Assignment 2 (ML for TS) - MVA 2023/2024

Gauthier Multari gauthier.multari@ens-paris-saclay.fr

December 3, 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 Tuesday 5$^{\text{th}}$ 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: docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2cBHQM

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

- Let $(Y_n)_{n\leq 1}$ a sequence of i.i.d variables of variance σ^2 and $\mu = \mathbb{E}[Y_i]$.

  $$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

  We know that $\mathbb{E}[\bar{Y}_n] = \mu$ and $Var(\bar{Y}_n) = \frac{\sigma^2}{n}$. By using the Bienaymé-Tchebychev inequality with $\epsilon > 0$ we have:

  $$\mathbb{P}(|\bar{Y}_n - \mu| \geq \epsilon) \leq \frac{2\sigma^2}{n\epsilon^2} \underset{n\to+\infty}{\longrightarrow} 0$$

  Thus $\bar{Y}_n \longrightarrow \mu$ with a rate of $\frac{1}{n}$
- We want to prove the L_2 convergence :

  $$\begin{aligned}
  \mathbb{E}[(\bar{Y}_n - \mu)^2] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)\right)\left(\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu)\right)\right] \\
  &= \frac{1}{n^2}\left(2\sum_{i=1}^{n}\sum_{j=i}^{n}\mathbb{E}[(Y_i - \mu)(Y_j - \mu)] + \sum_{i=1}^{n}\mathbb{E}[(Y_i - \mu)^2]\right) \\
  &= \frac{1}{n^2}\left(2\sum_{i=1}^{n}\sum_{k=1}^{n-i}\gamma(k) + n\gamma(0)\right) \\
  &= \frac{2}{n^2}\sum_{i=1}^{n}(n-i)\gamma(i) + \frac{1}{n}\gamma(0) \\
  &\leq \frac{2}{n}\sum_{i=0}^{n}\gamma(k) \\
  &\leq \frac{2}{n}\sum_{i=0}^{+\infty}|\gamma(k)|
  \end{aligned}$$

  Let C > 0 s.t C = $\sum_{i=0}^{+\infty}$. Then : $\mathbb{E}[(\bar{Y}_n - \mu)^2] \leq \frac{2C}{n} \underset{n\to+\infty}{\longrightarrow} 0$.

  So $\bar{Y}_n \xrightarrow[n\to+\infty]{L_2} \mu$. Thus $\xrightarrow[n\to+\infty]{\mathbb{P}} \mu$.

We once again use the Bienaymé-Tchebychev inequality to find the rate of consistency. Let $\epsilon > 0$:

$$\begin{aligned}
\mathbb{P}(|\bar{Y}_n - \mu| \geq \epsilon) &\leq \frac{2\mathbb{E}[(\bar{Y}_n - \mu)^2]}{\epsilon^2} \\
&\leq \frac{4C}{n\epsilon^2} \underset{n \to +\infty}{\longrightarrow} 0
\end{aligned}$$

So $\bar{Y}_n \xrightarrow[n \to +\infty]{\mathbb{P}} \mu$ with a rate of $\frac{1}{n}$.

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

- $$\begin{aligned}
\mathbb{E}(Y_t) &= \sum_{k=0}^{+\infty} \mathbb{E}[\psi_k \epsilon_{t-k}] \\
&= \sum_{k=0}^{+\infty} \psi_k \mathbb{E}[\epsilon_{t-k}] \qquad \text{(Because } \psi_k \text{ is deterministic)} \\
&= 0 \qquad \text{(Because } \epsilon_k \text{ is a zero mean white noise)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(Y_t Y_{t-k}) &= \mathbb{E}\left[\left(\sum_{i=0}^{+\infty} \psi_i \epsilon_{t-i}\right)\left(\sum_{j=0}^{+\infty} \psi_j \epsilon_{t+k-j}\right)\right] \\
&= \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \psi_i \psi_j \mathbb{E}[\epsilon_{t-i}\epsilon_{t+k-j}] \qquad (\mathbb{E}[\epsilon_{t-i}\epsilon_{t+k-j}] = 0 \text{ if } t-i \neq t+k-j) \\
&= \sigma_\epsilon^2 \gamma(k)
\end{aligned}$$

We see that $\mathbb{E}(Y_t Y_{t-k})$ only depends on k, so Y_t is weakly stationary.

- First, with $N \in \mathbb{N}$:

$$\begin{aligned}
\left|\sum_{j=0}^{N} \psi_j e^{-2i\pi f j}\right|^2 &= \left(\sum_{j=0}^{N} \psi_j e^{-2i\pi f j}\right)\left(\sum_{l=0}^{N} \psi_l e^{2i\pi f l}\right) \\
&= \sum_{j=0}^{N}\sum_{l=0}^{N} \psi_j \psi_l e^{-2i\pi f(j-l)} \\
&= \sum_{\tau=-N+1}^{N-1} \sum_{n=0}^{N-\tau-1} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau} \quad \text{(like in Assignment 1)}
\end{aligned}$$

When $N \longrightarrow +\infty$ we get:

$$\left|\phi\left(e^{-2\pi i f}\right)\right|^2 = \sum_{\tau=-\infty}^{+\infty} \sum_{n=0}^{+\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau}$$

Finally:

$$\begin{aligned}
S(f) &= \sum_{\tau=-\infty}^{\tau=+\infty} \gamma(\tau) e^{-2i\pi f \tau} \\
&= \sigma_\epsilon^2 \sum_{\tau=-\infty}^{\tau=+\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau} \\
&= \sigma_\epsilon^2 \left|\phi\left(e^{-2\pi i f}\right)\right|^2
\end{aligned}$$

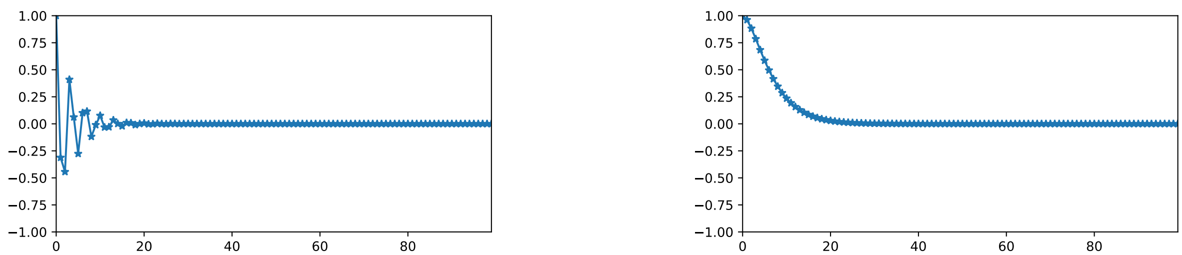
Question 3 *AR(2) process*

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 **??**. 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. To compute the covariance we suppose that $\mathbb{E}[\epsilon] = 0$:

$$\begin{aligned}\gamma(\tau) &= \mathbb{E}\left[Y_t Y_{t+\tau}\right] \\ &= \mathbb{E}\left[Y_t \phi_1 Y_{t+\tau-1} + Y_t \phi_2 Y_{t+\tau-2} + Y_t \epsilon_t\right] \\ &= \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2)\end{aligned}$$

  γ is solution to the characteristic polynomial, so:

  1. if $r_1, r_2 \in \mathbb{R}$, then it exists a unique $\lambda, \mu \in \mathbb{R}$ such that for all τ :

  $$\gamma(\tau) = \frac{\lambda}{r_1^\tau} + \frac{\mu}{r_2^\tau}$$

  2. if $r_1, r_2 \in \mathbb{C}$, then it exits a unique $\lambda, \mu \in \mathbb{R}$ such that for all τ :

  $$\gamma(\tau) = \frac{1}{r^\tau}(\lambda \cos(\tau\theta) + \mu \sin(\tau\theta))$$

- Correlograms. The correlogram on the left represents an oscillating system fading to zero, therefore it corresponds to the complex roots. The one on the right corresponds to the real roots.

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

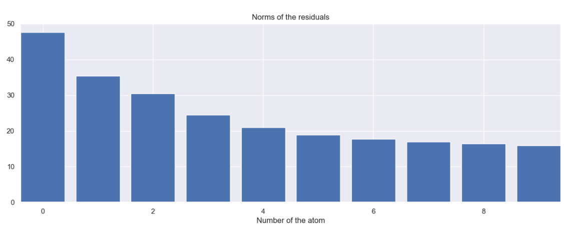
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

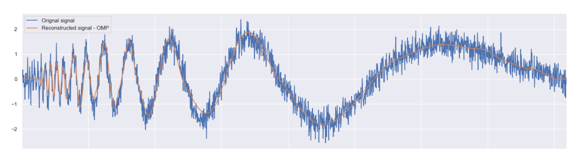
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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 2: Question 4