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

Simon Queric simon.queric@telecom-paris.fr
Vincent Herfeld vincent.herfeld@telecom-paris.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

Let $(X_n)_n$ be a sequence of i.i.d random variables of mean m and variance σ^2. Markov's inequality gives us

$$\begin{aligned}
\mathbb{P}(|S_n - m| \geqslant \varepsilon) &\leqslant \frac{\mathbb{E}[(S_n - m)^2]}{\varepsilon^2} \\
&= \frac{\frac{1}{n^2} Var(\sum_{k=1}^n X_k - m)}{\varepsilon^2}, \text{ since } \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^n X_k - m\right] = 0 \\
&= \frac{\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}[(X_k - m)^2]}{\varepsilon^2} \text{ by independance} \\
&= \frac{\sigma^2}{n\varepsilon^2}
\end{aligned}$$

where $S_n = \frac{1}{n}\sum_{k=1}^n X_k$, so that the rate of convergence of the sample mean in the i.i.d case is $1/n$.

We have ;

$$\begin{aligned}
\mathbb{E}[|\overline{Y_n} - \mu|^2] &= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^n (Y_i - \mu)\right) \\
&= \frac{1}{n^2}\text{Cov}\left(\sum_{i=1}^n (Y_i - \mu), \sum_{j=1}^n (Y_j - \mu)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \text{Cov}(Y_i, Y_j) \\
&= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \gamma(i-j) \\
&\leqslant \frac{1}{n^2}\sum_{i=1}^n \|\gamma\|_1 \\
&= \frac{\|\gamma\|_1}{n} \underset{n \longrightarrow +\infty}{\longrightarrow} 0
\end{aligned}$$

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 $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ $(\psi_0 = 1)$ 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

Using Cauchy-Schwartz inequality we can prove that convergence in L^2 implies convergence in L^1 :

$$\mathbb{E}[|Y_t - Y_{n,t}|] \leq \mathbb{E}[|Y_t - Y_{n,t}|^2]^{\frac{1}{2}}, \text{ by Cauchy-Schwarz}$$

where $Y_{n,t} = \sum_{k=0}^{n} \psi_k \varepsilon_{t-k}$ and by the definition given above of the infinite sum of random variables and taking the limit we have the convergence in L^1.

That said we also have that $\mathbb{E}[Y_t] = \lim\limits_{n\to+\infty} \mathbb{E}[Y_{n,t}]$ (this is given by the triangular inequality of the absolute value).

So to conclude we have that : $\mathbb{E}[Y_{n,t}] = \mathbb{E}[\sum_{k=0}^{n} \psi_k \varepsilon_{t-k}] = \sum_{k=0}^{n} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0 \Rightarrow \mathbb{E}[Y_t] = 0$

For the autocorrelation, we will use Holder's inequality :

$$\begin{aligned}
|\mathbb{E}[Y_{t,n}Y_{t-k,n}] - \mathbb{E}[Y_tY_{t-k}]| &= |\mathbb{E}[Y_{t,n}(Y_{t-k,n} - Y_{t-k})] + \mathbb{E}[Y_{t-k,n}(Y_{t,n} - Y_t)]| \\
&\leq \mathbb{E}[|Y_{t,n}|^2]^{\frac{1}{2}}\mathbb{E}[|Y_{t-k,n} - Y_{t-k}|^2]^{\frac{1}{2}} + \mathbb{E}[|Y_{t-k,n}|^2]^{\frac{1}{2}}\mathbb{E}[|Y_{t,n} - Y_t|^2]^{\frac{1}{2}} \xrightarrow{n\to+\infty} 0
\end{aligned}$$

Again using the definition of the infinite sum. This allows us to intervert limit and expectation.So finally we have :

$$\begin{aligned}
\forall k,\ \mathbb{E}[Y_tY_{t-k}] = \lim_{n\to+\infty} \mathbb{E}[Y_{t,n}Y_{t-k,n}] &= \lim_{n\to+\infty} \sum_{l=0}^{n}\sum_{m=0}^{n} \psi_l\psi_m \mathbb{E}[\varepsilon_{t-l}\varepsilon_{t-k-m}] \\
&= \lim_{n\to+\infty} \sum_{l\geq k}^{n} \psi_l\psi_{l-k}\mathbb{E}[\varepsilon_{t-l}^2] = \sigma_\varepsilon^2 \sum_{l\geq k}^{+\infty} \psi_l\psi_{l-k}
\end{aligned}$$

The last sum is well defined and this is proven by using Cauchy-Schwarz inequality.

Let's calculate the power spectrum of $(Y_t)_t$:

$$\begin{aligned}
S(f) &= \sum_{k\in\mathbb{Z}} \gamma(k) e^{-2i\pi kf} \\
&= \sigma_\varepsilon^2 \sum_{k\in\mathbb{Z}} e^{-2i\pi kf} \sum_{l\geqslant k} \psi_l \psi_{l-k} \\
&= \sigma_\varepsilon^2 \sum_{l\geqslant 0} \psi_l \sum_{k\leqslant l} \psi_{l-k} e^{-2i\pi kf} \\
&= \sigma_\varepsilon^2 \sum_{l\geqslant 0} \psi_l \sum_{m\geqslant 0} \psi_m e^{2i\pi mf} e^{-2i\pi lf} \text{ with the change of variable } m = l-k \\
&= \sigma_\varepsilon^2 \left| \sum_{l\geqslant 0} \psi_l e^{-2i\pi lf} \right|^2 \\
&= \sigma_\varepsilon^2 |\phi(e^{-2i\pi f})|^2
\end{aligned}$$

It seems to us that we also need $(\psi_k)_{k\in\mathbb{N}}$ to be in ℓ^1.

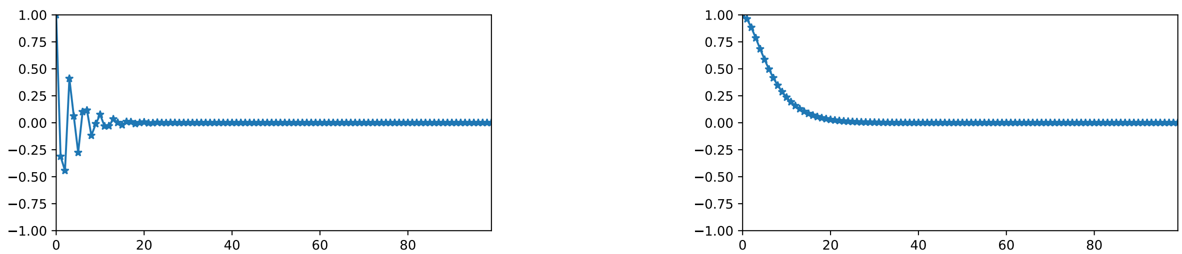
Question 3 *AR(2) process*

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

We want to express the autocovariance of the AR(2) process. We assume that Y_1 and Y_2 are centered. Let's multiply by Y_{t-k} :

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

By taking the expectation we get

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2)$$

We can conclude that $\gamma(k) = a r_1^{-k} + b r_2^{-k}$ with $a + b = \gamma(0), a r_1^{-1} + b r_2^{-1} = \gamma(1)$.

In the two cases above, we have $\gamma(k) \underset{k\to+\infty}{\longrightarrow} 0$ which confirms the hypothesis $|r_i| > 1$. Now in the left correlogram we observe a pseudo-periodic behavior which indicates that the roots of the polynomial are complex. Let's show that.

If the roots are complex, they are conjugate. Then, $\gamma(1) \in \mathbb{R}$ so that $a = b$ or $a = -b$. But $\gamma(0) \neq 0$, hence $a = b$. We write $r_1 = \rho e^{i\theta}$ so that $\gamma(k) = 2a\rho^k \cos(\theta k)$. If the roots are real there is no pseudo-periodicity, with eventually a bump in the graph of the correlogram. It proves that the first AR(2) has complex roots.

Let's calculate the power spectrum of the AR(2) process. We set $s(z) = \sum_{k=0}^{+\infty} \gamma(k) z^k$. We easily show the relation $s(z) = \frac{\gamma(0) - z\phi_1\gamma(0) + \gamma(1)z}{\phi(z)}$.

Then :

$$S(f) = s(e^{2i\pi f}) + s(e^{-2i\pi f}) - \gamma(0)$$

Finally we obtain $S(f) = \dfrac{\sigma^2}{|\phi(e^{-2i\pi f})|^2}$.

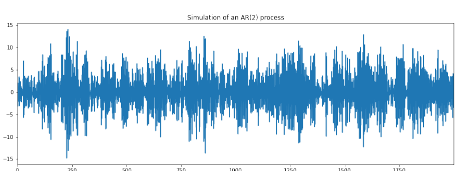
We want to find ϕ_1, ϕ_2 such that

$$1 - \phi_1 z - \phi_2 z^2 = -\phi_2 (re^{-i\theta} - z)(re^{i\theta} - z)$$

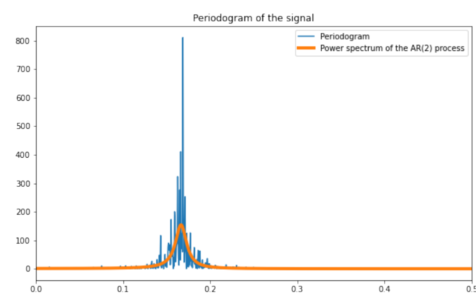
$\iff$

$$\begin{cases} -\phi_2 r^2 & = 1 \\ \phi_2 r & = -\phi_1 \end{cases}$$

We can chose $\phi_2 = -1/r^2$ and $\phi_1 = 1/r$.



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

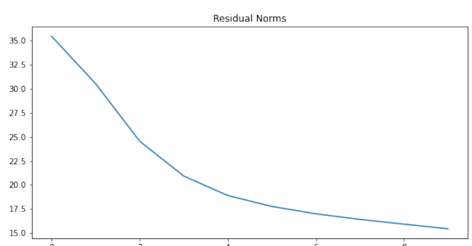
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

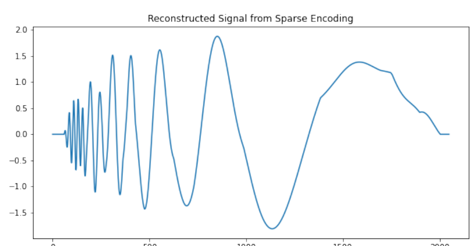
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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