partial fraction decomposition. Finally, since $\frac{1}{1-x} = \lim_{n \to +\infty} \sum_{k=0}^{n} x^k$ when $|x| < 1$, by noticing that $|\xi_1| = |\frac{1}{r_1}| < 1$ and $|\xi_2| = |\frac{1}{r_2}| < 1$ we have

$$\begin{aligned} Y_t &= \left(\frac{a}{(1 - \xi_1 B)} + \frac{b}{(1 - \xi_2 B)} \right) \epsilon_t \\ &= \left(\lim_{n \to +\infty} \sum_{k=0}^{n} a \xi_1^k B^k + \lim_{n \to +\infty} \sum_{k=0}^{n} b \xi_2^k B^k \right) \epsilon_t \\ &= \left(\lim_{n \to +\infty} \sum_{k=0}^{n} a \xi_1^k B^k + b \xi_2^k B^k \right) \epsilon_t \\ &= \lim_{n \to +\infty} \sum_{k=0}^{n} (a \xi_1^k + b \xi_2^k) \epsilon_{t-k} \end{aligned}$$

We recognize a MA(∞) process with $\tilde{\psi}_k = a\xi_1^k + b\xi_2^k$. Hence,

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

where $\tilde{\phi}(z) = \sum_j \tilde{\psi}_j z^j$.

- The roots are given by $z_1 = re^{i\theta}$ and $z_2 = re^{-i\theta}$. $r = 1.05$ and $\theta = \frac{\pi}{3}$. Let's consider a norm equal to $\frac{1}{1.05}$. Then the product $(z - z_1)(z - z_2)$ gives:

$$\begin{aligned} (z - z_1)(z - z_2) &= (z - \frac{1}{1.05} \left(\frac{1}{2} + i\frac{\sqrt{3}}{2} \right))(z - \frac{1}{1.05} \left(\frac{1}{2} - i\frac{\sqrt{3}}{2} \right)) \\ &= z^2 - z\frac{1}{1.05} + \frac{1}{1.1025} \\ &= z^2 - z\phi_1 - \phi_2 \end{aligned}$$

z_1, z_2 are solutions of $z^2 - z\frac{1}{1.05} + \frac{1}{1.1025} = 0$ so their inverse (with norm 1.05) are solutions of $1 - \frac{1}{1.05}z + \frac{1}{1.1025}z^2 = 0$. This leads to

$$\phi_1 = \frac{1}{1.05}, \phi_2 = -\frac{1}{1.1025}.$$

- The signal is stationary with $\mu_Y = 0$, we saw in the first question that it's because the modulus of the roots are strictly larger than one. The periodogram shows a flat spectrum with a single peak at a specific frequency ($\sim$ 1.75 Hz). It typically indicates the presence of a dominant periodic component at that frequency.

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

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

where w_L is a modulating window given by

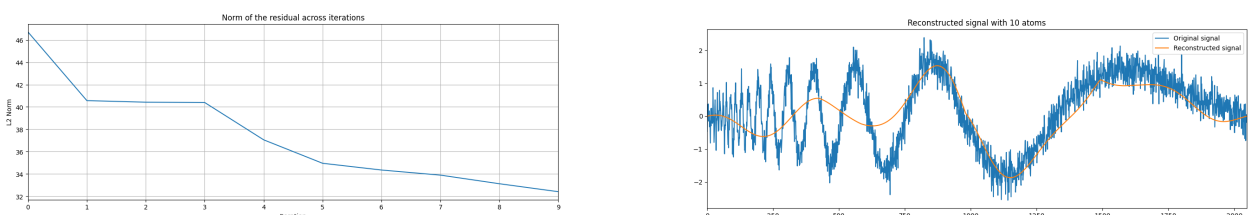
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{9}$$

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

We see that the reconstructed signal follows roughly follows the true signal for $K_0 = 10$. The lack of accuracy (in particular for early times) could be linked to the quality of the selection of the atoms, which may not represent the signal well. The result is more satisfying if we increase the number of selected atoms K_0. Nonetheless we see that the residual is decreasing.