

Assignment 1 (ML for TS) - MVA

Raphaël Faure raphael.faure@student-cs.fr
Victor J         victor.jesequel@gmail.com

October 28, 2024

1 Introduction

Objective. This assignment has three parts: questions about convolutional dictionary learning, spectral features, and a data study using the DTW.

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 28th October 23:59 PM.
- Rename your report and notebook as follows:
FirstnameLastname1_FirstnameLastname2.pdf and
FirstnameLastname1_FirstnameLastname2.ipynb.
For instance, LaurentOudre_CharlesTruong.pdf.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: [LINK](#).

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \quad (1)$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ the design matrix, $\beta \in \mathbb{R}^p$ the vector of regressors and $\lambda > 0$ the smoothing parameter.

Show that there exists λ_{\max} such that the minimizer of (1) is $\mathbf{0}_p$ (a p -dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

On a :

$$F = f + g$$

avec

$$F(\beta) = \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1,$$

$$f(\beta) = \frac{1}{2} \|y - X\beta\|_2^2,$$

et

$$g(\beta) = \lambda \|\beta\|_1.$$

\mathbb{R}^p est un espace de Hilbert. Nous avons :

$$\text{dom}(g) = \mathbb{R}^p = \text{dom}(F),$$

et

$$\text{int}(\text{dom}(g)) \cap \text{dom}(F) \neq \emptyset.$$

Ainsi :

$$\partial F = \partial f + \partial g,$$

et en particulier en 0 :

$$\partial F(0) = \partial f(0) + \partial g(0).$$

Or, f est différentiable en 0 et

$$\partial f(0) = \{\nabla f(0)\} = \{-X^T y\}.$$

et

$$\partial g(0) = \lambda \partial \|\beta\|_1(0),$$

avec

$$\partial \|\beta\|_1(0) = \{u \in \mathbb{R}^p \mid u_j \in [-1, 1] \forall j \in \{1, \dots, p\}\}.$$

Donc :

$$\partial F(0) = \{-X^T y + \lambda v, \quad v_j \in [-1, 1] \forall j\}$$

D'après la règle de Fermat :

$$0 \in \text{argmin } F \iff 0 \in \partial F(0) \iff \lambda > |(X^T y)_j| \quad \text{pour tout } j.$$

Ainsi, la plus grande valeur de λ est donnée par :

$$\boxed{\lambda_{\max} = \max_j |(X^T y)_j|}. \quad (2)$$

Question 2

For a univariate signal $\mathbf{x} \in \mathbb{R}^n$ with n samples, the convolutional dictionary learning task amounts to solving the following optimization problem:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k, \|\mathbf{d}_k\|_2 \leq 1} \left\| \mathbf{x} - \sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{z}_k\|_1 \quad (3)$$

where $\mathbf{d}_k \in \mathbb{R}^L$ are the K dictionary atoms (patterns), $\mathbf{z}_k \in \mathbb{R}^{N-L+1}$ are activations signals, and $\lambda > 0$ is the smoothing parameter.

Show that

- for a fixed dictionary, the sparse coding problem is a lasso regression (explicit the response vector and the design matrix);
- for a fixed dictionary, there exists λ_{\max} (which depends on the dictionary) such that the sparse codes are only 0 for any $\lambda > \lambda_{\max}$.

Answer 2

Le problème de codage parcimonieux se réécrit comme un Lasso standard, où la matrice de conception D est donnée par les convolutions des atomes d_k avec les coefficients z_k , comme suit :

La matrice D_k est définie de telle sorte que chaque coefficient $D_k[i, j]$ est donné par :

$$D_k[i, j] = \begin{cases} d_k[i - j] & \text{si } j + 1 \leq i \leq j + L, \\ 0 & \text{sinon.} \end{cases}$$

où i varie de 1 à n et j de 1 à $N - L + 1$.

Ainsi, $z_k * d_k$ peut être écrit sous la forme du produit matriciel suivant pour i varie de 1 à n :

$$z_k * d_k[i] = \sum_{j=1}^{N-L+1} z_k[j] d_k[i - j] = (D_k z_k)_i.$$

Nous pouvons alors écrire le problème d'optimisation complet comme suit :

$$\min_{z \in \mathbb{R}^{K(N-L+1)}} \|\mathbf{x} - D\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1$$

avec

$$D = [D_1, D_2, \dots, D_K] \in \mathbb{R}^{n \times K(N-L+1)}$$

où les D_k sont les matrices associées à la convolution $d_k * z_k$.

Ainsi, d'après ce qui précède, nous pouvons définir λ_{\max} comme :

$$\lambda_{\max} = \max_{1 \leq j \leq K(N-L+1)} |(D^T \mathbf{x})_j| = \max_{1 \leq k \leq K} \left(\max_{1 \leq j \leq N-L+1} |(D_k^T \mathbf{x})_j| \right) = \max_{1 \leq k \leq K} \left(\max_{1 \leq j \leq N-L+1} \left| \sum_{i=1}^n x[i] d_k[i - j] \right| \right)$$

3 Spectral feature

Let X_n ($n = 0, \dots, N-1$) be a weakly stationary random process with zero mean and autocovariance function $\gamma(\tau) := \mathbb{E}(X_n X_{n+\tau})$. Assume the autocovariances are absolutely summable, i.e. $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$, and square summable, i.e. $\sum_{\tau \in \mathbb{Z}} \gamma^2(\tau) < \infty$. Denote the sampling frequency by f_s , meaning that the index n corresponds to the time n/f_s . For simplicity, let N be even.

The *power spectrum* S of the stationary random process X is defined as the Fourier transform of the autocovariance function:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau / f_s}. \quad (4)$$

The power spectrum describes the distribution of power in the frequency space. Intuitively, large values of $S(f)$ indicate that the signal contains a sine wave at the frequency f . There are many estimation procedures to determine this important quantity, which can then be used in a machine-learning pipeline. In the following, we discuss the large sample properties of simple estimation procedures and the relationship between the power spectrum and the autocorrelation.

(Hint: use the many results on quadratic forms of Gaussian random variables to limit the number of calculations.)

Question 3

In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise.

- Calculate the associated autocovariance function and power spectrum. (By analogy with the light, this process is called “white” because of the particular form of its power spectrum.)

Answer 3

Comme X_n est un bruit blanc gaussien, X_n et $X_{n+\tau}$ sont indépendantes et d’espérance 0 ce qui donne :

$$\gamma(\tau) = \begin{cases} \sigma^2, & \text{si } \tau = 0, \\ 0, & \text{si } \tau \neq 0. \end{cases}$$

ou sous forme d’une fonction delta de Dirac :

$$\gamma(\tau) = \sigma^2 \delta(\tau),$$

où $\delta(\tau)$ est la fonction delta de Dirac.

Le spectre de puissance est donné par la transformée de Fourier de la fonction d’autocovariance :

$$S(f) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau / f_s}.$$

En utilisant $\gamma(\tau) = \sigma^2 \delta(\tau)$, on obtient :

$$\boxed{S(f) = \sigma^2.}$$

Le spectre de puissance constant signifie que toutes les fréquences ont la même contribution au signal, d'où l'appellation de bruit blanc, par analogie avec la lumière blanche, qui contient toutes les fréquences avec la même intensité

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \quad (5)$$

for $\tau = 0, 1, \dots, N-1$ and $\hat{\gamma}(\tau) := \hat{\gamma}(-\tau)$ for $\tau = -(N-1), \dots, -1$.

- Show that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$ but asymptotically unbiased. What would be a simple way to de-bias this estimator?

Answer 4

L'estimateur empirique de la fonction d'autocovariance est défini comme :

$$\hat{\gamma}(\tau) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}.$$

Calculons maintenant l'espérance de cet estimateur :

$$E[\hat{\gamma}(\tau)] = \frac{1}{N} \sum_{n=0}^{N-\tau-1} E[X_n X_{n+\tau}] = \frac{N-\tau}{N} \gamma(\tau).$$

Nous avons donc :

$$E[\hat{\gamma}(\tau)] = \frac{N-\tau}{N} \gamma(\tau) \neq \gamma(\tau)$$

Ainsi, $\hat{\gamma}(\tau)$ est un estimateur biaisé.

Comme :

$$E[\hat{\gamma}(\tau)] - \gamma(\tau) \rightarrow 0 \quad \text{lorsque } N \rightarrow \infty,$$

$\hat{\gamma}(\tau)$ est un estimateur asymptotiquement sans biais.

Pour débiaiser $\hat{\gamma}(\tau)$, il suffit de diviser par $N-\tau$ au lieu de N , d'où la définition de l'estimateur débiaisé :

$$\hat{\gamma}_{\text{sans biais}}(\tau) = \frac{1}{N-\tau} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}.$$

Question 5

Define the discrete Fourier transform of the random process $\{X_n\}_n$ by

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \quad (6)$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$ where $f_k = f_s k / N$. (They can be efficiently computed using the Fast Fourier Transform.)

- Write $|J(f_k)|^2$ as a function of the sample autocovariances.
- For a frequency f , define $f^{(N)}$ the closest Fourier frequency f_k to f . Show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Answer 5

$$|J(f_k)|^2 = \left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-2\pi i k n / N} \right) \left(\frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} X_m e^{2\pi i k m / N} \right),$$

car $f_k = \frac{k f_s}{N}$.

En multipliant ces deux sommes, on obtient :

$$|J(f_k)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m e^{-2\pi i k (n-m) / N}.$$

Maintenant, nous introduisons un changement de variable $\tau = n - m$, et réécrivons la somme double en fonction de τ :

$$|J(f_k)|^2 = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{n=\max(0,\tau)}^{N-1} X_n X_{n-\tau} \right) e^{-2\pi i k \tau / N}.$$

Sachant que $X_n X_{n-\tau}$ correspond à l'autocovariance échantillonnée $\hat{\gamma}(\tau)$, nous pouvons exprimer le périodogramme comme une somme en fonction de $\hat{\gamma}(\tau)$:

$$|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{-2\pi i k \tau / N}.$$

Ainsi, le périodogramme $|J(f_k)|^2$ est la transformée de Fourier discrète des autocovariances échantillonnées $\hat{\gamma}(\tau)$.

Nous avons de plus :

$$E \left(|J(f_k)|^2 \right) = \frac{1}{N} \sum_{l=-(N-1)}^{N-1} \left(\sum_{n=\max(0,l)}^{N-1-\max(0,l)} E[X_n X_{n-l}] \right) e^{\frac{2i\pi f l}{f_s}}$$

$$\begin{aligned}
&= \frac{1}{N} \sum_{l=-(N-1)}^{N-1} (N - |l|) \gamma(l) e^{\frac{2i\pi fl}{f_s}} \\
&= \sum_{l=-(N-1)}^{N-1} \frac{N - |l|}{N} \gamma(l) e^{\frac{2i\pi fl}{f_s}} \\
&= \sum_{l=-\infty}^{\infty} \frac{N - |l|}{N} \gamma(l) e^{\frac{2i\pi fl}{f_s}} \delta_{|l| < N-1}
\end{aligned}$$

Or

$$\left| \frac{N - |l|}{N} \gamma(l) \right| \delta_{|l| < N-1} \leq |\gamma(l)| \quad (\text{sommable})$$

D'après le théorème de convergence dominée, on obtient le résultat.

$$\boxed{\lim_{N \rightarrow \infty} E \left(|J(f_{(N)})|^2 \right) = \sum_{l=-\infty}^{\infty} \gamma(l) e^{\frac{2i\pi fl}{f_s}} = S(f)}.$$

Donc, $|J(f_{(N)})|^2$ est un estimateur asymptotiquement sans biais de $S(f)$.

Question 6

In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs τ) for 100 simulations of X . Plot the average value as well as the average \pm , the standard deviation. What do you observe?
- For $N \in \{200, 500, 1000\}$, compute the *periodogram* ($|J(f_k)|^2$ vs f_k) for 100 simulations of X . Plot the average value as well as the average \pm , the standard deviation. What do you observe?

Add your plots to Figure 1.

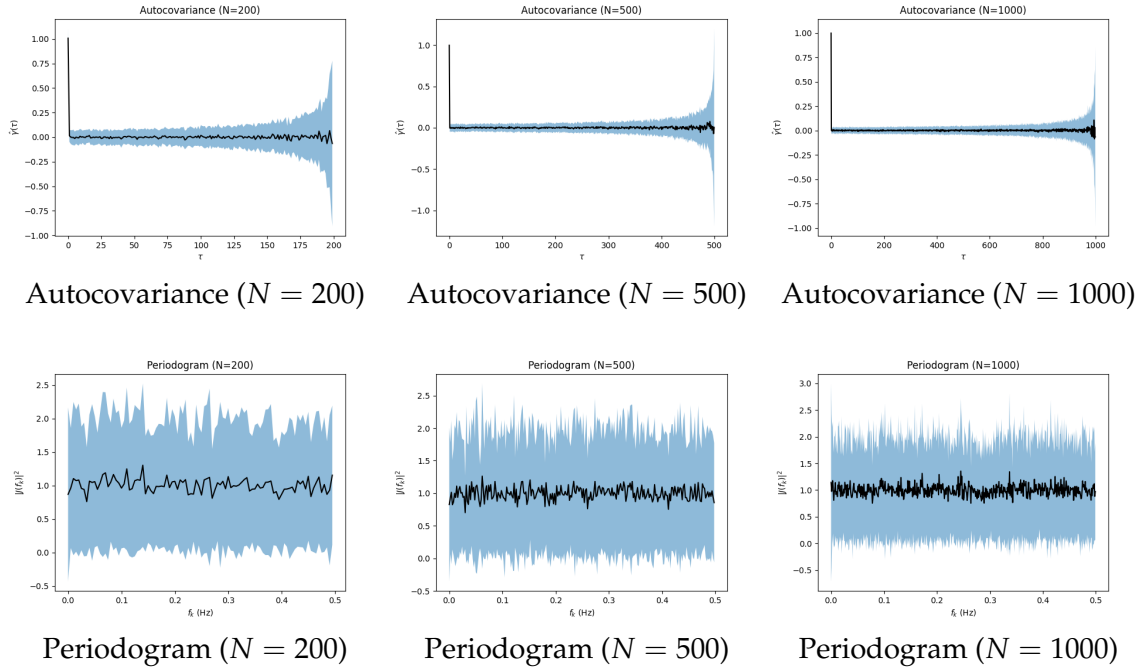


Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question ??).

Answer 6

Autocovariances :

On observe que, pour τ petit, les autocovariances restent proches de 0. Cependant, les valeurs fluctuent davantage pour des τ plus grandes, en particulier aux bords des séries temporelles. Cette augmentation des fluctuations est capturée par la bande d'incertitude (moyenne \pm l'écart-type) et devient plus importante à mesure que τ se rapproche de N .

Periodogrammes :

En moyenne, le périodogramme reste relativement stable autour de 1, reflétant la nature du bruit blanc gaussien (puisque son spectre de puissance est théoriquement plat). La variance du périodogramme ne décroît pas lorsque N augmente.

Question 7

We want to show that the estimator $\hat{\gamma}(\tau)$ is consistent, i.e. it converges in probability when the number N of samples grows to ∞ to the true value $\gamma(\tau)$. In this question, assume that X is a wide-sense stationary *Gaussian* process.

- Show that for $\tau > 0$

$$\text{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) [\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)]. \quad (7)$$

(Hint: if $\{Y_1, Y_2, Y_3, Y_4\}$ are four centered jointly Gaussian variables, then $\mathbb{E}[Y_1 Y_2 Y_3 Y_4] = \mathbb{E}[Y_1 Y_2] \mathbb{E}[Y_3 Y_4] + \mathbb{E}[Y_1 Y_3] \mathbb{E}[Y_2 Y_4] + \mathbb{E}[Y_1 Y_4] \mathbb{E}[Y_2 Y_3]$.)

- Conclude that $\hat{\gamma}(\tau)$ is consistent.

Answer 7

On a :

$$\text{Var}(\hat{\gamma}(\tau)) = E[\hat{\gamma}(\tau)^2] - E[\hat{\gamma}(\tau)]^2$$

Or :

$$E[\hat{\gamma}(\tau)] = \left(\frac{N-\tau}{N} \right) \gamma(\tau)$$

Calculons maintenant $E[\hat{\gamma}(\tau)^2]$:

$$\begin{aligned} E[\hat{\gamma}(\tau)^2] &= E \left(\frac{1}{N^2} \sum_{m=0}^{N-\tau-1} X_m X_{m+\tau} \sum_{m=0}^{N-\tau-1} X_m X_{m+\tau} \right) \\ &= \frac{1}{N^2} \sum_{m=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} E[X_m X_{m+\tau} X_m X_{m+\tau}] \end{aligned}$$

En utilisant l'espérance pour des processus gaussiens centrés, on obtient :

$$E[X_m X_{m+\tau} X_n X_{n+\tau}] = \gamma(\tau)^2 + \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)$$

On peut alors réécrire :

$$\begin{aligned} E[\hat{\gamma}(\tau)^2] &= \frac{1}{N^2} \sum_{m=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} (\gamma(\tau)^2 + \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)) \\ &= \frac{1}{N^2} \sum_{m=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} (\gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)) + \left(\frac{N-\tau}{N} \right) \gamma(\tau)^2 \end{aligned}$$

Ainsi :

$$\text{Var}(\hat{\gamma}(\tau)) = \frac{1}{N^2} \sum_{m=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} (\gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau))$$

On pose $k = n - m$. n et m varient de 0 à $N - \tau - 1$ donc k varie de $-(N - \tau - 1)$ à $N - \tau - 1$.

Nous avons maintenant :

$$E[\hat{\gamma}(\tau)^2] = \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N - \tau - |k|) (\gamma^2(\tau) + \gamma^2(k) + \gamma(k+\tau)\gamma(k-\tau)).$$

Pour chaque différence $k = n - m$, il y a $N - \tau - |k|$ paires (n, m) qui partagent la même valeur de k .

La variance de $\hat{\gamma}(\tau)$ devient alors :

$$\text{Var}(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau + |k|}{N}\right) (\gamma^2(k) + \gamma(k-\tau)\gamma(k+\tau)).$$

On a :

$$|\text{Var}(\hat{\gamma}(\tau))| \leq \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} [|\gamma^2(k)| + |\gamma(k-\tau)| |\gamma(k+\tau)|]$$

Or, comme la somme converge vers $\sum_l \gamma^2(l)$, lorsque $N \rightarrow \infty$, la variance $\text{Var}(\hat{\gamma}(\tau)) \rightarrow 0$.

Deplus, $E[\hat{\gamma}(\tau)] \rightarrow \gamma(\tau)$ donc $\hat{\gamma}(\tau)$ converge en moyenne quadratique donc en proba. C'est donc un estimateur consistant.

Contrary to the correlogram, the periodogram is not consistent. It is one of the most well-known estimators that is asymptotically unbiased but not consistent. In the following question, this is proven for Gaussian white noise, but this holds for more general stationary processes.

Question 8

Assume that X is a Gaussian white noise (variance σ^2) and let $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n / f_s)$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n / f_s)$. Observe that $J(f) = (1/N)(A(f) + iB(f))$.

- Derive the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, \dots, f_{N/2}$ where $f_k = f_s k / N$.
- What is the distribution of the periodogram values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$.
- What is the variance of the $|J(f_k)|^2$? Conclude that the periodogram is not consistent.
- Explain the erratic behavior of the periodogram in Question 6 by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

Moyenne et variances

$$A(f_k) = \sum_{n=0}^{N-1} X_n \cos\left(-2\pi \frac{kn}{N}\right).$$

Comme X_n est un bruit blanc gaussien avec $E[X_n] = 0$, nous avons :

$$E[A(f_k)] = E\left[\sum_{n=0}^{N-1} X_n \cos\left(-2\pi \frac{kn}{N}\right)\right] = \sum_{n=0}^{N-1} E[X_n] \cos\left(-2\pi \frac{kn}{N}\right).$$

$E[X_n] = 0$ pour tout n , donc :

$$\boxed{E[A(f_k)] = 0}.$$

De même :

$$E[B(f_k)] = E \left[\sum_{n=0}^{N-1} X_n \sin \left(-2\pi \frac{kn}{N} \right) \right] = \sum_{n=0}^{N-1} E[X_n] \sin \left(-2\pi \frac{kn}{N} \right) = 0.$$

Pour les calculs de variance :

$$\text{var}(A(f_k)) = E[A(f_k)^2] - (E[A(f_k)])^2.$$

Comme $E[A(f_k)] = 0$, la variance est simplement $E[A(f_k)^2]$:

$$A(f_k)^2 = \left(\sum_{n=0}^{N-1} X_n \cos \left(-2\pi \frac{kn}{N} \right) \right)^2.$$

$$A(f_k)^2 = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m \cos \left(-2\pi \frac{kn}{N} \right) \cos \left(-2\pi \frac{km}{N} \right).$$

En prenant l'espérance, et en utilisant le fait que $E[X_n X_m] = \sigma^2 \delta_{nm}$ (car X_n est un bruit blanc), nous obtenons :

$$E[A(f_k)^2] = \sum_{n=0}^{N-1} \sigma^2 \cos^2 \left(2\pi \frac{kn}{N} \right).$$

La somme des carrés des cosinus donne $\frac{N}{2}$, donc pour $k \neq 0$:

$$\boxed{\text{var}(A(f_k)) = \sigma^2 \cdot \frac{N}{2}.}$$

Des calculs similaires donnent, pour $k \neq 0$:

$$\boxed{\text{var}(B(f_k)) = \sigma^2 \cdot \frac{N}{2}.}$$

Pour le cas particulier $k = 0$, on obtient :

$$\text{var}(A(f_0)) = \sigma^2 \cdot N \quad \text{et} \quad \text{var}(B(f_0)) = 0.$$

Pour chaque fréquence $f_k = \frac{k f_s}{N}$, nous avons :

$$E[A(f_k)] = 0, \quad E[B(f_k)] = 0,$$

et

$$\text{var}(A(f_k)) = \frac{N\sigma^2}{2}, \quad \text{var}(B(f_k)) = \frac{N\sigma^2}{2}.$$

Ces résultats montrent que les moyennes sont nulles et que les variances sont proportionnelles à N , avec un facteur $\sigma^2/2$.

Distribution du périodogramme

Nous avons :

$$|J(f_k)|^2 = \frac{1}{N} (A(f_k)^2 + B(f_k)^2)$$

avec

$$A(f_k) \sim \mathcal{N}(0, \frac{N\sigma^2}{2}), \quad B(f_k) \sim \mathcal{N}(0, \frac{N\sigma^2}{2}).$$

Montrons que $A(f_k)$ et $B(f_k)$ sont indépendants. Par propriété du processus X_n , le vecteur aléatoire $(A(f_k), B(f_k))$ est gaussien. $A(f_k)$ et $B(f_k)$ sont indépendants, si et seulement si $A(f_k)$ et $B(f_k)$ sont non-corrélés.

Nous avons :

$$\text{cov}(A(f_k), B(f_k)) = E[A(f_k)B(f_k)] - E[A(f_k)]E[B(f_k)].$$

Or :

$$E[A(f_k)] = E[B(f_k)] = 0,$$

et

$$E[A(f_k)B(f_k)] = \sigma^2 \sum_{m=0}^{N-1} \cos\left(-2\pi \frac{km}{N}\right) \sin\left(-2\pi \frac{km}{N}\right).$$

En développant l'expression, nous obtenons :

$$E[A(f_k)B(f_k)] = \sigma^2 \sum_{m=0}^{N-1} \sin\left(-4\pi \frac{km}{N}\right) = 0.$$

Donc $A(f_k)$ et $B(f_k)$ sont indépendants.

Par conséquent, nous avons :

$$\left(\frac{1}{\sqrt{N}} \frac{\sqrt{2}}{\sigma}\right) A(f_k) \sim \mathcal{N}(0, 1), \quad \left(\frac{1}{\sqrt{N}} \frac{\sqrt{2}}{\sigma}\right) B(f_k) \sim \mathcal{N}(0, 1).$$

La somme du carré de deux variables gaussiennes indépendantes suit une loi du χ^2 à deux degrés de liberté, donc :

$$\frac{2}{\sigma^2} |J(f_k)|^2 \sim \chi^2_2.$$

Or la loi du χ^2_2 est une loi exponentielle de paramètre $\frac{1}{2}$, donc :

$$|J(f_k)|^2 \sim \sigma^2 \cdot \text{Exp}(1).$$

La variance de $|J(f_k)|^2$ est donc :

$$\boxed{\text{Var}(|J(f_k)|^2) = \frac{1}{\lambda^2} = \frac{1}{\left(\frac{1}{\sigma^2}\right)^2} = \sigma^4.}$$

Covariance

On a pour tout $i, j \in \{1, \dots, \frac{N}{2}\}$ avec $i \neq j$, les termes $J(f_i)$ et $J(f_j)$ sont indépendants. Cela découle directement du calcul de la covariance qui vaut 0 (comme $i + j \neq N$) et en remarquant que $J(f_i)$ et $J(f_j)$ sont des combinaisons linéaires de vecteurs gaussiens (car X_n est un bruit blanc), donc le vecteur $(J(f_i), J(f_j))$ est lui-même gaussien. Ainsi, si les variables sont décorréélées, elles sont indépendantes.

Donc, la covariance entre $|J(f_i)|^2$ et $|J(f_j)|^2$ est nulle :

$$\text{Cov}(|J(f_i)|^2, |J(f_j)|^2) = 0.$$

En effet, si deux variables aléatoires X et Y sont indépendantes, alors pour toutes fonctions mesurables f et g telles que $f(X)$ et $g(Y)$ sont intégrables (i.e., L^1), les variables $f(X)$ et $g(Y)$ sont également indépendantes.

Dans notre cas, en prenant $f = g = |\cdot|^2$, on obtient le résultat souhaité :

$$\boxed{\text{Cov}(|J(f_i)|^2, |J(f_j)|^2) = 0.}$$

Le comportement erratique du périodogramme est donc expliqué par l'absence de corrélation entre les valeurs pour des fréquences différentes.

Question 9

As seen in the previous question, the problem with the periodogram is the fact that its variance does not decrease with the sample size. A simple procedure to obtain a consistent estimate is to divide the signal into K sections of equal durations, compute a periodogram on each section, and average them. Provided the sections are independent, this has the effect of dividing the variance by K . This procedure is known as Bartlett's procedure.

- Rerun the experiment of Question 6, but replace the periodogram by Bartlett's estimate (set $K = 5$). What do you observe?

Add your plots to Figure 2.

Answer 9

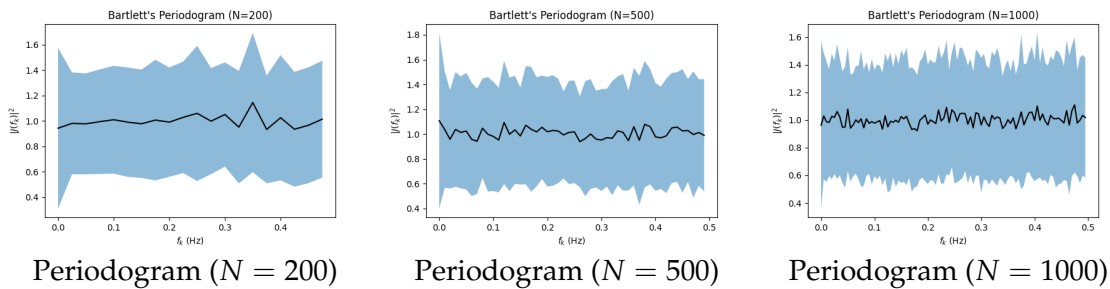


Figure 2: Bartlett's periodograms of a Gaussian white noise (see Question 6).

À mesure que N augmente, l'intervalle d'incertitude autour de la moyenne (représenté par la zone bleue) diminue. Cette réduction de la variance est conforme à l'effet attendu de la méthode de Bartlett, qui divise la variance par le nombre de sections K .

4 Data study

4.1 General information

Context. The study of human gait is a central problem in medical research with far-reaching consequences in the public health domain. This complex mechanism can be altered by a wide range of pathologies (such as Parkinson's disease, arthritis, stroke,...), often resulting in a significant loss of autonomy and an increased risk of falls. Understanding the influence of such medical disorders on a subject's gait would greatly facilitate early detection and prevention of those possibly harmful situations. To address these issues, clinical and bio-mechanical researchers have worked to objectively quantify gait characteristics.

Among the gait features that have proved their relevance in a medical context, several are linked to the notion of step (step duration, variation in step length, etc.), which can be seen as the core atom of the locomotion process. Many algorithms have, therefore, been developed to automatically (or semi-automatically) detect gait events (such as heel-strikes, heel-off, etc.) from accelerometer and gyrometer signals.

Data. Data are described in the associated notebook.

4.2 Step classification with the dynamic time warping (DTW) distance

Task. The objective is to classify footsteps and then walk signals between healthy and non-healthy.

Performance metric. The performance of this binary classification task is measured by the F-score.

Question 10

Combine the DTW and a k-neighbors classifier to classify each step. Find the optimal number of neighbors with 5-fold cross-validation and report the optimal number of neighbors and the associated F-score. Comment briefly.

Answer 10

D'après les résultats de la cross-validation, le nombre de voisins optimal est $n_{opt} = 7$ avec un F1-score moyen sur le train de 0.77 et un écart-type sur chaque pli de 0.08. Le F1-score obtenu sur le test set est de 0.49.

$$n_{opt} = 7 \quad \text{et} \quad F1_{\text{score}} = 0.49$$

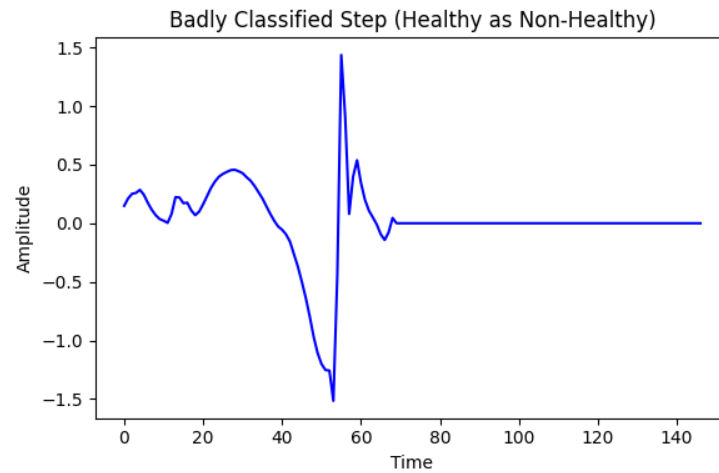
La valeur obtenue est faible comparée à la valeur obtenue sur les plis lors de la cross-validation. Cela laisse supposer une mauvaise généralisation des données d'entrée. Plusieurs raisons peuvent expliquer cela : un modèle trop complexe, un train set trop grand, ou des données de train et de test qui diffèrent énormément. Ce dernier point est d'autant plus plausible que la taille maximale

des signaux dans X_{train} est de 99 valeurs, alors qu'elle est de 147 valeurs dans X_{test} . Néanmoins le score reste correct.

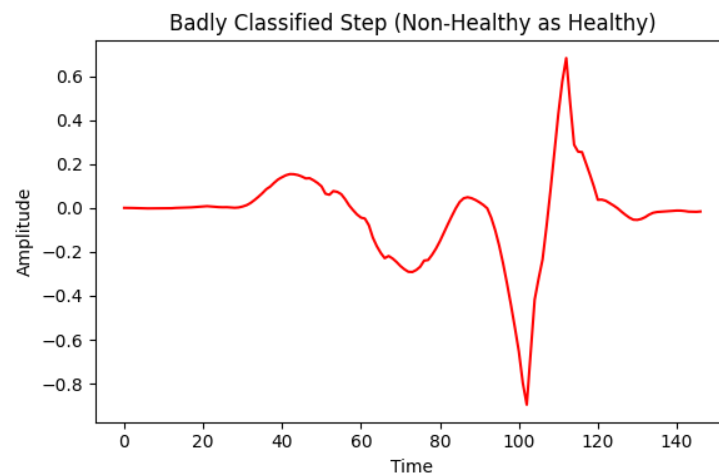
Question 11

Display on Figure 3 a badly classified step from each class (healthy / non-healthy).

Answer 11



Badly classified healthy step



Badly classified non-healthy step

Figure 3: Examples of badly classified steps (see Question 11).