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

Denys Sikorskyi dsikora369@gmail.com
Alisa Kugusheva kugusheva.ad@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 27th 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

- According to the CLT if the samples $Y_1, ..., Y_n$ are i.i.d. with the mean μ and variance σ^2 then

$$\sqrt{n}\left(\frac{\hat{Y}_n - \mu}{\sigma}\right) \xrightarrow{\mathcal{D}} N(0,1)$$

  where $\hat{Y}_n = \frac{1}{n}(Y_1 + ... + Y_n)$ the sample mean. The convergence rate is therefore $\frac{1}{\sqrt{n}}$.

- $\{Y_t\}_{t\geq 1}$ is a wide-sense stationary process:

$$\forall t \mathbb{E}[Y_t] = \mu$$

$$\forall t_1, t_2 \mathbb{E}[Y_{t_1} Y_{t_2}] = \gamma(|t_2 - t_1|)$$

  Then

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \mathbb{E}[\bar{Y}_n^2] - \mathbb{E}[\bar{Y}_n]^2 = \mathbb{E}[\bar{Y}_n^2] - \mu^2 \leq \mathbb{E}[\bar{Y}_n^2] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{n}\sum_{j=1}^{n} Y_j] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[Y_i Y_j] =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(|i-j|) = \frac{1}{n^2}(n\gamma(0) + 2\sum_{k=1}^{n-1}(n-k)\gamma(k))$$

  In the last equality we replaced $i - j = k$ and considered the symmetry of $\gamma(k)$ to calculate the number of combinations for (i, j).
  Now let's bound this expression:

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

  We know that $\sum_k |\gamma(k)| < +\infty$, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\gamma(k) = 0$$

  And therefore

$$\lim_{n\to\infty}\mathbb{E}[(\bar{Y}_n - \mu)^2] = 0$$

We have established the convergence of the sample mean to the expectation in L_2 and since the convergence in L_2 implies the convergence in probability $\hat{Y}_n \xrightarrow{\mathcal{D}} \mu$ meaning that $\bar{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$ is a consistent estimator. The convergence rate of $\hat{Y}_n \xrightarrow{\mathcal{D}} \mu$ is the same as the convergence rate of $\frac{1}{\sqrt{n}}\sqrt{\sum_{k=0}^{n} \gamma(k)}$ which is equal to $\frac{1}{\sqrt{n}}$.

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

- $$\mathbb{E}[Y_t] = \mathbb{E}[\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \epsilon_{t-2} + ...] = \mathbb{E}[\epsilon_t] + \psi_1 \mathbb{E}[\varepsilon_{t-1}] + \psi_2 \mathbb{E}[\varepsilon_{t-2}] + ... = 0$$

  $$\mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}[\left(\sum_{l=0}^{\infty} \psi_l \varepsilon_{t-l}\right)\left(\sum_{p=0}^{\infty} \psi_p \varepsilon_{t-p-k}\right)] = \mathbb{E}[\sum_{l=0}^{\infty}\sum_{p=0}^{\infty} \psi_l \psi_p \varepsilon_{t-l} \varepsilon_{t-p-k}] = \sum_{l=0}^{\infty}\sum_{p=0}^{\infty} \psi_l \psi_p \mathbb{E}[\varepsilon_{t-l} \varepsilon_{t-p-k}]$$

  For the white noise if $i = j$:

  $$\mathbb{E}[\varepsilon_i \varepsilon_j] = \sigma_\varepsilon^2$$

  if $i \neq j$

  $$\mathbb{E}[\varepsilon_i \varepsilon_j] = 0$$

  Then defining $k + p = l$

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

  To define if the process is weakly stationary we have to check the properties:

  1. The mean is constant: $\mathbb{E}[Y_t] = 0 \ \ \forall t$.
  2. The variance is finite: $\mathbb{E}[Y_t^2] = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_l^2 < \infty$.
  3. The covariance only depends on k: $Cov(Y_t, Y_{t-k}) = \mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_{t-k} - \mathbb{E}[Y_{t-k}])] = \mathbb{E}[Y_t Y_{t-k}] = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_l \psi_{l-k} = \gamma(k)$. All conditions are satisfied, hence the process is weakly stationary.

- The power spectrum is defined as

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

where $\gamma(k)$ is the autocovariance defined in the previous computations. We can define the autocovariance-generating function as

$$g(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k = \sigma_{\varepsilon^2} \sum_{k=-\infty}^{\infty} \sum_{l=0}^{\infty} \psi_l \psi_{l-k} z^k = \sigma_{\varepsilon^2} \sum_{l=0}^{\infty} \sum_{p=0}^{\infty} \psi_l \psi_p z^{p-l} = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_l z^{-l} \sum_{p=0}^{\infty} \psi_p z^p = \sigma_\varepsilon^2 \phi(z^{-1}) \phi(z)$$

Then we can notice that

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-2\pi i f k} = g(e^{-2\pi i f}) = \sigma_\varepsilon^2 \phi(e^{-2\pi i f}) \phi(e^{2\pi i f}) = \sigma_{\varepsilon^2} |\phi(e^{2\pi i f}|^2$$

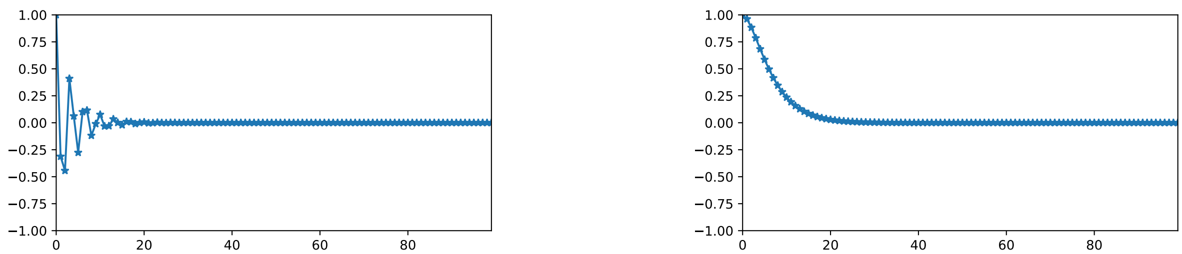
Question 3 *AR(2) process*

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 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

Autocovariance coefficients $\gamma(\tau)$: The autocovariance coefficients of the AR(2) process can be expressed in terms of its characteristic roots r_1 and r_2 as:

$$\begin{aligned}
\gamma(0) &= \frac{\sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2} \\
\gamma(1) &= \frac{\phi_1 \sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2}(r_1 + r_2) \\
\gamma(2) &= \frac{\phi_2 \sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2}(r_1^2 + r_2^2) + \frac{\phi_1 \sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2}(\phi_1 + r_1 r_2) \\
\gamma(\tau) &= \frac{\phi_1 \sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2}\left(r_1^{\tau-1} + r_2^{\tau-1}\right) + \frac{\phi_2 \sigma^2}{|1-\phi_1 r_1 - \phi_2 r_2|^2}\left(r_1^{\tau-2} + r_2^{\tau-2}\right), \quad \text{for } \tau > 2.
\end{aligned}$$

The parameter σ represents the standard deviation of the process noise ε_t.
An AR(2) process' correlogram can reveal details regarding the nature of its distinctive origins. A

damped sinusoidal pattern will be visible on the correlogram if the roots are actual and distinct. The correlogram will display an oscillating pattern that does not dampen over time if the roots are complex conjugates. As a result, in Figure 1, the correlogram on the left represents an AR(2) process with real roots, whereas the correlogram on the right represents an AR(2) process with complex conjugate roots.

S(f) = $1\frac{}{2\pi\left|\phi(e^{-i2\pi f})\right|^{-2}}$

where

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

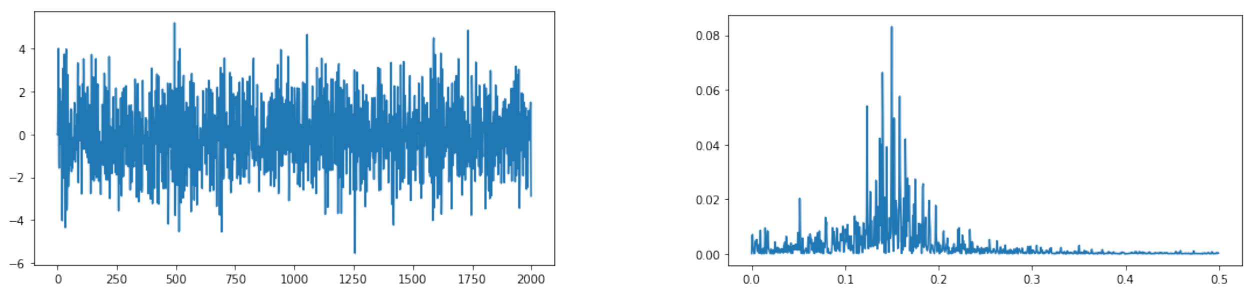
and f is the frequency. If we assume a sampling frequency of 1 Hz, we can write the formula as:

S(f) = $1\frac{}{2\pi\left|\phi(e^{-i2\pi f/1\,\mathrm{Hz}})\right|^{-2}}$

where f is measured in Hz. The spectral density function is given by:

$\left|\phi(e^{-i2\pi f/1\,\mathrm{Hz}})\right| = \left|1 - \phi_1 e^{-i2\pi f/1\,\mathrm{Hz}} - \phi_2 e^{-i4\pi f/1\,\mathrm{Hz}}\right|$

Substituting this expression into the power spectrum formula gives us the power spectrum of the AR(2) process as a function of frequency.



Signal Periodogram

Figure 2: AR(2) process

As we should we observe one big spike on the periodogram.

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

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

where w_L is a modulating window given by

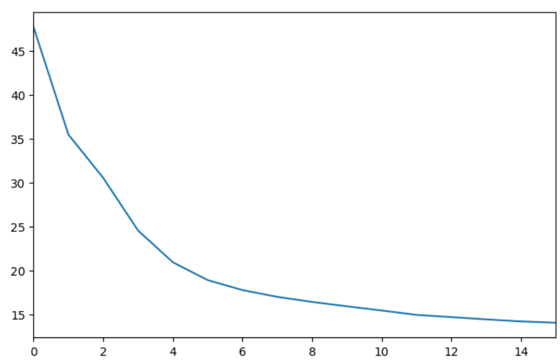
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

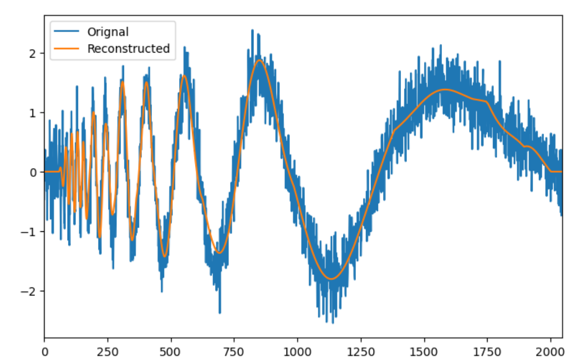
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

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

The error decreases steadily and monotonically, so there are no problems with the algorithms. As we can see the reconstructed signal with 10 atoms almost perfectly matches the original one.