

Assignment 2 (ML for TS) - MVA

Adonis Jamal adonis.jamal@student-cs.fr
Fotios Kapotos fotiskapotos@mail.com

November 28, 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 7th 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 (Ω, \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 $\bar{Y}_n = (Y_1 + \dots + Y_n)/n$ 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 $(X_i)_{i \geq 1}$ be i.i.d. with mean μ and variance $\sigma^2 < \infty$, and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Then

$$\mathbb{E}[\bar{X}_n] = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Hence $\bar{X}_n \rightarrow \mu$ in probability and the typical size of the error is $\sqrt{\text{Var}(\bar{X}_n)} = \sigma/\sqrt{n}$. Moreover, by the Central Limit Theorem,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

so the convergence rate of \bar{X}_n toward μ is $n^{-1/2}$.

- By stationarity, $\mathbb{E}[\bar{Y}_n] = \mu$. Moreover,

$$\text{Var}(\bar{Y}_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(j-i) = \frac{1}{n} \left(\gamma(0) + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k) \right).$$

Let $A_n := \gamma(0) + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k)$; we show $A_n \rightarrow A < \infty$.

Since $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$, in particular $\sum_{k=1}^{\infty} |\gamma(k)| < \infty$. Hence for any $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that

$$\sum_{k=K+1}^{\infty} |\gamma(k)| < \varepsilon.$$

Then, for all n ,

$$\sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k) = \sum_{k=1}^K \left(1 - \frac{k}{n} \right) \gamma(k) + \sum_{k=K+1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k).$$

The first finite sum converges to $\sum_{k=1}^K \gamma(k)$ by dominated convergence (since $|1 - k/n| \leq 1$). For the tail,

$$\left| \sum_{k=K+1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k) \right| \leq \sum_{k=K+1}^{\infty} |\gamma(k)| < \varepsilon.$$

Thus the whole sum converges to $\sum_{k=1}^{\infty} \gamma(k)$, and

$$A_n \xrightarrow[n \rightarrow \infty]{} A := \gamma(0) + 2 \sum_{k=1}^{\infty} \gamma(k) < \infty.$$

Therefore $\text{Var}(\bar{Y}_n) \sim \frac{A}{n}$, which implies $\text{Var}(\bar{Y}_n) \rightarrow 0$ and hence $\bar{Y}_n \xrightarrow{\mathbb{P}} \mu$, with rate $n^{-1/2}$ as in the i.i.d. case.

3 AR and MA processes

Question 2 *Infinite order moving average $MA(\infty)$*

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

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad (1)$$

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

- Since the series converges in L^2 and $\mathbb{E}[\varepsilon_t] = 0$,

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0.$$

For $h \in \mathbb{Z}$,

$$\mathbb{E}[Y_t Y_{t-h}] = \mathbb{E}\left[\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \sum_{m=0}^{\infty} \psi_m \varepsilon_{t-h-m}\right] = \sum_{j,m \geq 0} \psi_j \psi_m \mathbb{E}[\varepsilon_{t-j} \varepsilon_{t-h-m}].$$

By whiteness, $\mathbb{E}[\varepsilon_{t-j} \varepsilon_{t-h-m}] = \sigma_\varepsilon^2 \mathbf{1}_{\{t-j=t-h-m\}}$, i.e. nonzero only if $m = j - h$. Using the convention $\psi_k = 0$ for $k < 0$, we obtain

$$\gamma_Y(h) := \mathbb{E}[Y_t Y_{t-h}] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j-h},$$

which depends only on the lag h . Moreover, by Cauchy–Schwarz,

$$\sum_{j=0}^{\infty} |\psi_j \psi_{j-h}| \leq \left(\sum_{j=0}^{\infty} \psi_j^2\right)^{1/2} \left(\sum_{j=0}^{\infty} \psi_{j-h}^2\right)^{1/2} = \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

so $\gamma_Y(h)$ is finite for all h . Hence $(Y_t)_t$ is weakly stationary with mean 0 and autocovariance $\gamma_Y(h)$.

- The spectral density is the Fourier series of $\gamma_Y(h)$:

$$S(f) = \sum_{h=-\infty}^{\infty} \gamma_Y(h) e^{-2\pi i f h}, \quad f \in [-\frac{1}{2}, \frac{1}{2}].$$

Using $\gamma_Y(h) = \sigma_\varepsilon^2 \sum_{j \geq 0} \psi_j \psi_{j-h}$ and absolute convergence (from $\sum_k \psi_k^2 < \infty$), we can interchange the sums:

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

where we set $m = j - h$. Thus

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

where $\varphi(z) = \sum_{j=0}^{\infty} \psi_j z^j$.

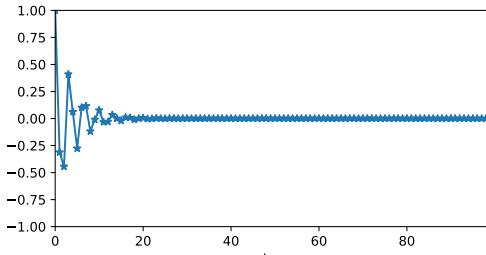
Question 3 AR(2) process

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

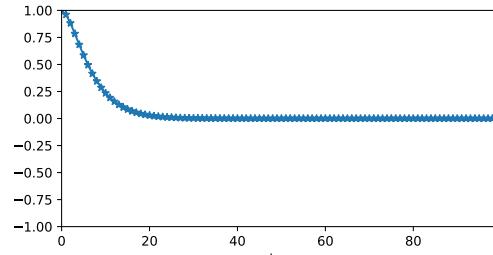
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad (2)$$

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

- **Autocovariance coefficients:** Assuming that Y_1 and Y_2 are centered, $k \geq 0$, we have

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

Taking the expectation and using the fact that ε_t is independent of Y_{t-k} for $k > 0$, we get:

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

The characteristic polynomial associated to this linear recurrence relation is the same as the one of the AR(2) process, i.e. $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$. Thus, the general solution of this recurrence relation is given by

$$\gamma(k) = A r_1^{-k} + B r_2^{-k},$$

where A and B are constants determined by initial conditions $\gamma(0)$ and $\gamma(1)$.

- **Correlograms:** The first AR(2) process has complex roots, as we can see from the oscillatory behavior of its correlogram. The second AR(2) process has real roots, as its correlogram

decays exponentially without oscillations. We can prove this by setting $r_{1,2} = \rho e^{\pm i\theta}$, with $\rho > 1$ and $\theta \in (0, \pi)$. Then, we have $r_{1,2}^{-k} = \rho^{-k} e^{\mp ik\theta}$, and thus

$$\gamma(k) = \rho^{-k}(C \cos(k\theta) + D \sin(k\theta))$$

for some constants $C, D \in \mathbb{C}$. This expression shows the oscillatory nature of the autocovariance function when the roots are complex.

- **Power spectrum:** We introduce a lag operator L such that $LY_t = Y_{t-1}$. The AR(2) process can be rewritten as

$$\phi(L)Y_t = \varepsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$. We can express $\phi(L)$ as

$$\phi(L) = \left(1 - \frac{L}{r_1}\right) \left(1 - \frac{L}{r_2}\right).$$

Since $|r_i| > 1$, we can rewrite the process as

$$Y_t = \frac{1}{\left(1 - \frac{L}{r_1}\right) \left(1 - \frac{L}{r_2}\right)} \varepsilon_t = \varepsilon_t \left(\sum_{k=0}^{\infty} \phi_1^k L^k \right) \left(\sum_{l=0}^{\infty} \phi_2^l L^l \right)$$

using the Taylor expansion. Then, we can express Y_t as an MA(∞) process:

$$Y_t = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{r_1^k r_2^l} \varepsilon_{t-(k+l)} = \sum_{i=0}^{\infty} \sum_{k,l=0}^i \frac{1}{r_1^k r_2^l} \varepsilon_{t-i}$$

From the previous question, we know that the power spectrum of an MA(∞) process is given by $S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi if})|^{-2}$. Therefore, for our AR(2) process, we have

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

- **Simulation:** We first find expressions for ϕ_1 and ϕ_2 in terms of r and θ . The roots of the characteristic polynomial are given by $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$, with $\phi(r_1) = \phi(r_2) = 0$. By expanding the polynomial, we end up with $\phi_1 = \frac{r_1+r_2}{r_1 r_2}$ and $\phi_2 = -\frac{1}{r_1 r_2}$. Using $r_{1,2} = re^{\pm i\theta}$, we substitute the values of r and θ and find $\phi_1 = \frac{2\cos(2\pi/6)}{1.05} \approx 0.952$ and $\phi_2 = -\frac{1}{1.05^2} \approx -0.907$. We then simulate the AR(2) process with these parameters and plot the signal and its periodogram. Looking at the obtained plots, we observe the oscillations of the signal, which is consistent with the presence of complex roots. The periodogram shows a peak at the frequency $f = 0.17$ Hz, which indicates a dominant frequency component in the signal, corresponding to the oscillatory behavior induced by the complex roots. Using $\theta = \frac{2\pi}{6}$ implies a frequency f verifying $2\pi f = \frac{2\pi}{6}$, which gives $f = \frac{1}{6} \approx 0.167$ Hz, matching the observed peak in the periodogram.

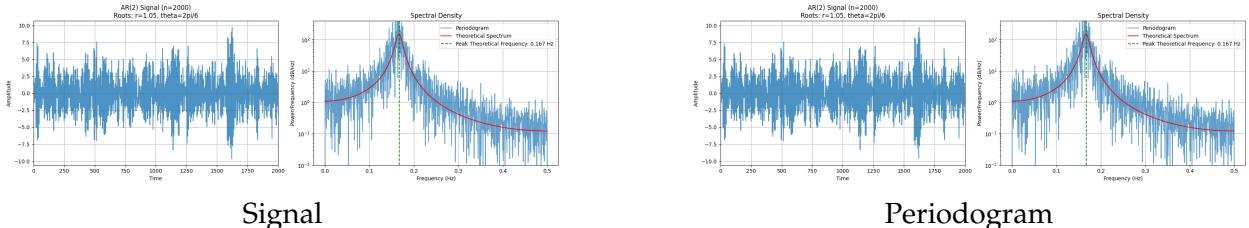


Figure 2: AR(2) process

4 Sparse coding

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

$$\forall u = 0, \dots, 2L - 1, \quad \phi_{L,k}[u] = w_L[u] \sqrt{\frac{2}{L}} \cos\left[\frac{\pi}{L} \left(u + \frac{L+1}{2}\right) \left(k + \frac{1}{2}\right)\right] \quad (3)$$

where w_L is a modulating window given by

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

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 MDCT 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

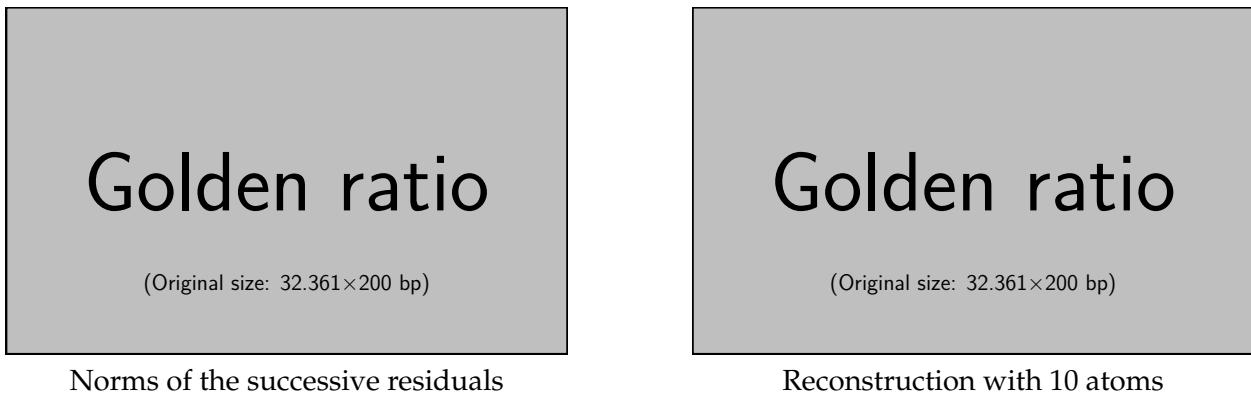


Figure 3: Question 4