

Assignment 2 (ML for TS) - MVA

Firstname Lastname youremail1@mail.com

Firstname Lastname youremail2@mail.com

November 19, 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 (Ω, \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 $\bar{Y}_n = (Y_1 + \dots + Y_n)/n$ 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

Let X_1, \dots, X_n be i.i.d. random variables with mean μ and variance σ^2 . The sample mean is $\bar{X}_n = \frac{(X_1 + \dots + X_n)}{n}$. Thanks to Bienaymé-Tchebychev inequality, we have that $\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} \leq \frac{\sigma^2}{n\varepsilon^2}$. The convergence rate is determined by the standard deviation of the sample mean, which is σ/\sqrt{n} .

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

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

$$= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\mathbb{E}(Y_i Y_j) - \mu^2) \quad (3)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\mathbb{E}(Y_i Y_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j)) \quad (4)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(i-j) \quad (5)$$

$$= \frac{1}{n^2} \sum_{k=-(n-1)}^{n-1} (n-|k|)\gamma(k) \quad (6)$$

$$\leq \frac{1}{n^2} \sum_{k=-(n-1)}^{n-1} n\gamma(k) \quad (7)$$

$$= \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \gamma(k) \quad (8)$$

$$\leq \frac{1}{n} \sum_{k=-\infty}^{\infty} \gamma(k) \xrightarrow{n \rightarrow \infty} 0 \quad (9)$$

We have proved that $\bar{Y}_n \xrightarrow{L^2} \mu$.

Moreover as the convergence in L_2 implies the convergence in probability, we have that $\bar{Y}_n \xrightarrow{P} \mu$.

3 AR and MA processes

Question 2 Infinite order moving average MA(∞)

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

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad (10)$$

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 (10).

Answer 2

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

$$\mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right)\left(\sum_{k'=0}^{\infty} \psi_{k'} \varepsilon_{t-k-k'}\right)\right] \quad (11)$$

$$= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \psi_k \psi_{k'} \mathbb{E}(\varepsilon_{t-k} \varepsilon_{t-k-k'}) \quad (12)$$

$$= \sum_{k=0}^{\infty} \psi_k^2 \sigma_\varepsilon^2 \quad (13)$$

$$(14)$$

This quantity exists, as $\sum_k \psi_k^2 < \infty$. Moreover, it only depends on $|k - k'|$, so the process is **weakly stationary**.

$$S(f) = \sum_{\tau=-\infty}^{\infty} e^{-2\pi f\tau} \quad (15)$$

$$= \sum_{\tau=-\infty}^{\infty} \left(\sum_{j=\tau}^{\infty} \psi_j \psi_{j-\tau} \right) \sigma_{\epsilon}^2 e^{-2i\pi f\tau} \quad (16)$$

$$= \sum_{\tau=-\infty}^{\infty} \left(\sum_{j=\tau}^{\infty} \psi_j \psi_{j-\tau} \right) \sigma_{\epsilon}^2 e^{-2i\pi f\tau} \quad (17)$$

$$= \sum_{j=-\infty}^{\infty} \psi_j \sum_{\tau=-\infty}^j \psi_{j-\tau} \sigma_{\epsilon}^2 e^{-2i\pi f\tau} \quad (18)$$

$$= \sum_{j=-\infty}^{\infty} \psi_j \sum_{\tau=0}^{\infty} \psi_{\tau} \sigma_{\epsilon}^2 e^{-2i\pi f(j-\tau)} \quad (19)$$

$$= \sum_{j,j'} \psi_j \psi_{j'} \sigma_{\epsilon}^2 e^{-2i\pi f(j-j')} \quad (20)$$

$$= \sum_{j,j'} \psi_j \psi_{j'} \sigma_{\epsilon}^2 e^{-2i\pi f j} e^{-2i\pi f j'} \quad (21)$$

$$= \sigma_{\epsilon}^2 \sum_j \psi_j e^{-2i\pi f j} \sum_{j'} \psi_{j'} e^{2i\pi f j'} \quad (22)$$

$$= \sigma_{\epsilon}^2 \left| \sum_j \psi_j e^{-2i\pi f j} \right|^2 \quad (23)$$

$$= \sigma_{\epsilon}^2 |\phi(e^{-2\pi i f})|^2 \quad (24)$$

Which is the expected result.

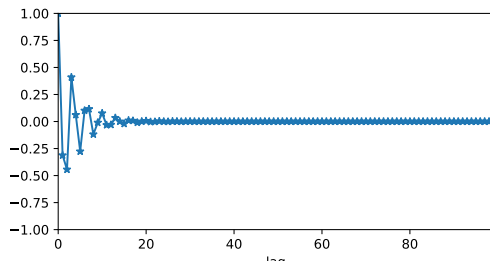
Question 3 AR(2) process

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

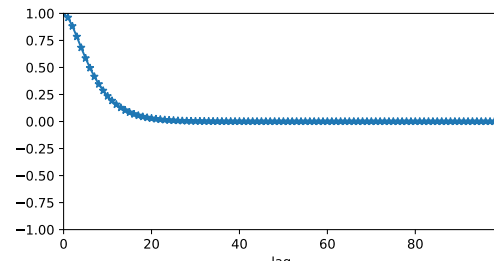
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad (25)$$

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



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

$$\forall u = 0, \dots, 2L - 1, \quad \phi_{L,k}[u] = w_L[u] \sqrt{\frac{2}{L}} \cos\left[\frac{\pi}{L} \left(u + \frac{L+1}{2}\right) \left(k + \frac{1}{2}\right)\right] \quad (26)$$

where w_L is a modulating window given by

$$w_L[u] = \sin\left[\frac{\pi}{2L} \left(u + \frac{1}{2}\right)\right]. \quad (27)$$

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 correlation coefficients.)
- Display the norm of the successive residuals and the reconstructed signal with 10 atoms.

Answer 4

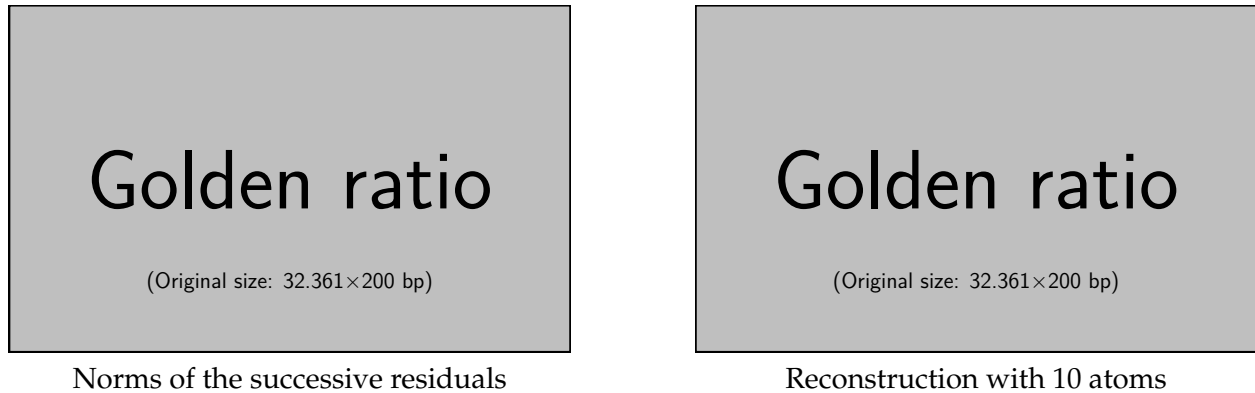


Figure 3: Question 4