Assignment 2 (ML for TS) - MVA 2022/2023

Théo Di Piazza theo.dipiazza@gmail.com
Stanislas du Ché stanislasduche@gmail.com

February 27, 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 Monday 27$^{\text{th}}$ February 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: .

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

- The rate of convergence quantifies how fast the estimation error decreases when increasing the sample size n.

  Let $Y_1, Y_2, ..., Y_n$ i.i.d random variables with finite mean μ and finite variance σ^2.

  By Central Limit Theorem, it comes that :

$$\sqrt{n}\frac{\bar{Y}_n - \mu}{\sigma} \underset{n}{\rightarrow} \mathcal{N}(0,1)$$

  Hence, $\boxed{\text{the convergence rate is in } \frac{1}{\sqrt{n}} : \mathcal{O}(\frac{1}{\sqrt{n}})}$

- By using the hint, it comes that :

$$\mathbb{E}\left[(\bar{Y}_n - \mu)^2\right] = \mathbb{E}\left[\bar{Y}_n{}^2 - 2\bar{Y}_n\mu + \mu^2\right] = \mathbb{E}\left[\bar{Y}_n{}^2\right] - \mu^2 \leq \mathbb{E}\left[\bar{Y}_n{}^2\right]$$

  Moreover by distributive law,

$$\mathbb{E}\left[\bar{Y}_n{}^2\right] = \frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} Y_j\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[Y_iY_j\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{|j-i|}$$

$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k}|\gamma(k)| \leq \frac{1}{n}\sum_{k}|\gamma(k)| \underset{n}{\rightarrow} 0 \text{ since } \sum_{k}|\gamma(k)| < \infty$$

  It has been shown that : $\mathbb{E}\left[(\bar{Y}_n - \mu)^2\right] \underset{n}{\rightarrow} 0 \Longrightarrow \bar{Y}_n \xrightarrow{L^2} \mu \Longrightarrow \bar{Y}_n \xrightarrow{p} \mu$

  Hence, $\boxed{\text{the sample mean is consistent.}}$

  About the rate of convergence, it comes that :

$$(\mathbb{E}\left[\bar{Y}_n - \mu\right])^2 \underset{Jensen}{\leq} \mathbb{E}\left[(\bar{Y}_n - \mu)^2\right] \leq \frac{1}{n}\sum_{k}|\gamma(k)|$$

$$\Longrightarrow \mathbb{E}\left[\bar{Y}_n - \mu\right] \leq \sqrt{\frac{1}{n}\sum_{k}|\gamma(k)|}$$

  Hence, $\boxed{\text{the rate of convergence is still in } \frac{1}{\sqrt{n}}}$

3 AR and MA processes

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

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

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

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

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

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

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

$$\mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-k-j}\right] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}(\varepsilon_{t-i})\mathbb{E}(\varepsilon_{t-k-j})$$

  by Fubini and independance of the ϵ_t Where $\mathbb{E}(\varepsilon_{t-i})\mathbb{E}(\varepsilon_{t-k-j}) = \begin{cases} \sigma_\epsilon^2 & \text{if } i = k+j \\ 0 & \text{otherwise} \end{cases}$

  It comes that :

$$\mathbb{E}(Y_t Y_{t-k}) = \sum_{i=k}^{\infty} \psi_i \psi_{i-k} \sigma_\epsilon^2 \text{ which is independant of the variable t !}$$

  Finally,

$$var(Y_t) = \sum_{k=0}^{\infty} {\psi_k}^2 \sigma_\epsilon^2 < \infty$$

  Hence, this process is $\boxed{\text{weakly stationary.}}$

- By defining γ_ϵ as the autocovariance function of $\{\epsilon_t\}_t$, it comes that :

$$\mathbb{E}\left[Y_t Y_{t-h}\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \epsilon_{t-h-j}\right] = \sum_{i,j=0}^{\infty} \psi_i \psi_j \mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-h-j}\right] = \sum_{i,j=0}^{\infty} \psi_i \psi_j \gamma_\epsilon(h+j-i)$$

Then, we have that :

$$S(f) = \sum_{h=\infty}^{\infty} \gamma(h) e^{-2i\pi fh} = \sum_{h=\infty}^{\infty} e^{-2i\pi fh} \sum_{k,j=0}^{\infty} \psi_k \psi_j \gamma_\epsilon(h+j-k)$$

With the change of variable (l = h+k-j), it comes that :

$$S(f) = \sum_{j=0}^{\infty} e^{-2i\pi fj} \psi_j \sum_{k=0}^{\infty} e^{2i\pi fk} \psi_k \sum_{l=\infty}^{\infty} e^{-2i\pi fl} \gamma_\epsilon(l) = \sigma_\epsilon^2 | \sum_{j=0}^{\infty} \psi_j e^{-2i\pi fj} |^2$$

Hence, finally :

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

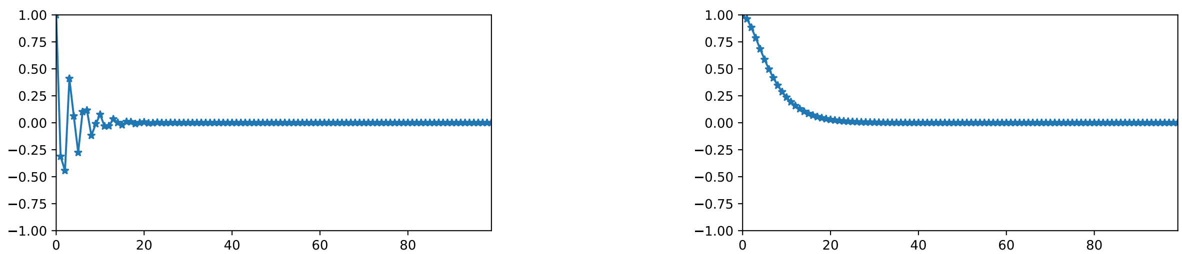
Question 3 *AR(2) process*

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

- We have for $0 \leq \tau$:

$$\begin{aligned}
&Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t \\
&\Rightarrow Y_t Y_{t-\tau} - \phi_1 Y_{t-1} Y_{t-\tau} - \phi_2 Y_{t-2} Y_{t-\tau} = \epsilon_t Y_{t-\tau} \\
&\Rightarrow \mathbb{E}[Y_t Y_{t-\tau}] - \phi_1 \mathbb{E}[Y_{t-1} Y_{t-\tau}] - \phi_2 \mathbb{E}[Y_{t-2} Y_{t-\tau}] = \mathbb{E}[\epsilon_t Y_{t-\tau}] \\
&\Rightarrow \gamma(\tau) - \phi_1 \gamma(\tau - 1) - \phi_2 \gamma(\tau - 2) = \sigma_\epsilon^2 \delta_0(\tau)
\end{aligned}$$

  We have a recurrent series of order 2, and is characteristic polynomial is $P(z) = z^2 - \phi_1 z - \phi_2 = z^2 \phi(\frac{1}{z})$ annealed by r_1 and r_2.

  Finally, we have the following expression:

$$\gamma(\tau) = a(\frac{1}{r_1})^\tau + b(\frac{1}{r_2})^\tau + \sigma_\epsilon^2 \delta_0(\tau), \quad a, b \in \mathbb{R} \tag{3}$$

- With $r_{1/2} = re^{\pm i\theta}$, we have $\frac{1}{r_{1/2}}^{\tau} = \frac{1}{r}^{\tau} e^{\mp i\tau\theta}$, so we have $\gamma(\tau) = \left(\frac{1}{r}\right)^{\tau} (a'cos(\tau\theta) + b'sin(\tau\theta))$, with $a', b' \in \mathbb{C}$.

  In case of real roots, the equation 3 have the one of a decreasing exponential, whereas in the complex case, the second form shows us the one of a sinusoidal function. Thus, the first is has complex roots and the second graph has real ones.

- To start, it is known that :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \text{ and } p(x) = 1 - \phi_1 x - \phi_2 x^2$$

  For the lag operator L, it comes that :

$$p(L)Y_t = (1 - \phi_1 L - \phi_2 L^2)Y_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t$$

  For $P(L) > 0$ and with Taylor expansion, it comes that :

$$P(L)Y_t = \epsilon_t \Longleftrightarrow Y_t = \frac{\epsilon_t}{p(L)} = \frac{\epsilon_t}{(1 - \frac{L}{r1})(1 - \frac{L}{r1})} = \epsilon_t (\sum_{i=0}^{\infty} \phi_1^i L^i)(\sum_{j=0}^{\infty} \phi_2^j L^j)$$

  Thus, the characteristic polynomial associated with the AR(2) process is equal to the inverse of the characteristic polynomial associated with the MA(∞) process.

  It comes that for the lag operator L :

$$\psi(L) = \frac{1}{\phi(L)}$$

  where $\psi(L) = \sum_j^{\infty} \psi_j L^j$ and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$

  Question 2, it has been shown that for a $MA(\infty)$ process :

$$S(f)_{MA(\infty)} = \sigma_\epsilon^2 |\psi(e^{-2i\pi f})|^2$$

  Hence with what with just explained, it comes that for a $AR(2)$ process that :

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

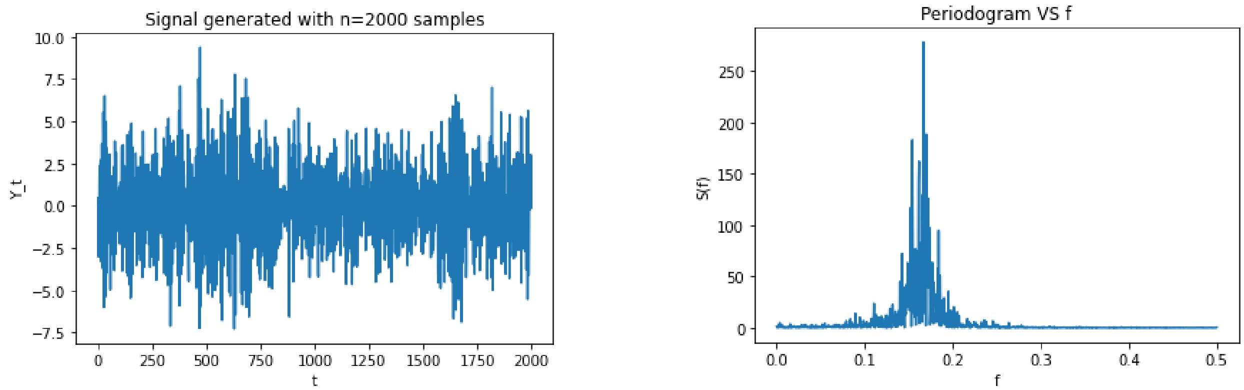
- To start, let's express ϕ_1 and ϕ_2 as a function of r_1 and r_2 knowing that $\phi(r_1) = \phi(r_2) = 0$

$$\begin{cases} \phi(r_1) = 0 \\ \phi(r_2) = 0 \end{cases} \Longrightarrow \begin{cases} 1 - \phi_1 r_1 - \phi_2 r_1^2 = 0 \\ 1 - \phi_1 r_2 - \phi_2 r_2^2 = 0 \end{cases} \Longrightarrow \begin{cases} \phi_1(r_2 - r_1) + \phi_2(r_2^2 - r_1^2) = 0 \\ 1 - \phi_1 r_2 - \phi_2 r_2^2 = 0 \end{cases}$$

$$\Longrightarrow \begin{cases} \phi_1 = -\phi_2(r_2 + r_1) \\ 1 + \phi_2(r_2 + r_1)r_1 - \phi_2 r_1^2 = 0 \end{cases} \Longrightarrow \begin{cases} \phi_1 = \frac{r_1 + r_2}{r_1 r_2} \\ \phi_2 = \frac{-1}{r_1 r_2} \end{cases}$$

  With $r_{1/2} = re^{\pm i\theta}$, it comes that :

$$\phi_1 = \frac{2cos(\theta)}{r} \text{ and } \phi_1 = \frac{-1}{r^2}$$



Signal

Periodogram

Figure 2: AR(2) process

We observe a peak around the frequency : $f = 0.16$. It would suggest that there is a dominant frequency component in the time series 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 ($k = 0, \ldots, L-1$) 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

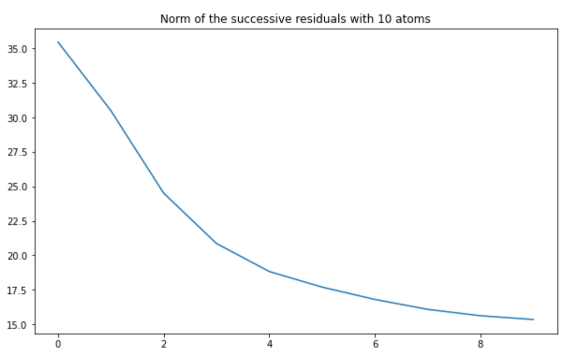
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{5}$$

Question 4 *Sparse coding with OMP*

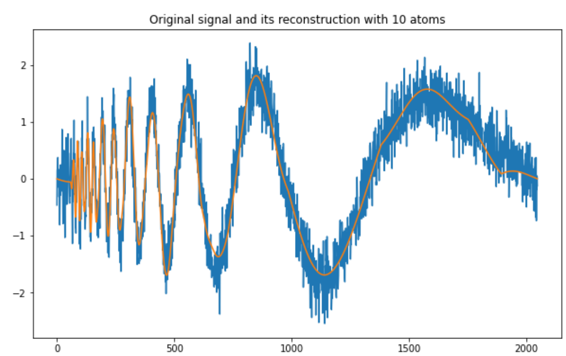
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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



Norms of the successive residuals



Reconstruction with 10 atoms

Figure 3: Question 4