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

Louis-Marie Lovichi lm.lovichi@me.com
Benjamin Dahan Monsonego benjamin.dahan_monsonego@ens-paris-saclay.fr

December 5, 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 5th 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

- Puisque les variables aléatoires sont i.i.d. et de variance finie, on a bien la consistance de l'estimateur $\hat{\theta}_n$ d'après la loi faible des grands nombres. Par ailleurs, cette convergence a lieu en $O(\frac{1}{\sqrt{n}})$.
- Notons alors $Z_i = Y_i - \mu$. On est alors ramenés à estimer $\mathbb{E}(\overline{Z}_n^2)$. En utilisant l'hypothèse de stationnarité et sachant qu'il existe $n-k$ couples (i,j) tels que $1 \leq i < j \leq n$ et $j - i = k$, on obtient :

$$\begin{aligned}
\mathbb{E}(\overline{Z}_n^2) &= \frac{1}{n^2}\mathbb{E}(\sum_{i=1}^{n} Z_i^2 + 2\sum_{1\leq i<j\leq n} Z_i Z_j) \\
&= \frac{1}{n}\gamma(0) + \frac{2}{n^2}\sum_{1\leq i<j\leq n}\gamma(j-i) \\
&= \frac{1}{n}\gamma(0) + \frac{2}{n}\sum_{k=1}^{n-1}(1-\frac{k}{n})\gamma(k) \\
&\leq \frac{2}{n}\sum_{k=1}^{n-1}|\gamma(k)|
\end{aligned}$$

Du fait que $\sum_k |\gamma(k)| < +\infty$, on obtient que $\mathbb{E}(\overline{Z}_n^2) = O(\frac{1}{n})$. Donc la moyenne empirique converge dans L_2 avec une convergence en $O(\frac{1}{\sqrt{n}})$, ce qui implique la convergence en probabilité et donc la consistance de l'estimateur comme dans le cas i.i.d. ce qui conclut la preuve.

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

- Par les théorèmes de convegence usuels étant donné $\sum_k \psi_k^2 < \infty$, on calcule $\mathbb{E}(Y_t) = \sum_{k=0}^{\infty} \psi_k^2$. De même, on a par les théorèmes de convergence usuels et sachant que $t - j = t - k - i$ si et seulement si $j = k + i$, on obtient :

$$\begin{aligned}
Y_t Y_{t-k} &= \sum_{1\leq i<j\leq n} \psi_i \psi_j \varepsilon_{t-j} \varepsilon_{t-k-i} \\
\mathbb{E}(Y_t Y_{t-k}) &= \sum_{1\leq i<j\leq n} \psi_i \psi_j \mathbb{E}(\varepsilon_{t-j} \varepsilon_{t-k-i})
\end{aligned}$$

  D'où $\mathbb{E}(Y_t Y_{t-k}) = \sum_{i=0}^{\infty} \psi_i \psi_{k+i} \sigma_\varepsilon^2$ et on en conclut que le processus MA(∞) est bien faiblement stationnaire.

- Étant donné que la fréquence d'échantillonnage est de 1 Hz, on peut écrire par le théorème de Fubini et avec le changement de variable $u = k + j$:

$$\begin{aligned}
S(f) &= \sum_{k=-\infty}^{+\infty} \gamma(k) \exp(-2i\pi f k) \\
&= \sigma_\varepsilon^2 \sum_{k=-\infty}^{+\infty} \sum_{j=0}^{+\infty} \psi_j \exp(2i\pi f j) \psi_{k+j} \exp(-2i\pi f(k+j)) \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \psi_j \exp(2i\pi f j) \sum_{u=0}^{+\infty} \psi_u \exp(-2i\pi f u) \\
&= \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2
\end{aligned}$$

  D'où le résultat demandé.

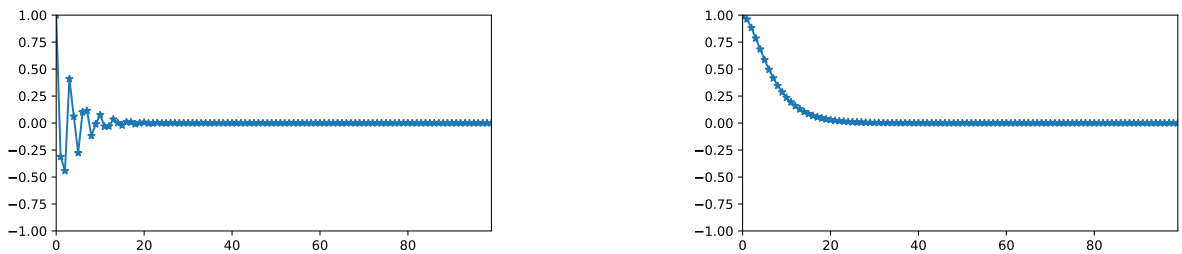
Question 3 *AR(2) process*

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

- Dans la mesure où les racines sont de module > 1, le processus est stationnaire. En multipliant par Y_{t-k}, on a :

$$\begin{aligned}
Y_t Y_{t-k} &= \phi_1 Y_{t-1} T_{t-k} + \phi_2 Y_{t-2} Y_{t-k} + \varepsilon_t Y_{t-k} \\
\mathbb{E}(Y_t Y_{t-k}) &= \phi_1 \mathbb{E}(Y_{t-1} T_{t-k}) + \phi_2 \mathbb{E}(Y_{t-2} Y_{t-k}) + \mathbb{E}(\varepsilon_t Y_{t-k}) \\
\mathbb{E}(Y_t Y_{t-k}) &= \phi_1 \mathbb{E}(Y_{t-1} T_{t-k}) + \phi_2 \mathbb{E}(Y_{t-2} Y_{t-k}) + \mathbb{E}(\varepsilon_t) \mathbb{E}(Y_{t-k}) \\
\gamma(k) &= \phi_1 \gamma(k) + \phi_2 \gamma(k-2)
\end{aligned}$$

  Donc $\gamma(k)$ est une suite récurrente d'ordre 2, et son expression générale s'écrit alors :

$$\gamma(k) = a \frac{1}{r_1^k} + b \frac{1}{r_2^k}$$

On détermine les coefficients a et b comme vérifiant $a + b = \gamma(0)$ et $\frac{a}{r_1} + \frac{b}{r_2} = \gamma(1)$, d'où :

$$\gamma(k) = \frac{1}{1/r_2 - 1/r_1} \left(\frac{\gamma(0)/r_2 - \gamma(1)}{r_1^k} + \frac{\gamma(1) - \gamma(0)/r1}{r_2^k} \right)$$

- Les oscillations du premier corrélogramme sont associées à des racines complexes tandis que le second corrélogramme à décroissance exponentielle stricte est associé à des racines réelles.

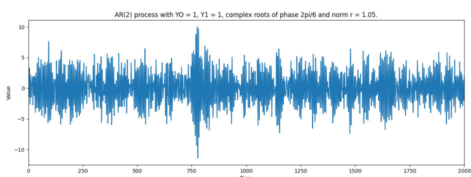
- On peut écrire :

$$Y_t = \sum_{i=0}^{t} a_i \varepsilon_{t-i}$$

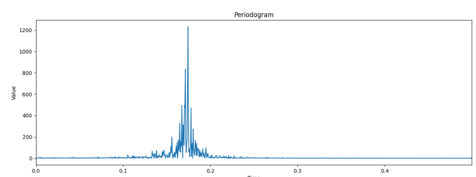
où $a_i = \sum_{j,k \geq 0, j+2k=i} \phi_1^j \phi_2^k$. Il s'agit des coefficients de la série $\sum_{i=0}^{+\infty} (\phi_1 z + \phi_2 z^2)^i = \frac{1}{1 - \phi_1 z - \phi_2 z^2} = \frac{1}{\phi(z)}$, d'où en utilisant les résultats de la question précédentes, on a :

$$S(f) = \frac{\sigma_\varepsilon^2}{|\phi(z)|^2}$$

- Nous observons que le signal présente des oscillations périodiques ce qui est confirmé par le périodogramme.



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 a 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

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

where w_L is a modulating window given by

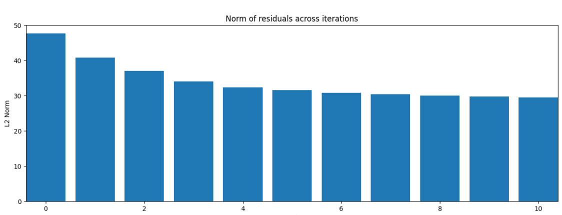
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

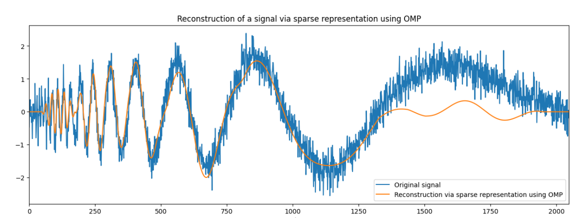
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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