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

Martin Brosset martin.brosset@student-cs.fr
Ben Kabongo ben.kabongo_buzangu@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 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 θ_i ($i \in \{1, ..., n\}$) a n-sample of i.i.d random variables of expectation θ and variance σ^2.
We have

$$\begin{aligned}\mathbb{E}[\hat{\theta}_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\theta_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\theta_i] \\ &= \theta\end{aligned}$$

The sample mean converges to the expectation.
So we have using the independance of the sample :

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_n) &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\theta_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\theta_i) \\ &= \frac{\sigma^2}{n}\end{aligned}$$

Using the Bienaymé-Tchebychev's inequality we get:

$$\forall \epsilon > 0, P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow_{n\rightarrow\infty} 0$$

So $\overline{X_n} \rightarrow_p \mu$ and because μ is a constant, we get $\overline{X_n} \rightarrow_D \mu$ It is thus a consistent estimator.
For the convergence rate, we have from the Central Limit Theorem:

$$\sqrt{n}(\overline{X_n} - \mu) \rightarrow_D N(0, \sigma)$$

This means by definition a rate of convergence of $\sqrt{n}$.

Let (Y_t) be a wide-sense stationary process. We also have $\mathbb{E}(Y_t) = \mu$ (since the expectation is constant).
Using the previous point: $\overline{Y_n}$ is consistent using Tchebychev and we have a rate of convergence

$\sqrt{n}$.
Nevertheless, as suggested in the subject, we have:

$$E((\overline{Y_n}-\mu)^2) = V(\overline{Y_n}) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} Y_i\right)$$

With:

$$\mathrm{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(X_i)^2 = \gamma(0) - \mu^2 = \sigma^2$$
$$\mathrm{Cov}(Y_i Y_j) = \mathbb{E}(Y_i Y_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j) = \gamma(|i-j|) - \mu^2$$

So:

$$\mathbb{E}((\overline{Y_n}-\mu)^2) = \frac{1}{n^2}\left(n\sigma^2 + 2\sum_{1\le i<j\le n} \gamma(|i-j|)\right)$$
$$\le \frac{1}{n}\left(\sigma^2 + 2\sum_{k=1}^{\infty} |\gamma(k)|\right)$$

So $\overline{Y_n} \to_{L_2} \mu$ and thus $\overline{Y_n} \to_D \mu$.

3 AR and MA processes

Question 2 *Infinite order moving average MA(∞)*

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

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

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

1. **Expectation**: Since ε_t has a zero mean, the expected value of Y_t would be:

$$\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t) + \sum_{k=1}^{\infty} \phi_k \mathbb{E}(\varepsilon_{t-k}) = 0$$

Given that $\mathbb{E}(\varepsilon_t) = 0$ for all t, and the ϕ_k are constants.

2. **Autocovariance**: Now, let's calculate the autocovariance function $\mathbb{E}(Y_t Y_{t-k})$.
For $k > 0$, we have:

$$\mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}\left((\varepsilon_t + \phi_1\varepsilon_{t-1} + \ldots)(\varepsilon_{t-k} + \phi_1\varepsilon_{t-k-1} + \ldots)\right)$$

Because ε_t is white noise with zero mean and variance σ_ε^2, $\mathbb{E}(\varepsilon_t\varepsilon_s) = 0$ for $t \neq s$ and $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$. The only terms that will contribute to the expectation are those where the ε's have the same subscript:

$$\mathbb{E}(Y_t Y_{t-k}) = \sigma_\varepsilon^2 \sum_{j=k}^{\infty} \phi_j \phi_{j-k}$$

For $k = 0$, the autocovariance is the variance of Y_t:

$$E(Y_t Y_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi_j^2$$

A process is weakly stationary if the mean is constant (does not depend on time) and the autocovariance function depends only on the lag k and not on time t. From the above calculations, it is evident that:

- The mean $\mathbb{E}(Y_t)$ is constant and equal to zero for all t.
- The autocovariance function $\mathbb{E}(Y_t Y_{t-k})$ depends only on the lag k and not on the specific time t.

Therefore, the process $\{Y_t\}$ is weakly stationary.

Let's now consider the power spectrum of $\{Y_t\}$ with a sampling frequency of 1 Hz.

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

let's pose $k' = j - k$

$$
\begin{aligned}
&= \sigma_\varepsilon^2 \sum_{k'=0}^{+\infty} \sum_{j=0}^{+\infty} \psi_j e^{-2i\pi f j} \, \psi_{k'} e^{2i\pi f k'} \\
&= \sigma_\varepsilon^2 \sum_{k'=0}^{+\infty} \psi_{k'} e^{2i\pi f k'} \, \sum_{j=0}^{+\infty} \psi_j e^{-2i\pi f j} \\
&= \sigma_\varepsilon^2 \, \overline{\phi(e^{-2i\pi f})} \, \phi(e^{-2i\pi f}) \\
&= \sigma_\varepsilon^2 \, |\phi(e^{-2i\pi f})|^2
\end{aligned}
$$

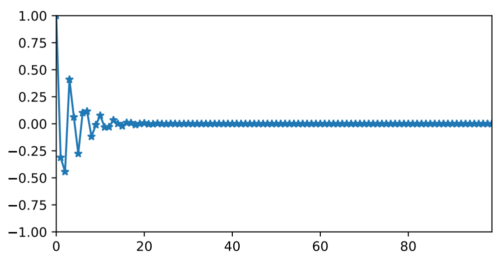
Question 3 *AR(2) process*

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

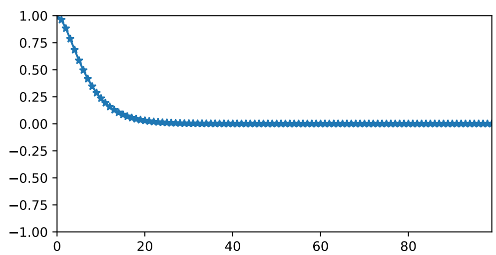
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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 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 have:

$$\gamma(\tau) = \mathbb{E}(Y_t Y_{t+\tau}) = \mathbb{E}\left(Y_t(\phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2} + \epsilon_{t+\tau})\right) = \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2) + \mathbb{E}(Y_t\epsilon_{t+\tau})$$

With:

$$\mathbb{E}(Y_t\epsilon_{t+\tau}) = \mathbb{E}\left(((\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t)\epsilon_{t+\tau})\right) = \phi_1\mathbb{E}(Y_{t-1}\epsilon_{t+\tau}) + \phi_2\mathbb{E}(Y_{t-2}\epsilon_{t+\tau}) + \mathbb{E}(\epsilon_t\epsilon_{t+\tau}) = \delta_\tau\sigma_\epsilon^2$$

So $(\gamma(\tau))_{\tau\in\mathbb{N}}$ is a sequence with recurrent schema of second order:

$$\gamma(0) = \mathbb{E}(Y_t^2) \quad \gamma(\tau+2) = \phi_1\gamma(\tau+1) + \phi_2\gamma(\tau-2)$$

Its characteristic polynomial is:

$$r^2 - \phi_1 r - \phi_2 = 0$$
$$\phi(\frac{1}{r}) = 0$$

We expect $\gamma(\tau)$ in the form: $\lambda(\frac{1}{r_1})^\tau + \mu(\frac{1}{r_2})^\tau$ with λ, μ two constants if the solutions are real. If the solutions are complex, we expect $\gamma(\tau)$ in the form: $\lambda(\frac{1}{r})^\tau \cos(n\alpha) + \mu(\frac{1}{r})^\tau \sin(n\alpha)$ with λ, μ two constants.

Using this, we can tell that the correlogram of the first AR(2) has complex roots since it oscillates and the second has real roots because it's damped.

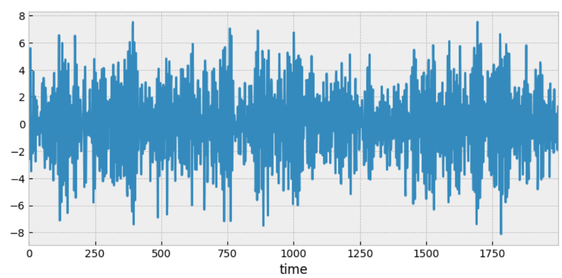
We have:

$$S(f) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau} = \sum_{\tau=-\infty}^{+\infty} (\phi_1 \gamma(\tau-1) + \phi_2 \gamma(\tau-2) + \delta_\tau \sigma_\epsilon^2) e^{-2\pi i f \tau}$$

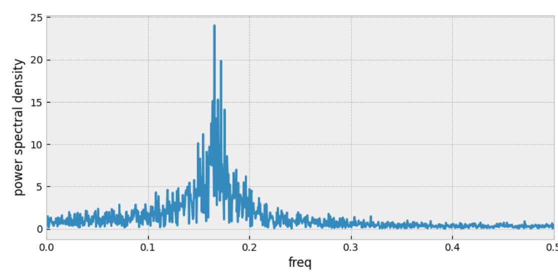
$$= (\phi_1 e^{2\pi i f} + \phi_2 e^{4\pi i f}) S(f) + \sigma_\epsilon^2$$

Thus we get:

$$S(f) = \frac{\sigma_\epsilon^2}{1 - \phi_1 e^{2\pi i f} - \phi_2 e^{4\pi i f}} = \frac{\sigma_\epsilon^2}{\phi(e^{2\pi i f})}$$



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 ($k = 0, \ldots, L-1$) by

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

where w_L is a modulating window given by

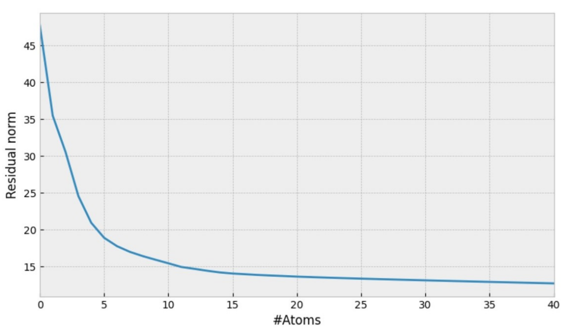
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

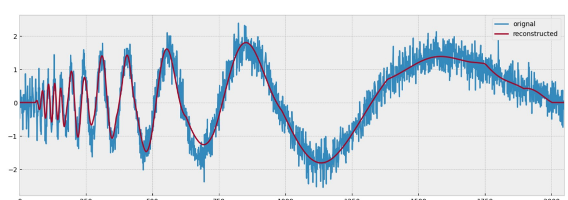
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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