Assignment 2 (ML for TS) - MVA

Aya Laajil aya.laajil@student-cs.fr
Martin Jolif martinjolif@gmail.com

December 2, 2024

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 2nd 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: https://docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2

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 realizations 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 be $(X_i)_{i\in[\![1,n]\!]}$ i.i.d random variables such that, for all $i \in [\![1, n]\!]$, $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(X_i^2) < +\infty$.

  Let's first show that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is a consistent estimator of μ: By the Bienaymé-tchebychev inequality for $\epsilon > 0$:

  $$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(X_1)}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0$$

  Therefore, $\bar{X}_n$ is a consistent estimator.

  Moreover by the central limit theorem,

  $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{law} \mathcal{N}(0, \mathrm{Var}(X_1))$$

  Therefore, the rate convergence of the consistent estimator $\bar{X}_n$ is $\frac{1}{\sqrt{n}}$.

- Let's prove the L_2 convergence of $\bar{Y}_n$,

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

$$
\begin{aligned}
&\leq \frac{1}{n^2}\left(2\sum_{i=1}^{n}|n-i||\gamma(i)| + \sum_{i=1}^{n}|\gamma(0)|\right) \\
&\leq \frac{1}{n}\left(2\sum_{i=1}^{n}|\gamma(i)| + |\gamma(0)|\right) \\
&\leq \frac{2}{n}\left(\sum_{i=0}^{n}|\gamma(i)|\right) \\
&\leq \frac{2}{n}\left(\sum_{i=0}^{+\infty}|\gamma(i)|\right) \xrightarrow[n\to\infty]{} 0
\end{aligned}
$$

As the $(\gamma(i))_i$ are summable. Therefore,

$$\bar{Y}_n \xrightarrow[n\to\infty]{L_2} \mu$$

As the L_2 convergence implies the convergence in probability, it shows that $\bar{Y}_n$ is a consistent estimator.

The L_2 convergence implies the L_1 convergence, and we have :

$$\mathbb{E}[(\bar{Y}_n - \mu)] \leq \sqrt{\mathbb{E}[(\bar{Y}_n - \mu)^2]} \leq \sqrt{\frac{2}{n}\left(\sum_{i=0}^{+\infty}|\gamma(i)|\right)} \leq \sqrt{\frac{1}{n}}\sqrt{2\left(\sum_{i=0}^{+\infty}|\gamma(i)|\right)}$$

As the $(\gamma(i))_i$ are summable, the second term is a $\frac{1}{\sqrt{n}}$ x constant. Therefore the convergence rate is $\frac{1}{\sqrt{n}}$.

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 2.

Answer 2

- As Y_t is in $L_2 \subset L_1$, Y_t is integrable and we can swap expectation and series:

  $$\mathbb{E}(Y_t) = \sum_{k=0}^{\infty} \psi_k \mathbb{E}(\varepsilon_{t-k}) = 0$$

  Moreover,

  $$\begin{aligned}\mathbb{E}(Y_t Y_{t-\tau}) &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \psi_k \psi_l \mathbb{E}(\varepsilon_{t-k}\varepsilon_{t-\tau-l}) \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \psi_k \psi_l \gamma_\varepsilon(\tau + l - k) \\ &= \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_{\tau+l}\psi_l\end{aligned}$$

  as $\gamma_\varepsilon(x) = \sigma_\varepsilon^2$ if $x = 0$ and 0 otherwise.

  From the previous expression, we get that $\mathbb{E}(Y_t Y_{t-\tau})$ is a function of τ. The variance of Y_t is:

  $$\mathrm{Var}(Y_t) = \mathrm{Cov}(Y_t, Y_t) = \mathbb{E}[Y_t^2].$$

  Expanding:

  $$\mathbb{E}[Y_t^2] = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}\right)^2\right].$$

Using the uncorrelatedness of ϵ_t, cross terms vanish, leaving:

$$\text{Var}(Y_t) = \sigma_\epsilon^2 \sum_{k=0}^{\infty} \psi_k^2.$$

Since $\sum_{k=0}^{\infty} \psi_k^2 < \infty$, $\text{Var}(Y_t)$ is finite and constant over time, satisfying the third condition for weak stationarity.

Therefore,the process is weakly stationary.

- As $f_s = 1$ Hz,

$$\begin{aligned}
S(f) &= \sum_{\tau \in \mathbb{Z}} \gamma(\tau) \exp(-i2\pi f \tau) \\
&= \sum_{\tau \in \mathbb{Z}} \mathbb{E}(Y_0 Y_\tau) \exp(-i2\pi f \tau) \\
&= \sigma_\epsilon^2 \sum_{\tau \in \mathbb{Z}} \sum_{l \in \mathbb{N}} \psi_l \psi_{l+\tau} \exp(-i2\pi f \tau) \\
&= \sigma_\epsilon^2 \sum_{\tau \in \mathbb{Z}} \sum_{(k,l) \in \mathbb{N}^2} \psi_l \psi_k \exp(-i2\pi f \tau) \mathbb{1}_{k=\tau+l} \\
&= \sigma_\epsilon^2 \sum_{(k,l) \in \mathbb{N}^2} \psi_l \psi_k \exp(-i2\pi f (k-l)) \\
&= \sigma_\epsilon^2 (\sum_{k \in \mathbb{N}} \psi_k e^{-2i\pi f k})(\sum_{l \in \mathbb{N}} \psi_l e^{2i\pi f l}) \\
&= \sigma_\epsilon^2 |\phi(e^{-2i\pi f})|^2
\end{aligned}$$

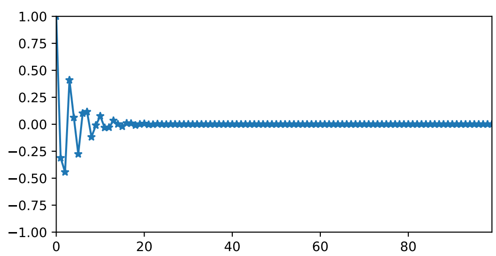
Question 3 *AR(2) process*

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

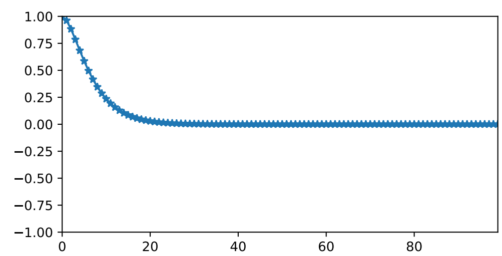
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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 behavior 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

- Expressing the autocovariance function $\gamma(\tau)$:

  The autocovariance function $\gamma(\tau) = \mathrm{Cov}(Y_t, Y_{t-\tau})$ satisfies the same recurrence relation as the AR(2) process:
  $$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2), \quad \tau \geq 2.$$

  The general solution to this recurrence relation can be expressed in terms of the roots r_1 and r_2 of the characteristic polynomial.

  When the roots r_1 and r_2 are real, it gives:

  $$\gamma(\tau) = \frac{C_1}{r_1^{|\tau|}} + \frac{C_2}{r_2^{|\tau|}},$$

  where C_1 and C_2 are constants determined by the initial conditions $\gamma(0)$ and $\gamma(1)$.

  $$\gamma(0) = \mathrm{Var}(Y_t), \quad \gamma(1) = \mathrm{Cov}(Y_t, Y_{t-1}).$$

Substitute $\tau = 0$ and $\tau = 1$ into the general solution:

$$\gamma(0) = C_1 + C_2,$$

$$\gamma(1) = \frac{C_1}{r_1} + \frac{C_2}{r_2}.$$

Solving for C_1 and C_2, we obtain:

$$C_1 = \frac{r_1\gamma(0) - r_1 r_2 \gamma(1)}{r_1 - r_2}, \quad C_2 = \frac{r_1 r_2 \gamma(1) - r_2 \gamma(0)}{r_1 - r_2}.$$

When the roots r_1 and r_2 are complex conjugates, let $r_1 = \rho e^{i\theta}$ and $r_2 = \rho e^{-i\theta}$, where $\rho = |r_1| = |r_2|$ (the modulus) and θ is the angular frequency. The autocovariance function can be expressed as:

$$\gamma(\tau) = \frac{1}{\rho^{|\tau|}} \left(A\cos(|\tau|\theta) + B\sin(|\tau|\theta) \right),$$

where A and B are related to C_1 and C_2 as follows:

$$A = \gamma(0), \quad B = \frac{\rho\gamma(1) - \gamma(0)\cos(\theta)}{\sin(\theta)}.$$

as $|r_i| > 1, \sin(\theta) \neq 0$

- If the roots are real, the correlogram does not oscillate, it simply decreases towards zero. If the roots are complex, it exhibits an oscillatory pattern, indicating the influence of the sinusoidal terms (cos and sin). Therefore, the correlogram of the first AR(2) has complex roots, and the second has real roots.

- Taking the Z-transform of the AR(2) process, we get:

$$Y(z) = \phi_1 z^{-1} Y(z) + \phi_2 z^{-2} Y(z) + \epsilon(z)$$

Rearranging terms, we have:

$$Y(z)(1 - \phi_1 z^{-1} - \phi_2 z^{-2}) = \epsilon(z)$$

The term in parentheses is the characteristic polynomial $\tilde{\phi}(z)$:

$$\tilde{\phi}(z) = 1 - \phi_1 z^{-1} - \phi_2 z^{-2}$$

Thus, we have:

$$Y(z)\tilde{\phi}(z) = \epsilon(z)$$

Solving for $Y(z)$:

$$Y(z) = \frac{\epsilon(z)}{\tilde{\phi}(z)}$$

The Fourier transform is a special case of the Z-transform evaluated on $z = e^{-2\pi if}$. Therefore, substituting $z = e^{-2\pi if}$ into the Z-transform expression, we get:

$$Y(f) = \frac{\epsilon(f)}{\tilde{\phi}(e^{-2\pi if})}$$

The power spectrum $S(f)$ is defined as the expected value of the squared magnitude of the Fourier transform of the process:

$$S(f) = \mathbb{E}[|Y(f)|^2]$$

Substituting the Fourier transform of Y_t into the power spectrum definition:

$$S(f) = \mathbb{E}\left[\left|\frac{\epsilon(f)}{\tilde{\phi}(e^{-2\pi if})}\right|^2\right]$$

Since $\epsilon(f)$ is the Fourier transform of a white noise process with variance σ_ε^2, we have:

$$\mathbb{E}[|\epsilon(f)|^2] = \sigma_\varepsilon^2$$

Therefore : The power spectrum $S(f)$ for an AR(2) process is given by:

$$\boxed{S(f) = \frac{\sigma_\varepsilon^2}{|\phi(e^{2\pi if})|^2}}$$

where $\phi(e^{2\pi if}) = 1 - \phi_1 e^{2\pi if} - \phi_2 e^{4\pi if} = \tilde{\phi}(e^{-2\pi if})$.

- To choose ϕ_1 and ϕ_2 such that the characteristic polynomial of the AR(2) process has two complex conjugate roots with a norm $r = 1.05$ and phase $\theta = \frac{2\pi}{6} = \frac{\pi}{3}$, we need to find the roots in polar form and then convert them to the coefficients of the characteristic polynomial.

  The roots of the characteristic polynomial are given by:

$$z_1 = re^{i\theta}$$

$$z_2 = re^{-i\theta}$$

  Given $r = 1.05$ and $\theta = \frac{\pi}{3}$, the roots are:

$$z_1 = 1.05e^{i\frac{\pi}{3}}$$

$$z_2 = 1.05e^{-i\frac{\pi}{3}}$$

  Using Euler's formula, $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, we get:

$$z_1 = 1.05\left(\cos\left(\frac{\pi}{3}\right) + i\sin\left(\frac{\pi}{3}\right)\right)$$

$$z_2 = 1.05\left(\cos\left(\frac{\pi}{3}\right) - i\sin\left(\frac{\pi}{3}\right)\right)$$

Since $\cos\left(\frac{\pi}{3}\right) = \frac{1}{2}$ and $\sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}$, the roots become:

$$z_1 = 1.05\left(\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) = 0.525 + 0.9093i$$

$$z_2 = 1.05\left(\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) = 0.525 - 0.9093i$$

The characteristic polynomial with roots z_1 and z_2 is:

$$\phi(z) = (z - \frac{1}{z_1})(z - \frac{1}{z_2})$$

Expanding this, we get:

$$\phi(z) = z^2 - (\frac{1}{z_1} + \frac{1}{z_2})z + \frac{1}{z_1 z_2}$$

1. Sum of the Roots :

$$\frac{1}{z_1} + \frac{1}{z_2} = 0.9523809524$$

2. Product of the Roots :

$$\frac{1}{z_1 z_2} = 0.9070294784$$

Thus, the characteristic polynomial is:

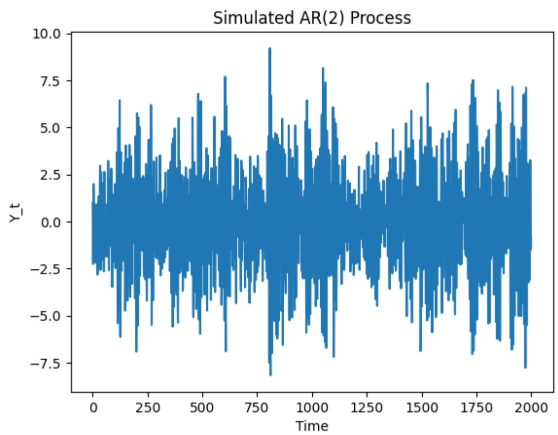
$$\phi(z) = z^2 - 0.9523809524z + 0.9070294784$$

From the characteristic polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$, we identify:

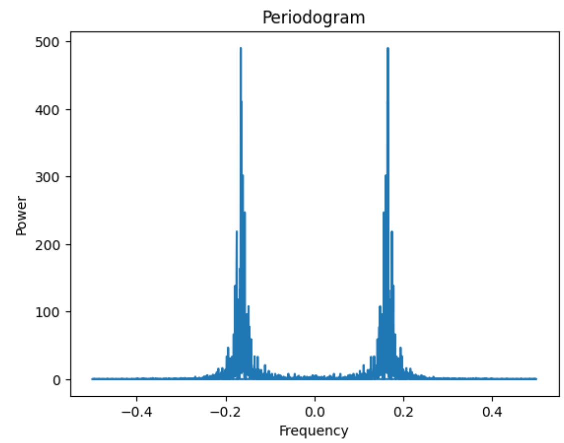
$$\phi_1 = 0.9523809524$$

$$\phi_2 = -0.9070294784$$

The simulated AR(2) process will exhibit oscillatory behavior due to the complex conjugate roots. Periodogram: The periodogram will show a peak at the frequency corresponding to the phase $\theta = \pi/3$ (frequency $= 1/6 \approx 0.166$) indicating the presence of a significant periodic component at that frequency. The use of Welch's method helps to smooth the periodogram, reducing variance and providing a clearer representation of the power spectral density.



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 an 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

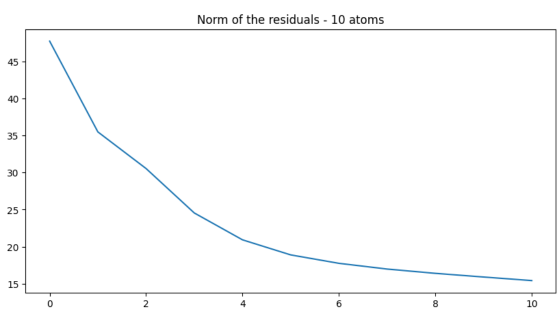
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

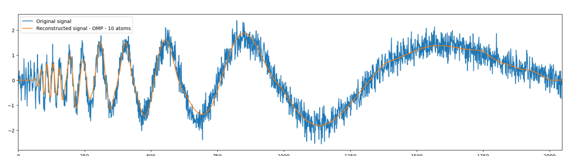
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- For the sparse coding, implement the Orthogonal Matching Pursuit (OMP). (Use convolutions to compute the correlation 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