Assignment 2 (ML for TS) - MVA

Antoine Le Maguet antoine.lemaguet@free.fr
Alexandre Mallez alexandre.mallez@gmail.com

December 7, 2025

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 Sunday 7$^{\text{th}}$ December 11:59 PM.
- Rename your report and notebook as follows:
`FirstnameLastname1_FirstnameLastname1.pdf` and
`FirstnameLastname2_FirstnameLastname2.ipynb`.
For instance, `LaurentOudre_ValerioGuerrini.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: https://forms.gle/J1pdeHspSs9zNfWAA.

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 realizations 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 i.i.d. random variables with finite variance $\mathrm{Var}[X_i] = \sigma^2$ and mean $\mathbb{E}[X_i] = \mu$, the sample mean $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is such that $\mathbb{E}[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$. Then :

$$\boxed{\|\bar{X}_n - \mu\|_2 = \sqrt{\mathbb{E}[(\bar{X}_n - \mu)^2]} = \sqrt{\mathrm{Var}(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to+\infty]{} 0$$

  Thus we have convergence in L_2 at rate $1/\sqrt{n}$, which implies convergence in probability at the same rate by Markov's inequality.

- For the wide-sense stationary process $\{Y_t\}_{t\geq 1}$, we have that $\mathbb{E}[Y_t] = \mu$, then :

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \mathrm{Var}(\bar{Y}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{t=1}^{n} Y_t\right) = \frac{1}{n^2}(\sum_{t=1}^{n}\mathrm{Var}(Y_t) + 2\sum_{1\leq t<s\leq n}\mathrm{Cov}(Y_t, Y_s))$$

  Using that $Cov(Y_t, Y_s) = \gamma(s-t)$, we can rewrite this as

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \frac{1}{n^2}\left(n\gamma(0) + 2\sum_{k=1}^{n-1}(n-k)\gamma(k)\right) = \frac{1}{n}\left(\gamma(0) + \frac{2}{n}\sum_{k=1}^{n-1}(n-k)\gamma(k)\right)$$

  We can bound the sum:

$$|\frac{1}{n}\sum_{k=1}^{n-1}(n-k)\gamma(k)| \leq \frac{n}{n}\sum_{k=1}^{\infty}|\gamma(k)| = O(1)$$

  We have then : $\frac{1}{n}\sum_{k=1}^{n-1}(n-k)\gamma(k) = O(1)$. Thus:

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \frac{\gamma(0)}{n} + \frac{2}{n}\frac{1}{n}\sum_{k=1}^{n-1}(n-k)\gamma(k) = O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right)$$

  Thus, $\boxed{\|\bar{Y}_n - \mu\|_2 = O(1/\sqrt{n})}$, which implies convergence in probability at the same rate by Markov's inequality.

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

- Noting $S_{N,t} = \sum_{k=0}^{N} \psi_k\varepsilon_{t-k}$, we have that $Y_t = \lim_{N\to\infty} S_{N,t}$ in L_2 (because the ψ_k are square summable and the ε_t have finite variance). Besides ε_t is zero mean. Thus:

$$\boxed{\mathbb{E}[Y_t] = \mathbb{E}[\lim_{N\to\infty} S_{N,t}] = \lim_{N\to\infty} \mathbb{E}[S_{N,t}] = \lim_{N\to\infty} \sum_{k=0}^{N} \mathbb{E}[\psi_k\varepsilon_{t-k}] = \lim_{N\to\infty} \sum_{k=0}^{N} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0}$$

Also using the L_2 convergence, we have:

$$\mathbb{E}[Y_tY_{t-k}] = \mathbb{E}[\lim_{N\to\infty} S_{N,t} \lim_{M\to\infty} S_{M,t-k}] = \lim_{N,M\to\infty} \mathbb{E}[S_{N,t}S_{M,t-k}] = \lim_{N,M\to\infty} \mathbb{E}\left[\left(\sum_{i=0}^{N} \psi_i\varepsilon_{t-i}\right)\left(\sum_{j=0}^{M} \psi_j\varepsilon_{t-k-j}\right)\right]$$

Using the linearity of expectation :

$$\mathbb{E}[Y_tY_{t-k}] = \lim_{N,M\to\infty} \sum_{i=0}^{N}\sum_{j=0}^{M} \psi_i\psi_j \mathbb{E}[\varepsilon_{t-i}\varepsilon_{t-k-j}]$$

Since ε_t is a white noise of variance σ_ε^2, we have that $\mathbb{E}[\varepsilon_{t-i}\varepsilon_{t-k-j}] = \sigma_\varepsilon^2 \delta_{i,k+j}$. Thus :

$$\boxed{\mathbb{E}[Y_tY_{t-k}] = \sigma_\varepsilon^2 \lim_{N,M\to\infty} \sum_{i=0}^{N}\sum_{j=0}^{M} \psi_i\psi_j\delta_{i,k+j} = \sigma_\varepsilon^2 \lim_{N,M\to\infty} \sum_{j=0}^{\min(M,N-k)} \psi_{j+k}\psi_j = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_{j+k}\psi_j}$$

$\mathbb{E}[Y_t]$ doesn't depend on t, $\mathbb{E}[Y_tY_{t-k}]$ only depends on k and $\mathbb{E}[Y_t^2] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$. Thus the process is weakly stationary.

- The power spectrum is defined as the Fourier transform of the autocovariance function $\gamma(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_{j+k}\psi_j$, using $f_s = 1$ Hz:

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-2\pi i f k} = \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \left(\sum_{j=0}^{\infty} \psi_{j+k}\psi_j \right) e^{-2\pi i f k}$$

Changing the order of summation (justified by the absolute convergence of the sums):

$$S(f) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \left(\sum_{k=-\infty}^{\infty} \psi_{j+k} e^{-2\pi i f k} \right) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j e^{2\pi i f j} \left(\sum_{k=-\infty}^{\infty} \psi_{j+k} e^{-2\pi i f (j+k)} \right)$$

Using the change of variable $m = j + k$ in the inner sum:

$$S(f) = \sigma_\varepsilon^2 \left(\sum_{j=0}^{\infty} \psi_j e^{2\pi i f j} \right) \left(\sum_{m=-\infty}^{\infty} \psi_m e^{-2\pi i f m} \right)$$

Since $\psi_m = 0$ for $m < 0$, we have:

$$\boxed{S(f) = \sigma_\varepsilon^2 \left| \sum_{j=0}^{\infty} \psi_j e^{-2\pi i f j} \right|^2 = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2}$$

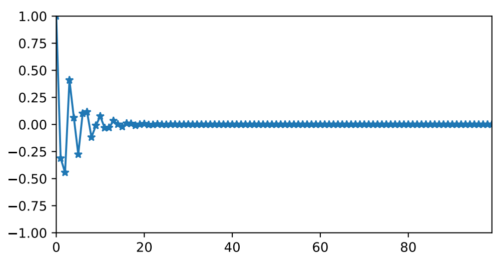
Question 3 *AR(2) process*

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

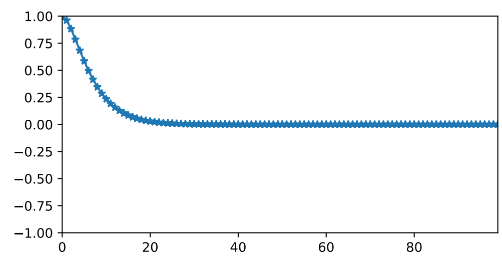
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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 behavior 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

- By taking equation (2) and multyplying by $Y_{t-\tau}$ (for $\tau \geq 2$) and taking the expectation we obtain, by using that ε_t is uncorrelated with the past :

  $$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2)$$

  The characteristic polynomial of this linear recurrence is : $\psi(z) = z^2 - \phi_1 z - \phi_2$

  We remark that $\psi(z) = -z^2 \phi(1/z)$, thus the roots of ψ are the inverse of the roots of ϕ, i.e. $1/r_1$ and $1/r_2$. Since $|r_i| > 1$, we have that $|1/r_i| < 1$. Thus the general solution of the linear recurrence is $\forall \tau \geq 0$:

  $$\gamma(\tau) = A(1/r_1)^\tau + B(1/r_2)^\tau$$

  where A and B are constants that can be determined using the initial conditions

  Replacing $\gamma(0)$ and $\gamma(1)$ in the general solution, we can solve for A and B:

  $$\begin{cases} A + B = \gamma(0) \\ A(1/r_1) + B(1/r_2) = \gamma(1) \end{cases}$$

$$\boxed{A = \frac{\gamma(1) - \gamma(0)/r_2}{1/r_1 - 1/r_2}, \quad B = \frac{\gamma(0)/r_1 - \gamma(1)}{1/r_1 - 1/r_2}}$$

To compute $\gamma(0)$ and $\gamma(1)$, we use again equation (2) multiplied by Y_t and Y_{t-1} respectively and taking the expectation :

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma_\varepsilon^2$$
$$\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1)$$

With the second equation, we have that $\gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0)$. Replacing in the first equation and using that $\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0)$, we have :

$$\gamma(0) = \phi_1\frac{\phi_1}{1-\phi_2}\gamma(0) + \phi_2(\phi_1\frac{\phi_1}{1-\phi_2}\gamma(0) + \phi_2\gamma(0)) + \sigma_\varepsilon^2$$

Thus :

$$\boxed{\gamma(0) = \frac{\sigma_\varepsilon^2(1-\phi_2)}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}}$$

And :

$$\boxed{\gamma(1) = \frac{\phi_1\sigma_\varepsilon^2}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}}$$

Morover, for $\tau < 0$, we have by symmetry that $\gamma(\tau) = \gamma(-\tau)$:

$$\boxed{\forall\tau \in \mathbb{Z}, \quad \gamma(\tau) = A(1/r_1)^{|\tau|} + B(1/r_2)^{|\tau|}}$$

with A and B that can be expressed using ϕ_1, ϕ_2 and σ_ε^2 as above.

- The first AR(2) process has complex roots (oscillatory behavior in the correlogram), while the second one has real roots (exponential decay in the correlogram).
- We can use the MA(∞) representation of the AR(2) process to compute the power spectrum (possible since the roots of the characteristic polynomial are outside the unit circle). Using B the backshift operator, we have :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \iff (1 - \phi_1 B - \phi_2 B^2)Y_t = \varepsilon_t \iff Y_t = \frac{1}{\phi(B)}\varepsilon_t$$

Since the root of ϕ are outside the unit circle, we can expand $\frac{1}{\phi(B)}$ as a power series in B to have ψ_k such that :

$$\frac{1}{\phi(B)} = \sum_{k=0}^{\infty}\psi_k B^k$$

noting $g(z) = \sum_{k=0}^{\infty}\psi_k z^k$, we have that $g(z)\phi(z) = 1$, thus $g(z) = 1/\phi(z)$. We have then the MA(∞) representation :

$$Y_t = g(B)\varepsilon_t = \sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}$$

Thus, the power spectrum is (using the result from the previous exercise) :

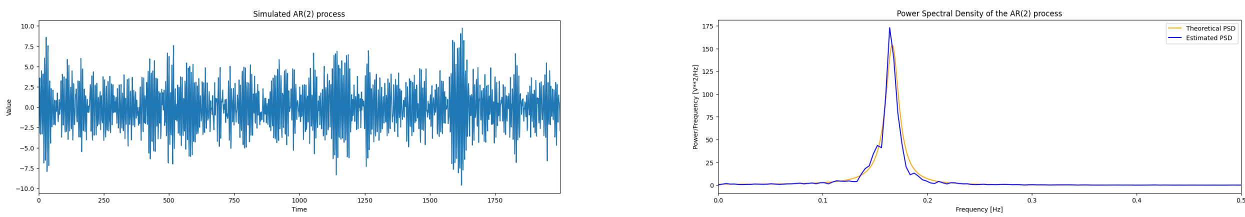
$$\boxed{S(f) = \sigma_\varepsilon^2|g(e^{-2\pi i f})|^2 = \frac{\sigma_\varepsilon^2}{|\phi(e^{-2\pi i f})|^2}}$$

- We want $r_1 = 1.05e^{2\pi i/6}$ and $r_2 = 1.05e^{-2\pi i/6}$, then since:

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

  We then take :

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



Signal　　Periodogram

Figure 2: AR(2) process

We observe an oscillatory behavior in the signal, which is coherent with the complex roots of the characteristic polynomial. In the periodogram, we can see a peak at frequency $1/6$ Hz, which corresponds to the phase of the roots $\theta = 2\pi/6$.

4 Sparse coding

The modulated discrete cosine transform (MDCT) is a signal transformation often used in sound processing applications (for instance, to encode an 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

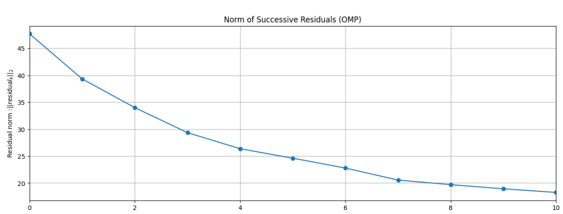
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

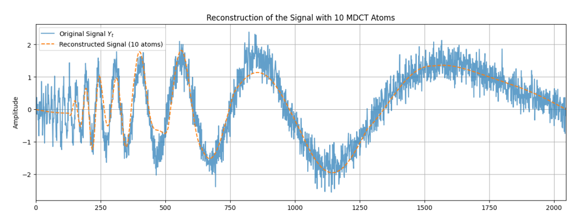
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- For the sparse coding, implement the Orthogonal Matching Pursuit (OMP). (Use convolutions to compute the correlation 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

These plots demonstrate that the chosen dictionary is very well suited ("consistent") with the signal being analyzed. The rapid convergence (left) and visual fidelity (right) prove that the OMP algorithm successfully isolated the essential structure of the signal in very few operations.