

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

Denis Duval denis.duval13@gmail.com
Dorian Gailhard dorian.gailhard@telecom-paris.fr

June 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 (Ω, \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 $\bar{Y}_n = (Y_1 + \dots + Y_n)/n$ 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

- From Bienaymé-Tchebychev, the convergence rate is at least $\frac{1}{n}$:

$$\mathbb{P}(|S_n - \mathbb{E}[Y_1]| > \epsilon) \leq \frac{V(S_n)}{\epsilon^2} = \frac{V(Y_1)}{n\epsilon^2}$$

•

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

We do the change of variables $k = j - i$:

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

Whether k is negative or positive there are $n - |k|$ terms in the sum:

$$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \frac{1}{n} \sum_{k=-(n-1)}^{(n-1)} \left(1 - \frac{|k|}{n}\right) \gamma(k) \leq \frac{1}{n} \sum_{k \in \mathbb{Z}} |\gamma(k)|$$

Hence using Markov inequality :

$$\boxed{\mathbb{P}(|\bar{Y}_n - \mathbb{E}[Y_1]| > \epsilon) \leq \frac{\sum_{k \in \mathbb{Z}} |\gamma(k)|}{n\epsilon^2}}$$

3 AR and MA processes

Question 2 Infinite order moving average $MA(\infty)$

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

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

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

- As we have an infinite sum, we first need to prove that we can invert the infinite sum and the expectation.

Going back to the partial sums $S_N(t) = \sum_{k=0}^N \psi_k \varepsilon_{t-k}$, the convergence in L_2 is defined as $\mathbb{E}[(S_N(t) - Y_t)^2] \rightarrow 0$. So we write :

$$|\mathbb{E}[Y_t] - 0| = |\mathbb{E}[Y_t - S_N(t)]| \leq \sqrt{\mathbb{E}[(Y_t - S_N(t))^2]} \rightarrow 0$$

The first equality comes from the fact that the noise has zero mean so the partial sums have also zero mean. The second inequality comes from the Cauchy-Schwarz inequality. Hence

$$\boxed{\mathbb{E}[Y_t] = 0}.$$

Moreover. If $x_n \rightarrow x$ and $x'_n \rightarrow x'$ we should write :

$$x_n x'_n - x x' = (x_n - x)(x'_n - x') + x'(x_n - x) + x(x'_n - x')$$

So if we take $x_N = S_N(t)$, $x'_N = S_N(t-k)$ and $x = Y_t$, $x' = Y_{t-k}$ and then take the expectation:

$$\begin{aligned} \mathbb{E}[S_N(t)S_N(t-k) - Y_t Y_{t-k}] &= \mathbb{E}[(S_N(t) - Y_t)(S_N(t-k) - Y_{t-k})] + \mathbb{E}[Y_{t-k}(S_N(t) - Y_t)] \\ &\quad + \mathbb{E}[Y_t(S_N(t-k) - Y_{t-k})] \end{aligned}$$

Using C-S inequality on the three terms leads to :

$$\mathbb{E}[Y_t Y_{t-k}] = \lim_N \mathbb{E}[S_N(t)S_N(t-k)]$$

$$\mathbb{E}[Y_t Y_{t-k}] = \lim_N \sum_{n=0}^N \sum_{m=0}^N \psi_n \psi_m \mathbb{E}[\epsilon_{t-n} \epsilon_{t-k-m}]$$

The expectation in the sum is zero if we don't have $t - n = t - k - m$ ie $m = n - k$ (only possible if $n \geq k$). So :

$$\mathbb{E}[Y_t Y_{t-k}] = \sigma_\epsilon^2 \sum_{n=\max(0,k)}^{\infty} \psi_n \psi_{n-k}$$

Because the above quantities don't depend on t , the process is weakly stationary.

$$\begin{aligned} \bullet S(f) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2i\pi f \tau / f_s} = \sum_{\tau=-\infty}^{+\infty} \mathbb{E}[Y_t Y_{t+\tau}] e^{-2i\pi f \tau} \\ &= \sum_{\tau=-\infty}^{+\infty} \mathbb{E}[Y_t Y_{t-\tau}] e^{2i\pi f \tau} \\ &= \sum_{\tau=-\infty}^{+\infty} \sigma_\epsilon^2 \sum_{n=\max(0,\tau)}^{\infty} \psi_n \psi_{n-\tau} e^{2i\pi f \tau} \\ &= \sigma_\epsilon^2 \sum_{n=0}^{\infty} \sum_{\tau=-\infty}^n \psi_n \psi_{n-\tau} e^{2i\pi f \tau} \\ &= \sigma_\epsilon^2 \sum_{n=0}^{\infty} \sum_{\tau=-\infty}^n \psi_n (e^{2i\pi f})^n \psi_{n-\tau} (e^{2i\pi f})^{\tau-n} \end{aligned}$$

By changing index $k = n - \tau$ in the second sum :

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

$$\text{So } \boxed{S(f) = \sigma_\epsilon^2 |\phi(e^{-2i\pi f})|^2}$$

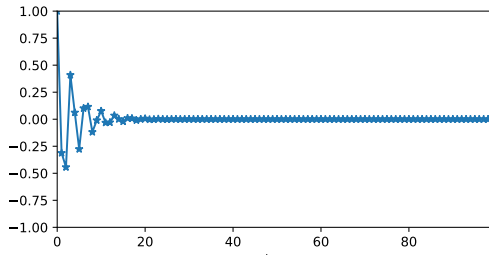
Question 3 AR(2) process

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

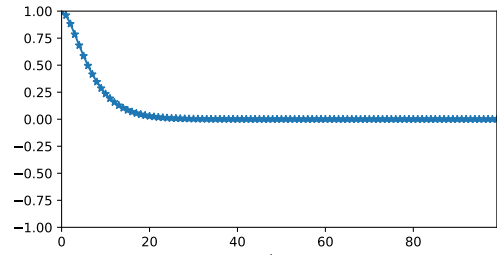
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad (2)$$

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

- Here we are assuming that the roots are outside the unit circle, which means that the process is weakly stationary. This implies also that the mean is zero. We have :

$$\gamma(\tau) = \mathbb{E}[Y_t Y_{t+\tau}] = \mathbb{E}[Y_t (\phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2} + \varepsilon_{t+\tau})]$$

- For $\tau = 0$, $\gamma(0) = \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + \sigma^2$
- For $\tau \geq 1$, $\gamma(\tau) - \phi_1 \gamma(\tau-1) - \phi_2 \gamma(\tau-2) = 0$

Hence the sequence can be solved with the roots of the characteristic polynomial $z^2 \phi(z^{-1})$, first for $\tau \geq 0$ and then for $\tau < 0$ by symmetrization :

$$\gamma(\tau) = a_1 \left(\frac{1}{r_1}\right)^{|\tau|} + a_2 \left(\frac{1}{r_2}\right)^{|\tau|}$$

We then need two initial conditions to find the parameters a_1 and a_2 .

- The first correlogram shows oscillations, so the roots are complex (they combine in an amortized cosinus), while the second is just an exponential decrease so the roots are real.

- Let's define the operator $L : Y_t \rightarrow Y_{t-1}$. Using the recurrence formula above we have :

$$\phi(L)Y_t = \epsilon_t$$

Then because $\phi(z) = (z - r_1)(z - r_2) = r_1 r_2 (1 - \frac{z}{r_1})(1 - \frac{z}{r_2})$, we have that :

$$\phi^{-1}(z) = \frac{1}{r_1 r_2 (1 - \frac{z}{r_1})(1 - \frac{z}{r_2})} = \frac{1}{r_1 r_2} \left(\sum_{i=0}^{+\infty} \left(\frac{z}{r_1} \right)^i \right) \left(\sum_{j=0}^{+\infty} \left(\frac{z}{r_2} \right)^j \right) \text{ for } z < \min(r_1, r_2)$$

That way we can invert the equality and get :

$$Y_t = \phi^{-1}(L)\phi(L)Y_t = \phi^{-1}(L)\epsilon_t = \frac{1}{r_1 r_2} \left(\sum_{i=0}^{+\infty} \frac{L^i}{(r_1)^i} \right) \left(\sum_{j=0}^{+\infty} \frac{L^j}{(r_2)^j} \right) \epsilon_t = \frac{1}{r_1 r_2} \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \frac{\epsilon_{t-i-j}}{(r_1)^i (r_2)^j}$$

Using the formula for the power spectrum for an MA(∞) process, we have ;

$$S(f) = \sigma^2 \left| \frac{1}{r_1 r_2} \left(\sum_{k=0}^{+\infty} \left(\frac{e^{2i\pi f}}{r_1} \right)^k \right) \left(\sum_{j=0}^{+\infty} \left(\frac{e^{2i\pi f}}{r_2} \right)^j \right) \right|^2 = \frac{\sigma^2}{|\phi(e^{2i\pi f})|^2}$$

- $\phi = 1 - \phi_1 X - \phi_2 X^2$ has the same roots as $X^2 + \frac{\phi_1}{\phi_2} X - \frac{1}{\phi_2}$.

Using the roots-coefficients relations, we have that :

$$\begin{cases} -\frac{\phi_1}{\phi_2} = r_1 + r_2 = 2 \times 1.05 \cos\left(\frac{2\pi}{6}\right) \\ -\frac{1}{\phi_2} = r_1 r_2 = 1.05^2 \end{cases} \Leftrightarrow \begin{cases} \phi_1 = \frac{2.1}{1.05^2} \cos\left(\frac{2\pi}{6}\right) \\ \phi_2 = -1.05^{-2} \end{cases}$$

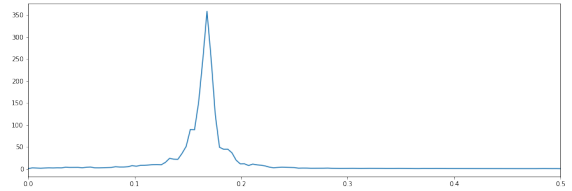
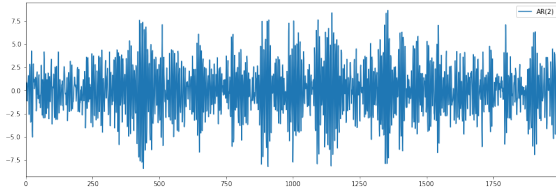


Figure 2: AR(2) process

- We can observe the stationarity of the process as well as a non negligible variance. Furthermore we can observe a frequency at about 0.16 of the sampling frequency.

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

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

where w_L is a modulating window given by

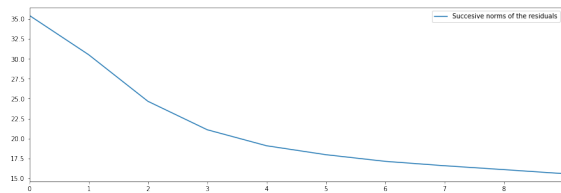
$$w_L[u] = \sin\left[\frac{\pi}{2L} \left(u + \frac{1}{2}\right)\right]. \quad (4)$$

Question 4 *Sparse coding with OMP*

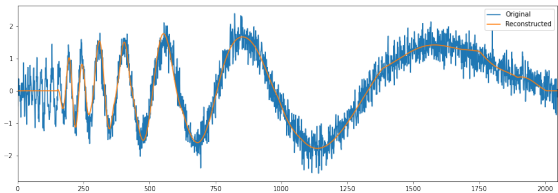
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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