Assignment 2 (ML for TS) - MVA

Samuel Lhayani samuel.lhayani@etu.minesparis.psl.eu
Hugo Pavy hugo.pavy@etu.minesparis.psl.eu

December 6, 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

Rate of convergence in the i.i.d. case. Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$. The sample mean is

$$\overline{X}n = \frac{1}{n} \sum i = 1^n X_i.$$

We have

$$\mathbb{E}\left[(\overline{X}_n - \mu)^2\right] = \mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Thus,

$$\overline{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right),$$

so the convergence rate is $1/\sqrt{n}$.

Wide-sense stationary case. Let $(Y_t)_{t\geq 1}$ be a wide-sense stationary process with mean μ and autocovariance function

$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t+k}),$$

and assume

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty.$$

Define the sample mean

$$\overline{Y}n = \frac{1}{n} \sum t = 1^n Y_t.$$

Variance bound. Using stationarity,

$$\mathrm{Var}(\overline{Y}_n) = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{s=1}^{n} \gamma(t-s) = \frac{1}{n^2} \sum_{k=-(n-1)}^{n-1} (n - |k|)\, \gamma(k). = \frac{1}{n} \sum_{k=-(n-1)}^{n-1} (1 - \frac{|k|}{n})\, \gamma(k)$$

Since $0 \leq 1 - \frac{|k|}{n} \leq 1$,

$$\mathrm{Var}(\overline{Y}_n) \leq \frac{1}{n} \sum_{k=-(n-1)}^{n-1} |\gamma(k)| \leq \frac{1}{n} \sum_{k=-\infty}^{\infty} |\gamma(k)|.$$

Let

$$C = \sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty.$$

Then

$$\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right] = \mathrm{Var}(\overline{Y}_n) \leq \frac{C}{n}.$$

Conclusion. Thus,

$$\overline{Y}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right),$$

which is the same rate of convergence as in the i.i.d. case.

Moreover, since

$$\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right] \to 0,$$

we have $\overline{Y}_n \to \mu$ in L^2, and therefore in probability. Thus, $\overline{Y}_n$ is a consistent estimator of μ.

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

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

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

where $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ are square–summable, $\sum_{k=0}^{\infty} \psi_k^2 < \infty$, and $\{\varepsilon_t\}_t$ is a zero–mean white noise with variance σ_ε^2.

(a) Mean and covariance, weak stationarity

Mean. Using linearity of expectation and $\mathbb{E}[\varepsilon_t] = 0$,

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

Hence the mean is constant in t and equal to 0.

Autocorrelation For any integer lag h,

$$\mathbb{E}[Y_t Y_{t-h}] = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right)\left(\sum_{\ell=0}^{\infty} \psi_\ell \varepsilon_{t-h-\ell}\right)\right].$$

Expanding the product and using Fubini's theorem (justified by square–summability),

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

Since $\{\varepsilon_t\}$ is white noise,

$$\mathbb{E}\left[\varepsilon_{t-k}\varepsilon_{t-h-\ell}\right] = \begin{cases} \sigma_\varepsilon^2, & \text{if } t-k = t-h-\ell \\ 0, & \text{otherwise.} \end{cases}$$

The condition $t-k = t-h-\ell$ is equivalent to $\ell = k-h$. Thus only terms with $\ell = k-h$ contribute:

$$\mathbb{E}[Y_t Y_{t-h}] = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h},$$

where we adopt the convention $\psi_j = 0$ for $j < 0$. Define the autocovariance function

$$\gamma(h) = \operatorname{Cov}(Y_t, Y_{t-h}) = \mathbb{E}[Y_t Y_{t-h}] - \mathbb{E}[Y_t]\mathbb{E}[Y_{t-h}] = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h}.$$

Because (ψ_k) is square–summable, the series defining $\gamma(h)$ is absolutely convergent and therefore finite for all h.

We see that $\mathbb{E}[Y_t] = 0$ for all t and that $\gamma(h)$ depends only on the lag h, not on t. Hence $\{Y_t\}$ is *weakly stationary*.

(b) Power spectrum

The power spectrum (spectral density) of a weakly stationary process is defined by

$$S(f) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i f h}, \qquad f \in [-\tfrac{1}{2}, \tfrac{1}{2}],$$

for a sampling frequency of 1 Hz.

Introduce the function

$$\phi(z) = \sum_{k=0}^{\infty} \psi_k z^k,$$

which is well-defined on the unit circle $|z| = 1$ by square–summability of ψ_k. Set $z = e^{-2\pi i f}$. Then

$$\phi(e^{-2\pi i f}) = \sum_{k=0}^{\infty} \psi_k e^{-2\pi i f k}.$$

Its complex conjugate is

$$\overline{\phi(e^{-2\pi i f})} = \sum_{j=0}^{\infty} \psi_j e^{2\pi i f j}.$$

Therefore

$$\begin{aligned} \left|\phi(e^{-2\pi i f})\right|^2 &= \phi(e^{-2\pi i f})\,\overline{\phi(e^{-2\pi i f})} \\ &= \left(\sum_{k=0}^{\infty} \psi_k e^{-2\pi i f k}\right)\left(\sum_{j=0}^{\infty} \psi_j e^{2\pi i f j}\right) \\ &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \psi_k \psi_j e^{-2\pi i f (k-j)}. \end{aligned}$$

Let $h = k - j$. Then

$$\left|\phi(e^{-2\pi i f})\right|^2 = \sum_{h=-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \psi_k \psi_{k-h} \right) e^{-2\pi i f h}.$$

But from above,

$$\gamma(h) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h}.$$

Hence

$$\left|\phi(e^{-2\pi i f})\right|^2 = \frac{1}{\sigma_\varepsilon^2} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i f h} = \frac{S(f)}{\sigma_\varepsilon^2}.$$

We conclude that

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

as required.

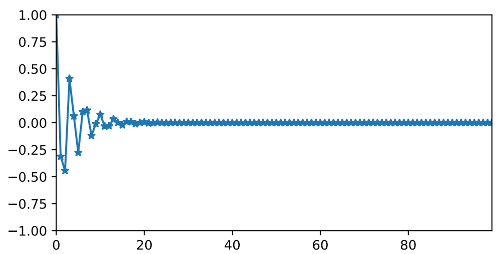
Question 3 *AR(2) process*

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

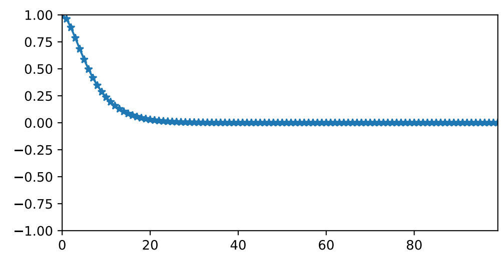
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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

- Let's compute $\gamma(\tau)$

$$\begin{aligned} \gamma(\tau) &= \mathbb{E}(Y_{t+\tau} Y_t) && (3) \\ &= \mathbb{E}((\phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2} + \epsilon_{t+\tau}) Y_t) && (4) \end{aligned}$$

Or since $t \neq t + \tau$, we have that $\epsilon_{t+\tau}$ is not correlated to Y_t

Thus,

$$\begin{aligned} \gamma(\tau) &= \mathbb{E}((\phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2}) Y_t) && (5) \\ &= \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2) && (6) \end{aligned}$$

This is a second-order linear recurrence. The space of all solutions is a vector space of dimension 2. If we find two solutions that are not proportionate, all the solutions are a linear combination of them.

Let's assume that $\gamma(\tau) = r^\tau$.

Thus, we have

$$\gamma(\tau) = \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2) \tag{7}$$
$$r^\tau = \phi_1 r^{\tau-1} + \phi_2 r^{\tau-2} \tag{8}$$
$$r^2 = \phi_1 r + \phi_2 \text{ (assuming that } r \neq 0) \tag{9}$$

Let's pose $r = \frac{1}{z}$

We have

$$1 - \phi_1 z - \phi_2 z^2 = 0 \tag{10}$$

Finally the autocovariance can be written as:

$$\gamma(\tau) = \frac{A}{r_1^\tau} + \frac{B}{r_2^\tau} \tag{11}$$

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 exists a unique $A, B \in \mathbb{R}$ such that for all τ:

$$\gamma(\tau) = \frac{1}{r^\tau}\big(A\cos(\tau\theta) + B\sin(\tau\theta)\big). \tag{12}$$

- Knowing that, it is obvious than the first figure with the oscillations is the one with complex roots and the other one has real roots.
- Let's use the lag operator L such that $LY_t = Y_{t-1}$. The properties of L give use that we can treat L "as if it was a polynomial". We can divide it, factorised by it...

  We have that

$$\epsilon_t = (1 - \phi_1 L - \phi_2 L^2)Y_t \tag{13}$$
$$= (1 - \frac{L}{r_1})(1 - \frac{L}{r_2})Y_t \tag{14}$$
$$Y_t = \frac{\epsilon_t}{(1 - \frac{L}{r_1})(1 - \frac{L}{r_2})} \tag{15}$$
$$\tag{16}$$

Because $|r_i| > 1$, let's assume that we will always have $|\frac{L}{r_i}| < 1$, we can use the geometric serie:

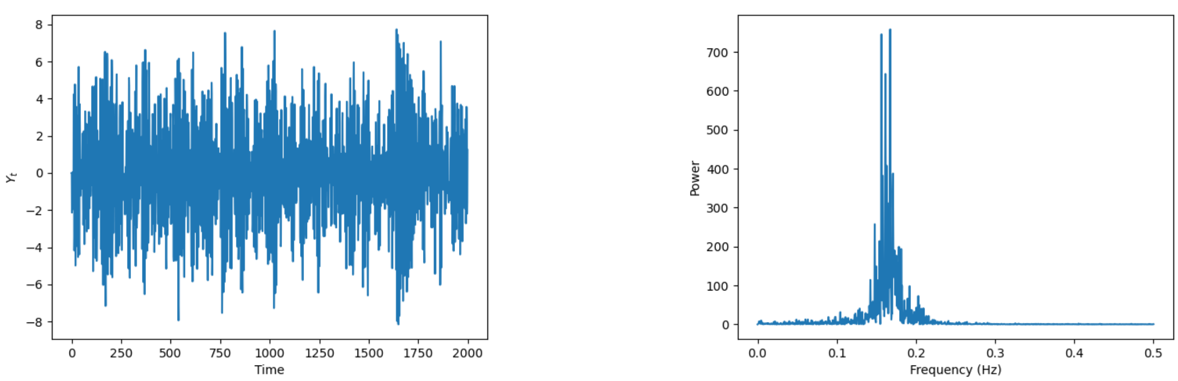
$$
\begin{aligned}
Y_t &= \left(\sum_{m=0}^{\infty} \frac{1}{r_1^m} L^m\right)\left(\sum_{n=0}^{\infty} \frac{1}{r_2^n} L^n\right)\epsilon_t && (17)\\
&= \left(\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{r_1^m r_2^n} L^{m+n}\right)\epsilon_t && (18)\\
&= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{r_1^m r_2^n} L^{m+n}\epsilon_t && (19)\\
&= \sum_{k=0}^{\infty} \underbrace{\sum_{i+j=k} \frac{1}{r_1^i r_2^j}}_{\theta_k} \epsilon_{t-k} && (20)\\
&= \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k} && (21)
\end{aligned}
$$

We finally have obtained an infinite order moving average MA(∞).

According to question 2, we know that: $S(f) = \sigma_\varepsilon^2 |\phi'(e^{-2\pi i f})|^2$. Where $\phi'(z) = \sum_j \theta_j z^j$ with $\theta_k = \sum_{i+j=k} \frac{1}{r_1^i r_2^j}$.
With same manipulations as before over the sum, we have $\phi = \phi'$

Finally,

$$S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2 \tag{22}$$



Signal Periodogram

Figure 2: AR(2) process

We can observe a pick in the peridodogram at $f \approx 0.17$Hz which makes sense according to the question 3 because this is the value of $\frac{\theta}{2\pi}$ and we have complex roots.

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

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{23}$$

where w_L is a modulating window given by

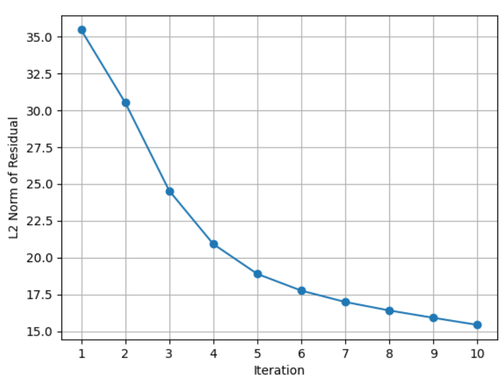
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{24}$$

Question 4 *Sparse coding with OMP*

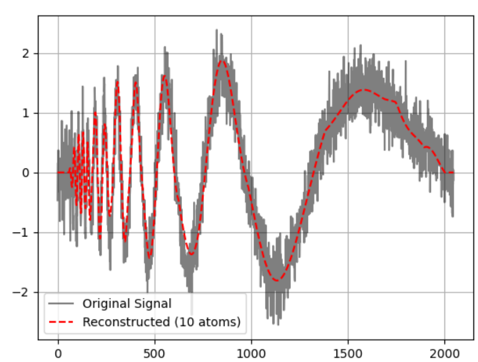
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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