Assignment 2 (ML for TS) - MVA

Samar RABEH samarrabeh.prepa@gmail.com
Hayder Hajlaoui Hayderhajlaoui.prepa@gmail.com

December 1, 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 Monday 2nd 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: https://docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2

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

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with:

- mean: $\mathbb{E}[X] = \mu$ that is finite,
- Finite variance: $\text{Var}(X) = \sigma^2$.

The sample mean is given by:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Law of Large Numbers (LLN)

The sample mean $\bar{X}_n$ is a consistent estimator of the true mean μ, as:

$$\bar{X}_n \xrightarrow{P} \mu \quad (\text{as } n \to \infty).$$

This is a convergence in probability

The Central Limit Theorem (CLT) states the rate of convergence for $\bar{X}_n$. It states that the distribution of the sample mean becomes close to a normal distribution when $n \to \infty$:

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

article amsmath, amssymb

We have : Let $\{Y_t\}_{t \geq 1}$ be a wide-sense stationary process with

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

Let's show that:

$$\bar{Y}_n = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

is consistent and has the same rate of convergence as the i.i.d. case.

By the definition

$$\mathbb{E}[Y_t] = \mu, \quad \text{for all } t,$$

and the autocovariance function is defined as:

$$\mathrm{Cov}(Y_t, Y_{t+k}) = \gamma(k), \quad \text{where } \gamma(k) \text{ depends only on the lag } k.$$

The goal is to show that:

$$\bar{Y}_n \xrightarrow{P} \mu.$$

By bounding the mean squared error (MSE):

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

which implies consistency if $\mathrm{Var}(\bar{Y}_n) \to 0$ as $n \to \infty$.

$$\mathrm{Var}(\bar{Y}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{t=1}^{n} Y_t\right).$$

Expanding the variance:

$$\mathrm{Var}\left(\sum_{t=1}^{n} Y_t\right) = \sum_{t=1}^{n} \mathrm{Var}(Y_t) + \sum_{t \neq s} \mathrm{Cov}(Y_t, Y_s).$$

Since Y_t is WSS:

$$\mathrm{Var}(Y_t) = \gamma(0), \quad \text{and} \quad \mathrm{Cov}(Y_t, Y_s) = \gamma(|t - s|).$$

Thus:

$$\mathrm{Var}\left(\sum_{t=1}^{n} Y_t\right) = n\gamma(0) + 2\sum_{1 \leq t < s \leq n} \gamma(|t - s|).$$

The covariance terms can be rewritten as:

$$\sum_{1 \leq t < s \leq n} \gamma(|t - s|) = \sum_{k=1}^{n-1} (n - k)\gamma(k),$$

where $k = |t - s|$.

Substituting back, the variance becomes:

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

$$\mathrm{Var}(\bar{Y}_n) = \frac{\gamma(0)}{n} + \frac{2}{n^2}\sum_{k=1}^{n-1} (n - k)\gamma(k).$$

For large n, the term $\frac{(n-k)}{n}$ approaches 1. :

$$\frac{2}{n^2}\sum_{k=1}^{n-1} (n - k)\gamma(k) \sim \frac{2}{n}\sum_{k=1}^{\infty} \gamma(k).$$

Thus:

$$\mathrm{Var}(\bar{Y}_n) \sim \frac{\gamma(0) + 2\sum_{k=1}^{\infty} \gamma(k)}{n},$$

The variance of $\bar{Y}_n$ decreases at the rate $\frac{1}{n}$, the same as in the i.i.d. case. Since convergence in L^2 implies convergence in probability:

$$\bar{Y}_n \xrightarrow{P} \mu.$$

The sample mean $\bar{Y}_n$ 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

The expectation of Y_t is given by

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

Since $\{\epsilon_t\}_{t\geq 0}$ is a white noise process with zero mean, i.e., $\mathbb{E}[\epsilon_t] = 0$ for all t, we have

$$\mathbb{E}[Y_t] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\epsilon_{t-k}] = \sum_{k=0}^{\infty} \psi_k \cdot 0 = 0.$$

Thus, $\mathbb{E}[Y_t] = 0$. The autocovariance is

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

Using the white noise property $\mathbb{E}[\epsilon_{t-i}\epsilon_{t-k-j}] = \sigma_\epsilon^2$ when $i = k + j$ and 0 otherwise, we obtain

$$\mathbb{E}[Y_t Y_{t-k}] = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

Thus, the autocovariance function is

$$\mathbb{E}[Y_t Y_{t-k}] = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

To determine if the process is weakly stationary, we need to check the following conditions:

- We already know that $\mathbb{E}[Y_t] = 0$, which is constant over time.
- **Constant Variance:** The variance of Y_t is

$$\text{Var}(Y_t) = \mathbb{E}[Y_t^2] = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2.$$

  Since $\sum_{k=0}^{\infty} \psi_k^2 < \infty$, the variance is finite and constant over time.
- **Autocovariance Depends Only on Lag:** We have already derived the autocovariance

$$\mathbb{E}[Y_t Y_{t-k}] = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

  This autocovariance depends only on the lag k, not the specific time t, since it involves sums over the coefficients ψ_i and ψ_{i+k}, and the process is stationary in this sense.

The process $\{Y_t\}_{t \geq 0}$ is **weakly stationary**

2/

We want to compute the power spectrum $S(f)$ of the process $\{Y_t\}$, given that the power spectrum of a stationary process is the Fourier transform of its autocovariance function.

We have already derived that the autocovariance function of the process $\{Y_t\}$ is:

$$E(Y_t Y_{t-k}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

Now, the power spectrum $S(f)$ is the Fourier transform of the autocovariance function $\gamma(k)$, i.e.,

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-2\pi i f k}.$$

Substituting $\gamma(k) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$, we get:

$$S(f) = \sum_{k=-\infty}^{\infty} \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} e^{-2\pi i f k}.$$

$$S(f) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i e^{-2\pi i f i} \sum_{k=-\infty}^{\infty} \psi_{i+k} e^{-2\pi i f k}.$$

Now, the second sum is the Fourier transform of the sequence ψ_k. We define:

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

Substituting $z = e^{-2\pi i f}$, we obtain:

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

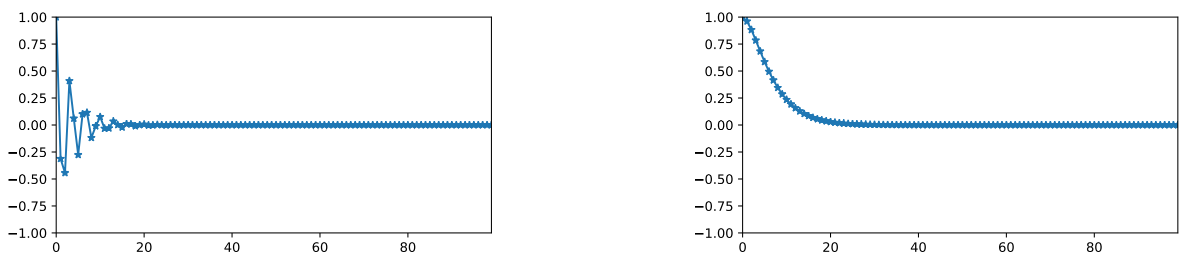
Question 3 *AR(2) process*

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

1. Autocovariance Coefficients $\gamma(k)$ Using Roots r_1 and r_2

An AR(2) process is defined as:

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

where ε_t is white noise with variance σ^2.

The autocovariance function $\gamma(k)$ satisfies the recurrence relation:

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2), \quad \forall k \geq 2.$$

The characteristic polynomial is:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2.$$

Assume $\phi(z)$ has two distinct roots r_1 and r_2 such that $|r_1| > 1$ and $|r_2| > 1$. The general solution for $\gamma(k)$ is:

$$\gamma(k) = \alpha r_1^{-k} + \beta r_2^{-k},$$

where α and β are determined by the initial conditions $\gamma(0)$ (variance) and $\gamma(1)$ (lag-1 autocovariance).

Step 1: Initial Conditions

- Variance $\gamma(0)$:

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2.$$

- Lag-1 Autocovariance $\gamma(1)$:

$$\gamma(1) = \frac{\phi_1\gamma(0)}{1-\phi_2}.$$

- Lag-2 Autocovariance $\gamma(2)$:

$$\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0).$$

Step 2: Variance $\gamma(0)$

By substituting $\gamma(1)$ and $\gamma(2)$ back into the equations, the variance $\gamma(0) = V$ is given by:

$$V = \frac{\sigma^2(1-\phi_2)}{(1+\phi_2)(1+\phi_1-\phi_2)(1-\phi_1-\phi_2)}.$$

Step 3: Coefficients α and β

From the initial conditions:

$$\begin{cases} \gamma(0) = \alpha + \beta = V, \\ \gamma(1) = \alpha r_1^{-1} + \beta r_2^{-1} = \frac{\phi_1 V}{1-\phi_2}. \end{cases}$$

Solving for α and β:

$$\alpha = \frac{\sigma^2 r_1^3 r_2^2}{(r_1r_2-1)(r_2-r_1)(r_1^2-1)}, \quad \beta = -\frac{\sigma^2 r_1^2 r_2^3}{(r_1r_2-1)(r_2-r_1)(r_2^2-1)}.$$

Thus, the autocovariance function is:

$$\gamma(k) = \frac{\sigma^2 r_1^2 r_2^2}{(r_1r_2-1)(r_2-r_1)}\left[\frac{r_1^{-k}}{r_1^2-1} - \frac{r_2^{-k}}{r_2^2-1}\right].$$

2. Identifying Complex and Real Roots from Correlograms

- Complex Roots: Oscillatory behavior is observed in the correlogram due to the complex terms in r_1^{-k} and r_2^{-k}, which include $\exp(i\theta)$.
- Real Roots: Monotonic exponential decay appears when roots are real.

Conclusion: The first correlogram corresponds to complex roots (oscillations), while the second corresponds to real roots (decay).

3. Power Spectrum $S(f)$

The power spectrum is:

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-2i\pi k f}.$$

For an AR(2) process:

$$S(f) = \frac{\sigma^2}{2\pi |\phi(e^{-2i\pi f})|^2},$$

where $\phi(e^{-2i\pi f}) = 1 - \phi_1 e^{-2i\pi f} - \phi_2 e^{-4i\pi f}$.

Expanding, the spectrum is:

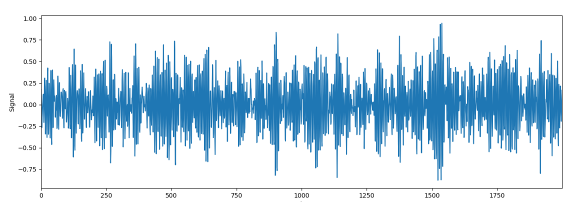
$$S(f) = \frac{\sigma^2}{2\pi \left[1 + \phi_1^2 + \phi_2^2 + 2\phi_1\phi_2 - 2\phi_1 \cos(2\pi f) - 4\phi_2 \cos^2(\pi f)\right]}.$$

4. Simulation with Complex Conjugate Roots

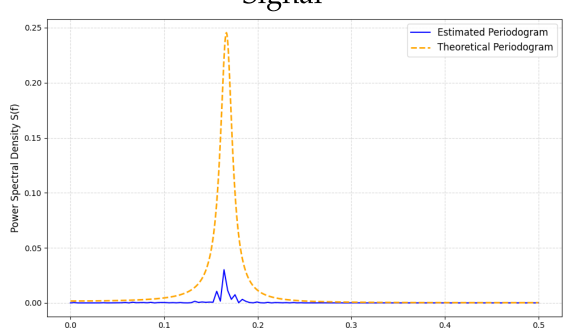
Let the roots have magnitude $r = 1.05$ and phase $\theta = \frac{2\pi}{6}$. Then:

$$\phi_1 = -2r\cos(\theta), \quad \phi_2 = r^2.$$

Simulating the process for $n = 2000$ samples produces:
- Oscillatory behavior in the time domain due to complex roots.
- A periodogram with distinct peaks at the frequencies corresponding to θ, reflecting spectral concentration.



Signal



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 ($k = 0, \ldots, L-1$) by

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

where w_L is a modulating window given by

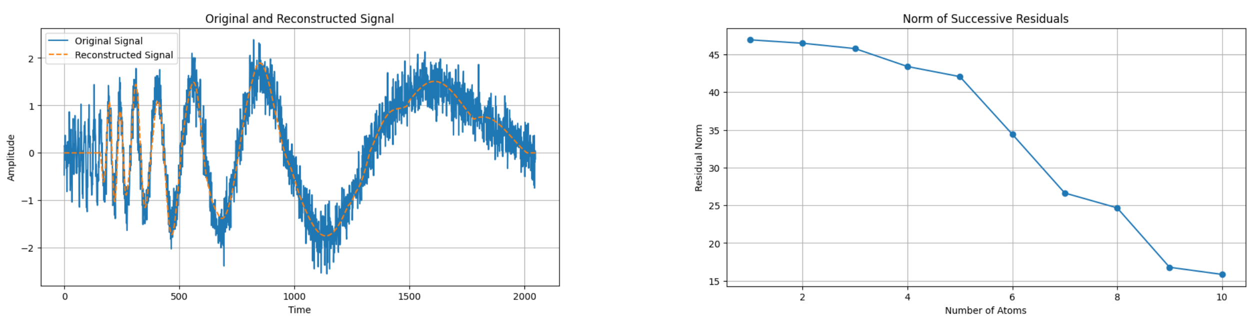
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

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