Assignment 2 (ML for TS) - MVA

Raphael Faure raphael.faure@student-cs.fr
Victor Jessequel victorjessequel@gmail.com

November 30, 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$^{\text{nd}}$ 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

Question a.

On note σ^2 la variance des X_i. Comme ils sont i.i.d., il vient :

$$\mathbb{V}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{\mathbb{V}(X_i)}{n^2} = \frac{\sigma^2}{n}.$$

La décomposition biais-variance nous donne (en notant μ l'espérance des X_i) :

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{X_i}{n} - \mu\right)^2\right] = \mathbb{V}\left[\sum_{i=1}^{n} \frac{X_i}{n} - \mu\right] = \mathbb{V}\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{\sigma^2}{n}.$$

En prenant la racine carrée, on a bien un $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

Question b.

$$\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \mu\right)^2 = \left(\sum_{i=1}^{n} \frac{Y_i - \mu}{n}\right)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} (Y_i - \mu)\right)^2.$$

D'où, en passant à l'espérance :

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \mu\right)^2\right] = \frac{1}{n^2} \sum_{i,j\leq n} \mathbb{E}[(Y_i - \mu)(Y_j - \mu)].$$

Avec la stationnarité du processus $\{Y_i\}$, on a :

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \mu\right)^2\right] = \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{p=1}^{n-k} \mathbb{E}[(Y_{p+k} - \mu)(Y_p - \mu)] = \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{p=1}^{n-k} \gamma[k] = \frac{1}{n} \sum_{k=0}^{n-1} \frac{(n-k)}{n} \gamma[k].$$

Or,

$$\delta_{k\leq n} \frac{n-k}{n^2} \gamma[k] \leq \gamma[k] \quad \text{et} \quad \lim_{n\to\infty} \delta_{k\leq n} \frac{n-k}{n^2} \gamma[k] = 0.$$

Le théorème de convergence dominée nous donne finalement :

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \mu\right)^2\right] = \lim_{n\to\infty} \sum_{k\in\mathbb{N}} \delta_{\,k\leq n} \frac{n-k}{n^2}\gamma[k] = \sum_{k\in\mathbb{N}} \lim_{n\to\infty} \delta_{\,k\leq n} \frac{n-k}{n^2}\gamma[k] = 0.$$

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

Question a. Soient $p \leq 0$ et $t \in \mathbb{R}$, on a d'après l'inégalité de Cauchy-Schwarz :

$$\left| \mathbb{E}\left[Y_t - \sum_{k\leq p} \psi_k \varepsilon_{t-k} \right] \right| \leq \mathbb{E}\left[\left(Y_t - \sum_{k\leq p} \psi_k \varepsilon_{t-k} \right)^2 \right].$$

Donc, comme on a convergence des sommes partielles dans L_2 :

$$\lim_{p\to\infty} \mathbb{E}\left[\sum_{k\leq p} \psi_k \varepsilon_{t-k} \right] = \mathbb{E}[Y_t].$$

Or :

$$\mathbb{E}\left[\sum_{k\leq p} \psi_k \varepsilon_{t-k} \right] = \sum_{k\leq p} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0.$$

D'où :

$$\boxed{\mathbb{E}[Y_t] = 0}.$$

Soit $k \in \mathbb{Z}$:

$$\left| \mathbb{E}\left[Y_t Y_{t-k} - \sum_{i,j\leq p} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t+k-j} \right] \right| \leq \left| \mathbb{E}\left[Y_t Y_{t-k} - Y_{t-k} \sum_{i\leq p} \psi_i \varepsilon_{t-i} \right] \right| + \left| \mathbb{E}\left[Y_{t-k} \sum_{i\leq p} \psi_i \varepsilon_{t-i} - \sum_{i,j\leq p} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t+k-j} \right] \right|$$

$$\leq \mathbb{E}[Y_{t-k}^2] \mathbb{E}\left[\left(Y_t - \sum_{i\leq p} \psi_i \varepsilon_{t-i} \right)^2 \right] + \mathbb{E}[(\sum_{i\leq p} \psi_i \varepsilon_{t-i})^2] \mathbb{E}\left[\left(Y_{t-k} - \sum_{i\leq p} \psi_i \varepsilon_{t+k-i} \right)^2 \right]$$

Or, $\{\varepsilon_t\}_t$ est un bruit blanc gaussien centré de variance σ_ε^2. Donc :

$$\mathbb{E}[(\sum_{i\leq p}\psi_i\varepsilon_{t-i})^2] = \mathbb{V}[(\sum_{i\leq p}\psi_i\varepsilon_{t-i})] = \sum_{i\leq p}\psi_i^2\mathbb{V}[\varepsilon_{t-i}] = \sum_{i\leq p}\psi_i^2\sigma_\varepsilon^2 \leq \sum_{i=0}^{\infty}\psi_i^2\sigma_\varepsilon^2$$

Il vient :

$$\left|\mathbb{E}\left[Y_tY_{t-k} - \sum_{i,j\leq p}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t+k-j}\right]\right| \leq \mathbb{E}[Y_{t-k}^2]\mathbb{E}\left[\left(Y_t - \sum_{i\leq p}\psi_i\varepsilon_{t-i}\right)^2\right] + (\sum_{i=0}^{\infty}\psi_i^2\sigma_\varepsilon^2)\mathbb{E}\left[\left(Y_{t-k} - \sum_{i\leq p}\psi_i\varepsilon_{t+k-i}\right)^2\right]$$

D'où, par comparaison :

$$\lim_{p\to\infty}\mathbb{E}\left[\sum_{i,j\leq p}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t+k-j}\right] = \mathbb{E}[Y_tY_{t-k}]$$

Or pour $k \leq p$,

$$\mathbb{E}\left[\sum_{i,j\leq p}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t+k-j}\right] = \sum_{i,j\leq p}\psi_i\psi_j\mathbb{E}[\varepsilon_{t-i}\varepsilon_{t+k-j}] = \sum_{i,j\leq p}\psi_i\psi_j\delta_{k,j-i}\sigma_\varepsilon^2 = \sum_{k\leq j\leq p}\psi_j\psi_{j-k}\sigma_\varepsilon^2$$

Donc

$$\boxed{\mathbb{E}[Y_tY_{t-k}] = \sum_{j=\max(k,0)}^{\infty}\psi_j\psi_{j-k}\sigma_\varepsilon^2}$$

La moyenne du processus ne dépend pas de t et son autocorrelation ne dépend que de l'écart k entre les instants considérés, le processus est donc stationnaire au sens large.

Question b.

Par défintion de S et d'après la question précédente, il vient :

$$S(f) = \sum_{k=-\infty}^{\infty}\gamma[k]\exp(-i2\pi fk) = \sum_{k=-\infty}^{\infty}\sum_{j=\max(k,0)}^{\infty}\exp(-i2\pi fk)\psi_j\psi_{j-k}\sigma_\varepsilon^2$$

$$= \sigma_\varepsilon^2\sum_{k=-\infty}^{\infty}\sum_{j=\max(k,0)}^{\infty}\exp(-i2\pi fk)\psi_j\psi_{j-k} = \sigma_\varepsilon^2\sum_{p=0}^{\infty}\sum_{j=0}^{\infty}\exp(-i2\pi f(j-p))\psi_j\psi_p$$

$$= \sigma_\varepsilon^2\phi(e^{2\pi if})\phi(e^{-2\pi if}) = \sigma_\varepsilon^2|\phi(e^{-2\pi if})|^2$$

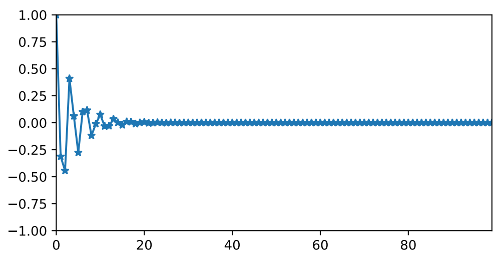
Question 3 *AR(2) process*

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

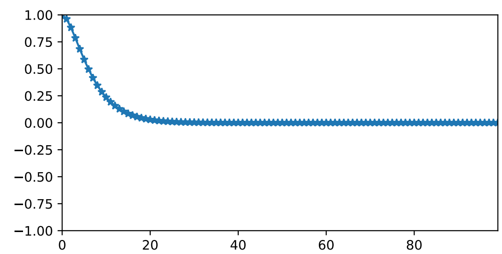
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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

Question a. Soit $\tau \geq 2$

$$\mathbb{E}[Y_t Y_{t-\tau}] = \phi_1 \mathbb{E}[Y_{t-1} Y_{t-\tau}] + \phi_2 \mathbb{E}[Y_{t-2} Y_{t-\tau}] + \mathbb{E}[\varepsilon_t Y_{t-\tau}]$$

Or, ε_t et $Y_{t-\tau}$ sont indépendant si $\tau \geq 1$ et $\mathbb{E}[\varepsilon_t] = 0$, d'où :

$$\gamma[\tau] = \phi_1 \gamma[\tau - 1] + \phi_2 \gamma[\tau - 2]$$

Donc il existe $C_1, C_2 \in \mathbb{C}$ tels que:

$$\gamma[\tau] = C_1 r_1^\tau + C_2 r_2^\tau$$

On utilise alors les valeurs de γ en $\tau = 0$ et $\tau = 1$ pour trouver C_1 et C_2 :

$$\begin{cases} C_1 + C_2 = \gamma[0] \\ C_1 r_1 + C_2 r_2 = \gamma[1] \end{cases}$$

On obtient comme $r_1 \neq r_2$: $C_1 = \frac{\gamma[1] - r_2\gamma[0]}{r_1 - r_2}$ et $C_2 = \frac{\gamma[1] - r_1\gamma[0]}{r_2 - r_1}$
Il vient finalement pour $\tau \geq 2$:

$$\boxed{\gamma[\tau] = \frac{1}{r_1 - r_2}[\gamma[0](r_1 r_2^\tau - r_2 r_1^\tau) + \gamma[1](r_1^\tau - r_2^\tau)]}$$

Question b. On observe des oscillations sur le graphique de gauche ce qui provient de la nature complexe des racines du polynômes caractéristiques. Celui de droite, sans oscillation, possède lui deux racines réels.

Question c. D'après le théorème de Wold, on peut écrire $\{Y_t\}_{t \geq 1}$ comme un terme déterministe et un terme stochastique.

D'après la question 2, on sait exprimer $S(f)$ en fonction des coefficients caractérisant le terme stochastique. Il suffit maintenant d'exprimer les coefficients du terme stochastique en fonction de ϕ.

Si on note $\mathcal{R}$ l'opérateur retard (c'est-à-dire : $\mathcal{R}(Y_t) = Y_{t-1}$), la relation $AR(2)$ se traduit par :

$$\phi(\mathcal{R})Y_t = \varepsilon_t.$$

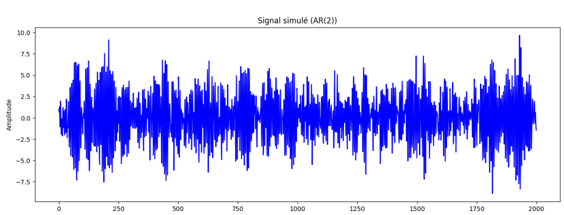
Ce qui revient à dire en inversant $\phi(\mathcal{R})$ (possible car le module des racines de ϕ est strictement plus grand que 1) :

$$Y_t = (\phi(\mathcal{R}))^{-1}\varepsilon_t$$

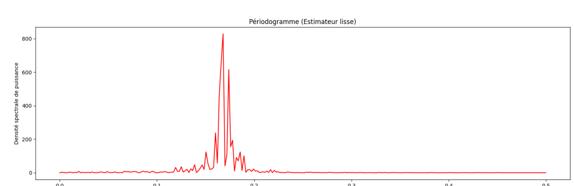
L'expression en série est obtenue en développant l'inverse de ϕ en série entière (décomposition en élément simple puis développement en série entière de chaque bloc comme les racines sont de module strictement plus grand que 1). Finalement, la série entière définie à la question 2 est donc $\frac{1}{\phi}$ ainsi :

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

Question d. On observe un pic à la fréquence correspond à $\frac{\theta}{2\pi} = \frac{1}{6}Hz$ ce qui est cohérent. L'amplitude est elle complètement aléatoire.



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

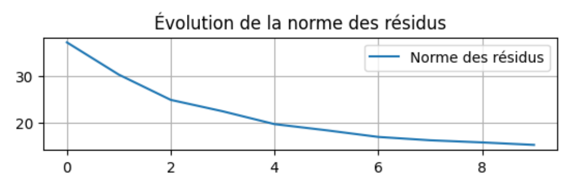
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

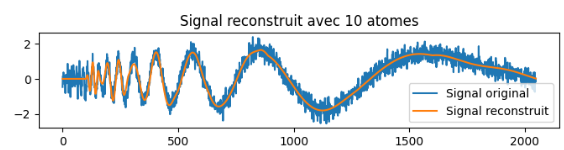
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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