Assignment 2 (ML for TS) - MVA

Soël Megdoud soel.megdoud@mail.com
Julien Delavande delavandejulien@mail.com

November 21, 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 2^{nd} 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

- For i.i.d. random variables $\{X_i\}$ with mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$, the Central Limit Theorem (CLT) states that:

  $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

  where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the sample mean.

  This implies that the sample mean $\bar{X}_n$ converges to the true mean μ at a rate of $1/\sqrt{n}$.

- First, note that $\mathbb{E}[\bar{Y}_n] = \mu$, so $\bar{Y}_n$ is an unbiased estimator of μ. To analyze the rate of convergence, we compute the mean squared error:

  $$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \mathrm{Var}(\bar{Y}_n) = \frac{1}{n^2}\left(\sum_{t=1}^n \mathrm{Var}(Y_t) + 2\sum_{s<t}\mathrm{Cov}(Y_t, Y_s)\right) = \frac{1}{n^2}\left(\sum_{t=1}^n \gamma(0) + 2\sum_{s<t}\gamma(t-s)\right),$$

  where $\gamma(t-s)$ is the autocovariance function.

  Since the process is wide-sense stationary, the variance $\mathrm{Var}(Y_t) = \gamma(0)$ is constant, and the covariance $\mathrm{Cov}(Y_t, Y_s)$ depends only on the lag $k = |t-s|$. The expression simplifies to:

  $$\mathrm{Var}(\bar{Y}_n) = \frac{1}{n^2}\left(n\gamma(0) + 2\sum_{k=1}^{n-1}(n-k)\gamma(k)\right).$$

  Using the assumption $\sum_{k=-\infty}^{\infty}|\gamma(k)| < \infty$, which implies $\sum_{k=1}^{\infty}|\gamma(k)| < \infty$, we can bound the second term:

  $$\sum_{k=1}^{n-1}(n-k)|\gamma(k)| \leq n\sum_{k=1}^{\infty}|\gamma(k)|.$$

  Thus, the variance becomes:

  $$\mathrm{Var}(\bar{Y}_n) \leq \frac{1}{n^2}\left(n\gamma(0) + 2n\sum_{k=1}^{\infty}|\gamma(k)|\right) = \frac{1}{n}\left(\gamma(0) + 2\sum_{k=1}^{\infty}|\gamma(k)|\right).$$

  Since the right-hand side decreases as $\mathcal{O}(1/n)$, we conclude:

  $$\mathbb{E}[(\bar{Y}_n - \mu)^2] = \mathrm{Var}(\bar{Y}_n) = \mathcal{O}\left(\frac{1}{n}\right).$$

Convergence in L_2 (mean squared error tending to zero) implies convergence in probability:

$$\bar{Y}_n \xrightarrow{P} \mu.$$

This proves that $\bar{Y}_n$ is a consistent estimator of μ.

Using Chebyshev's inequality, we get:

$$P(|\bar{Y}_n - \mu| \geq \epsilon) \leq \frac{\mathbb{E}[(\bar{Y}_n - \mu)^2]}{\epsilon^2} = \mathcal{O}\left(\frac{1}{n\epsilon^2}\right) = \mathcal{O}(1/\sqrt{n}).$$

This shows that $\bar{Y}_n$ converges to μ in probability at a rate of $1/\sqrt{n}$, which matches the rate of convergence in the i.i.d. case.

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 (1).

Answer 2

- Weakly stationary:
  - Deriving $\mathbb{E}[Y_t]$:

    Since $\{\varepsilon_t\}$ is a zero-mean white noise:

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

  - Computing $\mathbb{E}[Y_t Y_{t-h}]$:

    Express Y_t and Y_{t-h}:

    $$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad Y_{t-h} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-h-j}.$$

    Compute the autocovariance function $\gamma(h)$:

    $$\begin{aligned}\gamma(h) &= \mathbb{E}[Y_t Y_{t-h}] \\ &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-h-j}\right)\right] \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}\left[\varepsilon_{t-i}\varepsilon_{t-h-j}\right].\end{aligned}$$

    Since $\{\varepsilon_t\}$ is white noise with variance σ_ε^2:

$$\mathbb{E}\left[\varepsilon_{t-i}\varepsilon_{t-h-j}\right] = \sigma_\varepsilon^2 \delta_{i,h+j},$$

where $\delta_{i,h+j}$ is the Kronecker delta (equals 1 if $i = h + j$, 0 otherwise). Therefore:

$$\gamma(h) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_{h+j}\psi_j.$$

- Conclusion on Weak Stationarity:
  * Mean: $\mathbb{E}[Y_t] = 0$, constant over time.
  * Autocovariance: $\gamma(h)$ depends only on the lag h, not on t.

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

- Power Spectrum:

  The time series is:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

  where ε_t is white noise with variance σ_ε^2.

  The power spectrum $S(f)$ is the Fourier transform of the autocovariance function $\gamma(h)$:

$$S(f) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i f h}.$$

  The autocovariance is:

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{h+j}, & h \geq 0, \\ \gamma(-h), & h < 0. \end{cases}$$

  Substituting $\gamma(h)$ into $S(f)$, we reorder terms and let $m = h + j$, giving:

$$S(f) = \sigma_\varepsilon^2 \left| \sum_{j=0}^{\infty} \psi_j e^{-2\pi i f j} \right|^2.$$

  Define the transfer function:

$$\varphi(z) = \sum_{j=0}^{\infty} \psi_j z^j.$$

  For $z = e^{-2\pi i f}$, the power spectrum becomes:

$$S(f) = \sigma_\varepsilon^2 \left| \varphi(e^{-2\pi i f}) \right|^2,$$

where $\varphi(z)$ represents the frequency response of the process.

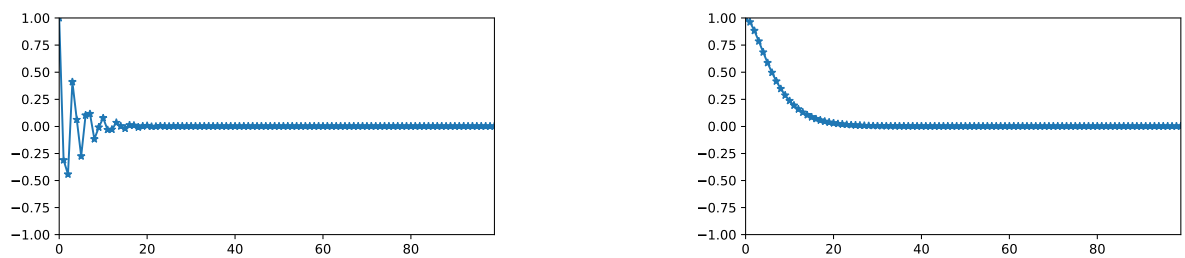
Question 3 *AR(2) process*

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

- The Yule-Walker equation for an AR(2) process is:

  $$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2), \quad \text{for } \tau \geq 2.$$

  This is a linear homogeneous recurrence relation of order 2.

  Let's assume a solution of the form $\gamma(\tau) = r^\tau$, where r is a constant to be determined. Substituting this into the recurrence relation gives:

  $$r^\tau = \phi_1 r^{\tau-1} + \phi_2 r^{\tau-2}.$$

  We obtain:

  $$r^2 = \phi_1 r + \phi_2.$$

  Therefore, r is a solution of the characteristic equation. Let r_1 and r_2 be the roots of $\phi(z) = 0$. Depending on the nature of the roots, we have:

1. If r_1 and r_2 are real and distinct, the solution is:

$$\gamma(\tau) = c_1 r_1^\tau + c_2 r_2^\tau,$$

where c_1 and c_2 are constants to be determined by the initial conditions $\gamma(0)$ and $\gamma(1)$.

2. If r_1 and r_2 are complex conjugates, i.e., $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$, Euler's formula gives:

$$\gamma(\tau) = r^\tau \left(c_1 \cos(\tau\theta) + c_2 \sin(\tau\theta)\right),$$

where $r = |r_1| = |r_2|$ is the modulus of the roots, and θ is their phase (argument).

3. The constants c_1 and c_2 are determined from the initial conditions:

$$\gamma(0) = c_1 + c_2,$$
$$\gamma(1) = c_1 r_1 + c_2 r_2.$$

The autocovariance $\gamma(\tau)$ for an AR(2) process is:

$$\gamma(\tau) = \begin{cases} c_1 r_1^\tau + c_2 r_2^\tau, & \text{if } r_1, r_2 \text{ are real,} \\ r^\tau \left(c_1 \cos(\tau\theta) + c_2 \sin(\tau\theta)\right), & \text{if } r_1, r_2 \text{ are complex conjugates.} \end{cases}$$

- In Figure 1, the correlogram of the first AR(2) process shows oscillatory behavior, which is characteristic of AR(2) processes with complex roots. The second correlogram exhibits exponential decay, indicative of AR(2) processes with real roots.

- The AR(2) process is defined as:

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

where ε_t is white noise with zero mean and variance σ_ε^2.

Taking the Fourier transform of both sides, we get:

$$\mathcal{F}[Y_t] = \phi_1 \mathcal{F}[Y_{t-1}] + \phi_2 \mathcal{F}[Y_{t-2}] + \mathcal{F}[\varepsilon_t].$$

In the frequency domain, the backshift operator B^k becomes a multiplication by $e^{-i2\pi fk}$. Therefore:

$$\tilde{Y}(f) = \phi_1 e^{-i2\pi f} \tilde{Y}(f) + \phi_2 e^{-i4\pi f} \tilde{Y}(f) + \tilde{\varepsilon}(f),$$

where $\tilde{Y}(f)$ and $\tilde{\varepsilon}(f)$ are the Fourier transforms of Y_t and ε_t, respectively.

Therefore we get:

$$\tilde{Y}(f)\left(1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f}\right) = \tilde{\varepsilon}(f).$$

and:

$$\tilde{Y}(f) = \frac{\tilde{\varepsilon}(f)}{\Phi(e^{-i2\pi f})}.$$

with the characteristic polynomial defined as:

$$\Phi(e^{-i2\pi f}) = 1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f}.$$

The power spectrum $S(f)$ is defined as:

$$S(f) = \mathbb{E}[|\tilde{Y}(f)|^2] = \frac{\sigma_\varepsilon^2}{|\Phi(e^{-i2\pi f})|^2}.$$

Finally, after calculation, we have:

$$S(f) = \frac{\sigma_\varepsilon^2}{(1 - \phi_1 \cos(2\pi f) - \phi_2 \cos(4\pi f))^2 + (\phi_1 \sin(2\pi f) + \phi_2 \sin(4\pi f))^2}.$$

- Let's write: $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$:

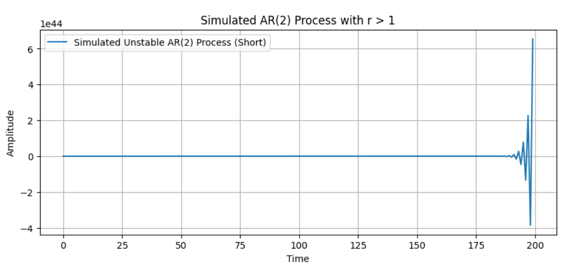
  Therefore :

  - $\phi_1 = -(re^{i\theta} + re^{-i\theta}) = -2r\cos(\theta)$
  - $\phi_2 = re^{i\theta} \cdot re^{-i\theta} = r^2$

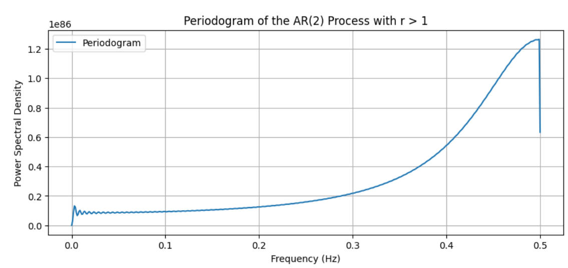
  Substituting $r = 1.05$ and $\theta = 2\pi/6$, we compute:

  $$\phi_1 = -2(1.05)\cos(2\pi/6), \quad \phi_2 = (1.05)^2.$$

  - Signal figure: The signal shows unstable oscillations that increase exponentially in amplitude over time. This instability is expected because the roots of the AR(2) characteristic polynomial have a magnitude $r > 1$, causing the recursive contributions from previous terms to grow instead of decay. The oscillatory nature of the process comes from the complex roots of the characteristic polynomial ($r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$). These roots introduce sinusoidal components in the dynamics.
  - Periodogram figure: The power spectral density $S(f)$ increases rapidly with frequency. This is due to the fact that, in the presence of instability ($r > 1$), the energy of higher-frequency components grows disproportionately.



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

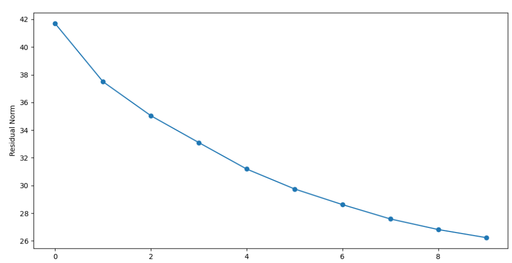
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

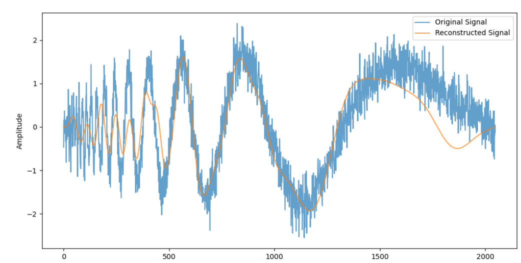
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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