



HAL
open science

Estimation paramétrique et non-paramétrique en utilisant une approche de régression quantile

Mohamed Aymen Meziani

► **To cite this version:**

Mohamed Aymen Meziani. Estimation paramétrique et non-paramétrique en utilisant une approche de régression quantile. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Paris-Est; Université des Sciences et de la Technologie Houari-Boumediène (Algérie), 2019. Français. NNT : 2019PESC0084 . tel-03467330

HAL Id: tel-03467330

<https://theses.hal.science/tel-03467330v1>

Submitted on 6 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ —
— PARIS-EST

THESE

EN VUE DE L'OBTENTION DU

DOCTORAT DE L'UNIVERSITE PARIS-EST

EN COTUTELLE AVEC

L'UNIVERSITE DE SCIENCE ET DE LA TECHNOLOGIE USTHB

Présentée et soutenue le 22/12/2019 par

Mohamed Aymen MEZIANI

Estimation paramétrique et non-paramétrique en utilisant une
approche de régression quantile

JURY

M. Roger Ceschi	Professeur	Rapporteur
M. Eric Monacelli	Professeur	Rapporteur
M. Zaher Mohdeb	Professeur	Rapporteur
Mme Khedidja Djaballah	Professeur	Examinatrice
Mme Ourida Sadki	Professeur	Examinatrice
M. Abdelghani Chibani	Maître de conférences	Examineur
M. Karim Djouani	Professeur	Directeur de thèse
M. Tarek Medkour	Maître de conférences	Co-Directeur de thèse

Ecole Doctorale : Mathématiques et Sciences et Technologies de l'Information et de la
Communication (MSTIC)

Spécialité : Signal, Image et Automatique

Laboratoire d'Accueil : Laboratoire Images, Signaux et Systèmes Intelligents (LISSI)

Résumé

Le quantile périodogramme développé dans [52] est une nouvelle approche qui fournit une information plus étendue et plus riche par rapport au périodogramme ordinaire. Cependant, il souffre d'instabilité de performance dans le cas de présence de plusieurs périodicités dominantes ou dans le cas de données bruitées. Cela peut être dû à la fuite spectrale (Leakage) produite sous forme de pics supplémentaires. Afin de remédier à ce problème, une version régularisée du quantile périodogramme est proposée. Les propriétés asymptotiques du nouvel estimateur sont développées. De plus, des simulations approfondies sont effectuées pour montrer l'efficacité de l'estimation proposée en termes de détection de périodicités cachées sous différents types de bruit.

Une première application de l'approche proposée a été conduite dans un cadre d'étude des signaux EEG (électroencéphalogramme). Les signaux EEG sont connus pour leurs non-stationnarité et non-linéarité. Le nouvel estimateur proposé a été utilisé ainsi que d'autres estimateurs spectraux comme étant des méthodes d'extraction de caractéristiques. Ces caractéristiques sont introduites par la suite à des classifieurs pour déterminer si le signal appartient à une des classes d'imagerie motrice.

La deuxième application est un travail qui a été fait dans le cadre du projet Européen MEDOLUTION pour l'étude des signaux accéléromètres. Le quantile périodogramme pénalisé ainsi que d'autres estimateurs spectraux ont été appliqués pour la classification des mouvements d'auto rééducation. Les résultats suggèrent que le quantile périodogramme pénalisé est une méthode prometteuse et robuste pour détecter des périodicités cachées, spécialement dans les cas de non-stationnarité et non-linéarité des données, ainsi que pour améliorer les performances de classification dans le cadre des signaux EEG et des signaux accéléromètres.

Abstract

The present thesis extends the quantile periodogram developed in [52] as it suffers from instabilities and spectral leakage. To overcome these drawbacks, a L_1 regularised version is proposed. The asymptotic properties of the new spectral estimator are discussed and simulations are performed in order to show its efficiency in the presence of hidden periodicities and under different types of noise.

A first application of the proposed estimator has been conducted in an EEG-based BCI study. The EEG signals are known for their non-stationarity and non-Gaussianity. In order to classify the motor imagery tasks, the proposed approach has been used along with other spectral estimators as feature extraction methods.

A dataset collected within the framework of the European MEDOLUTION project has been utilised for the second application. Motivated by the non-stationary nature of the 3D accelerometers signals, the penalised quantile periodogram has been applied for the recognition of self-rehabilitation exercises. A comparative study has been conducted between the newly introduced approach and the classical spectral estimators through four machine learning classifiers.

The results suggest that the penalised quantile periodogram is a promising and robust technique for the detection of hidden periodicities. Especially, in the case of non-stationary and non-linear signals. In addition, It can improve the classification performance in the case of EEG and 3D accelerometer signals.

Remerciements

Je tiens tout d'abord à remercier plus amplement mes directeurs de thèse M. Tarek MEDKOUR et M. Karim DJOUANI pour l'honneur qu'ils m'ont fait en acceptant la charge de m'encadrer en me consacrant beaucoup de leurs temps précieux, pour leur conseils et soutien pendant l'élaboration de ce travail.

Je voudrais remercier les rapporteurs de cette thèse M. Roger CESCHI professeur à South African Institute of Technology , M. Zaher MOHDEB professeur au Polytechnique de Constantine et M. Eric MONACELLI professeur à l'université de Versailles Saint-Quentin-en-Yvelines pour l'intérêt qu'ils ont porté à ce travail.

J'associe à ces remerciements Mme. Khedidja DJABALLAH professeur à l'université de l'USTHB, Mme Ourida SADKI professeur à l'USTHB et M. Abdelghani CHIBANI Maître de conférences à l'université de Paris-Est Créteil pour avoir accepté d'examiner mon travail.

Je souhaite aussi exprimer ma gratitude à l'ensemble des membres de la faculté de Mathématiques de l'université de l'USTHB en Algérie pour leur soutien logistique et moral que j'ai trouvée pendant toute la période de préparation de cette thèse. Je remercie également l'ensemble des membres du laboratoire LISSI en France pour l'intérêt dont ils ont fait preuve envers ma recherche, les moyens qu'ils ont mis en œuvre pour me donner accès aux données réels, plus particulièrement, je tiens à exprimer ma reconnaissance à M. Yacine AMIRAT directeur du laboratoire et M. Abdelghani CHIBANI.

Ces remerciements ne peuvent s'achever, sans une pensée pour ma première fan (et correctrice des fautes d'orthographe de cette thèse!) : ma mère. Sa présence et ses encouragements sont pour moi les piliers fondateurs de ce que je suis et de ce que je fais. Je dédie cette thèse à elle, à mon père, mes deux frères Rami et Abderrahman, mes deux sœurs, Randa et Nabila et ma chère fiancée. Enfin, je voudrais dire merci à tous ceux, famille ou amis, qui m'entourent et à l'ensemble des personnes présentes à la soutenance.

'Truth is ever to be found in the simplicity, and not in the multiplicity and confusion of things.'

Isaac Newton

Table des matières

Abstract	3
Abstract	4
Remerciements	5
1 Introduction	1
2 Préliminaires	5
2.1 Analyse spectrale	5
2.1.1 La périodicité et la représentation spectrale	7
2.1.2 La densité spectrale	9
2.1.3 L'estimation spectrale	11
2.1.4 Le périodogramme	12
2.2 La régression quantile	17
2.2.1 La régression linéaire et la régression quantile	17
2.2.2 Estimation des paramètres	23
2.3 Régression régularisée	25
2.3.1 Inférence	26
2.3.2 La régression quantile régularisée	28

3	Quantile périodogramme pénalisé pour l'estimation spectrale	31
3.1	Le quantile périodogramme	31
3.1.1	Propriétés asymptotiques	33
3.2	Le quantile périodogramme pénalisé	36
3.3	Propriétés asymptotiques	37
3.4	Simulation	39
3.5	Conclusion	49
3.6	Appendice	49
3.6.1	Théorème de régression quantile régularisé	49
3.6.2	Preuve du théorème 4	53
4	Le quantile périodogramme pénalisé pour la détection des patterns dans les signaux EEG	57
4.1	Introduction	57
4.2	Activité cérébrale et les systèmes BCI	60
4.2.1	Electroencephalography	60
4.2.2	L'imagerie motrice et la désynchronisation liée à un événement (ERD)	63
4.3	Description du dataset	65
4.4	Analyse des signaux EEG	66
4.5	La chaîne de traitement de signal	68
4.6	Les algorithmes de classification	72
4.7	Expérimentations et étude comparative	77
4.8	Conclusion	84

5	Lasso quantile périodogramme pour la reconnaissance d'activités humaines	87
5.1	Introduction	88
5.2	Contexte de l'assistance de la vie quotidienne	89
5.3	Capteurs portables et acquisition de données	90
5.4	Description du jeu de données	91
5.5	Analyse des signaux accélérométriques	93
5.6	Processus d'apprentissage automatique	96
5.7	Résultats et discussion	103
5.8	Conclusion	105
6	Conclusion	107
	Bibliography	108

Liste des tableaux

3.1	Puissance de test de Fisher pour 1000 simulations de Monte Carlo pour la série temporelle définie dans le premier exemple en utilisant une périodicité ($\omega_0 = 0.1 \times 2\pi$).	44
3.2	Puissance de test de Fisher pour le quantile périodogramme pénalisé avec différentes valeurs de régularisation. 1000 simulations utilisées avec le quantile $\tau = 0.8$	46
3.3	Puissance de test de Fisher pour 1000 simulations de Monte Carlo pour la série temporelle définie dans le premier exemple en utilisant deux périodicités ($\omega_0 = 0.1 \times 2\pi$, $\omega_1 = 0.12 \times 2\pi$).	46
4.1	Résultats de 4 estimateurs spectraux différents, avec 4 classifieurs	79
4.2	Quantile périodogramme avec différents quantiles en utilisant les données du premier sujet	80
4.3	Lasso quantile Périodogramme avec différentes valeurs de α en utilisant les données du premier sujet	81
5.1	Les résultats de classification pour les données d'auto-exercices	104

Table des figures

2.1	Signaux périodiques de l'exemple 1.	13
2.2	Périodogramme du signal à 3 période de l'exemple 1.	14
2.3	Régression linéaire de l'ensemble de données des accidents aux États-Unis (Matlab)	19
2.4	Croissance de poids selon l'age ([20]).	20
2.5	Dataset généré a partir d'un modèle linéaire hétéroscédastique ([7]).	21
2.6	Régression MCO et quantiles pour un modèle linéaire hétéroscédastique [7].	22
2.7	Données avec des Valeurs aberrantes [7].	22
2.8	Régression MCO et quantiles pour un dataset avec des valeurs aberrantes [7].	23
2.9	La fonction de perte ρ_τ	24
2.10	les formes géométriques des fonctions de régularisation, Lasso à gauche et Ridge à droite.	25
3.1	Moyenne de l'ensemble de 1000 périodogrammes lissés d'un processus AR(2) définie dans (3.14) avec un bruit blanc. (a) montre le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La taille de l'échantillon utilisée est $n = 200$. Le quantile $\tau = 0.8$	41

3.2	Estimateurs de la densité spectrale pour trois types de bruit pour une seule réalisation pour le modèle défini par (3.13). (a) montre le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La ligne pointillée (.) est pour le bruit gaussien standard, la ligne (-) est pour le bruit gaussien de grande variance et la ligne solide est pour le bruit de student-t. La taille de l'échantillon utilisé est $n = 200$. Le quantile utilisé est $\tau = 0.8$	42
3.3	Moyenne de 1000 réalisations de densités spectrales estimées en utilisant le bruit de student-t pour un processus AR (2) simple (équation (3.14)) (a) représente le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La taille de l'échantillon utilisé est $n = 200$. Le quantile utilisé $\tau = 0.8$	43
3.4	La distribution cumulative de l'ALQP et du QP calculés pour 1000 simulations de bruit blanc standard à une fréquence fixe.	45
3.5	séries chronologiques de nombre annuel de taches solaires avec bruit de student-t supplémentaire	47
3.6	Estimateurs spectraux pour les nombres annuels de taches solaires avec ajout de bruit student-t. (a) représente le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé	48
4.1	La région d'intérêt dans un système à 32 électrodes ([59]).	58
4.2	Un exemple illustratif d'un système BCI exogène ([59]).	61
4.3	les différentes parties du cerveau humain.	62
4.4	Les rythmes de l'activité cérébrale ([59])	63
4.5	ERD et ERS durant l'imagerie motrice [68]	64
4.6	Résumé graphique de l'expérimentation.	65
4.7	Scénario Off-ligne	66

4.8 La médiane et le 0.95-quartile des différents estimateurs spectraux pour chaque tache (sujet2, électrode C3). Pour les paramètres, (a) représente la FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO. Les paramètres ($\tau = 0.5, \alpha = 5$) ont été utilisés respectivement comme le quantile d'intérêt et le paramètre de régularisation pour les estimateurs quantiles. 68

4.9 La médiane des Quantiles périodogrammes du LASSO pour la classe d'imagerie motrice de la main droite (électrode C3) pour tous les essais. (a) représente la médiane du Quantile périodogramme du LASSO appliqué sur les données du sujet 2 avec les quantiles (0,25, 0,5, 0,75 et 0,95). (b) représente le Quantile périodogramme du LASSO appliqué sur les données du sujet 5 avec les quantiles (0,25, 0,5, 0,75 et 0,95). $\alpha = 5$ a été utilisé comme le paramètre de régularisation. 69

4.10 La médiane des différents estimateurs spectraux appliqués pour les classes d'imagerie motrice de la main droite et la main gauche (sujet 2, électrode C3) pour tous les essais. (a) représente la FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO. Les paramètres ($\tau = 0.5, \alpha = 5$) ont été utilisés respectivement comme le quantile d'intérêt et le paramètre de régularisation pour les estimateurs Quantile. 70

4.11 La chaine de traitement de Signal 71

4.12 K plus proches voisins 73

4.13 Support vector machine optimal hyperplane 74

4.14 L'algorithme de l'analyse discriminante linéaire (LDA) 76

4.15 Exemple d'une arbre de décision simple 77

4.16 Taux de précision pour l'algorithme de Knn avec différents estimateurs spectraux pour tous les sujets. Performance de session à session. 82

4.17 Taux de précision pour l'algorithme de Foret aléatoire avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.	83
4.18 Taux de précision pour l'algorithme d'analyse discriminante linéaire avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.	83
4.19 Taux de précision pour l'algorithme de Machine à vecteurs de support avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.	84
5.1 Illustration graphique pour un système d'assistance de la vie quotidienne [83]	89
5.2 Illustration graphique pour l'emplacement de capteurs [8]	91
5.3 Positions des exercices pour le protocole de rééducation	91
5.4 Emplacement des appareils	92
5.5 5 répétition pour les activités 1,2 et 7. Les capteurs d'accélération ont été utilisés en choisissant les données de l'axe x. (a) représente les données de l'exercice 1, (b) les données de l'exercices 2 et (c) les données de l'exercice 7	94
5.6 Le phénomène de Gibbs. (a) représente l'approximation de Fourier, (b) est la transformée de Fourier du signal à ondes carrés.	95
5.7 La médiane et l'intervalle de confiance des différents estimateurs spectraux sur toutes les répétitions de l'exercice 1, l'axe x du capteur de la main droite est utilisé. (a) représente les réalisations du FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO	97

5.8	La médiane du Quantile périodogrammes du LASSO effectué sur les données de l'exercice 1 et 4. l'axe x du capteur de la main droite a été utilisé. (a) représente la réalisation de l'exercice 1 et (b) la réalisation de l'exercice 4.	98
5.9	La médiane des différents estimateurs spectraux considérés appliqués sur les données des exercices 1,4 et 7. L'axe x du capteur de la main droite est utilisé pour toutes les expériences. (a) représente les réalisation du FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO.	99
5.10	Processus de reconnaissance d'activités	100
5.11	La matrice de confusion pour le modèle de réseau de neurone en utilisant les caractéristiques du Quantile périodogramme	104

Chapitre 1

Introduction

L'analyse des séries temporelles vise à développer des moyens quantitatifs permettant de caractériser et de modéliser des phénomènes qui se passent dans le temps. Une série temporelle est tout ensemble d'observations dans le temps. En effet, elle a de nombreuses applications importantes en économie, en astronomie, en médecine et beaucoup d'autres domaines.

Par ailleurs, l'analyse de comportement périodique d'une série est importante pour comprendre certains phénomènes.

Dans le but d'analyser la périodicité d'un processus, l'approche la plus commune est l'estimation de ce qu'on appelle la densité spectrale. Le problème d'estimation du spectre a été toujours un des sujets les plus importants dans le domaine d'analyse des séries temporelles.

Traditionnellement, l'analyse spectrale est connue comme la décomposition de la fonction d'autocovariance en somme des sinus et cosinus, une mesure de dépendance linéaire des observations à des moments différents. Le succès du périodogramme dans l'analyse des séries temporelles est lié à son calcul rapide et simple (à travers l'algorithme Fast Fourier transform) et son interprétation au sens de comportement cyclique, qui est beaucoup plus efficace que la représentation dans le domaine temporel. Le périodogramme est couramment utilisé comme un estimateur non-paramétrique du spectre.

Il peut être défini soit par la transformée de Fourier discrète de la fonction d'auto-covariance empirique, ou à travers une projection L^2 des observations dans une base harmonique (voir [13, 17]).

Être connecté à l'autocovariance peut être bénéfique pour avoir des caractéristiques intéressantes (comme l'optimalité en analyse des séries Gaussiennes). Cependant, il existe aussi des inconvénients à cette connexion, par exemple la non-robustesse par rapport à des données aberrantes ou des données bruitées. De plus, l'estimateur ne peut pas capturer des caractéristiques dynamiques importantes, comme le changement dans la forme conditionnelle (asymétrie, kurtosis), la dépendance dans les extrêmes et la dépendance non-linéaire. Cela a été réalisé par beaucoup de recherches dans la littérature, et diverses extensions et modifications ont été proposées afin de remédier à ces inconvénients.

L'utilisation de la régression quantile en estimation spectrale a attiré beaucoup d'attention après sa première proposition par Koenker[45].

Dans un premier temps, Li[51] a proposé d'utiliser la méthode des moindres écarts absolus comme une alternative des moindres carrés ordinaires dans un modèle de régression harmonique. Il a construit un estimateur asymptotiquement non biaisé, après en 2012, il a étendu ses résultats à un quantile quelconque [52]. Le quantile périodogramme résume le comportement cyclique du signal pas seulement autour de la moyenne, mais surtout sur l'ensemble de la distribution. Cela nous permet de voir quelle partie de distribution représente mieux l'information.

Divers travaux récents ont utilisé la régression quantile dans l'analyse spectrale. Hagemann[37] a proposé un nouveau spectre appelé Quantile Spectra, il a utilisé une transformation de la série temporelle en utilisant une fonction quantile, puis il a appliqué le spectre de puissance ordinaire. Il a construit un estimateur consistant sans impliquer le terme de distribution marginale qui a été produit dans l'approche de Li[51]. Dans [23] et [47], les auteurs ont introduit une étude similaire basée sur les copules dans le domaine fréquentiel. Inspiré par le quantile périodogramme établi par Li[52], les auteurs

dans [54] ont proposé de faire la somme sur les quantiles, pour surmonter le problème du choix du quantile d'intérêt. Dans le domaine temporel, le quantilogramme a été introduit dans [55] pour tester la prévisibilité directionnelle dans le domaine temporel.

Les techniques de régularisation ont été proposées notamment dans la littérature pour l'estimation spectrale. Sous le terme de Basis pursuit denoising, La norme L_1 a été utilisée comme une technique de régularisation dans l'estimation spectrale classique. Dans [18] et [42], les auteurs ont obtenu une estimation robuste du spectre contre les fuites spectrale (leakage), les données bruitées et inégales (uneven data).

Motivé par la fuite spectrale produite par le quantile périodogramme, cette thèse introduit une nouvelle approche appelée le quantile périodogramme pénalisé, qui est une version régularisée du quantile périodogramme de Li[52]. En utilisant une norme L_1 , le quantile périodogramme pénalisé peut être considéré comme une représentation creuse (sparse) du spectre du quantile. Nous montrons que l'approche proposée pourrait également être utilisée comme représentation de la dépendance en série dans le domaine fréquentiel, de même que le périodogramme ordinaire et le quantile périodogramme. De plus, nous montrons que le quantile périodogramme pénalisé pourrait réduire la fuite spectrale produite par le quantile périodogramme en réduisant les éléments bruyants vers zéro. Ceci est démontré en utilisant le test de Fisher des périodicités cachées. Nous montrons également que le quantile périodogramme pénalisé est lié au spectre de passage à niveau. Ainsi, il partage l'invariance contre les distorsions non linéaires.

La thèse est organisée comme suite. Dans un premier temps, une courte présentation du contexte et des termes utilisés dans le document est donnée dans le chapitre 2. Dans le 3ème chapitre, notre contribution est introduite. Dans le chapitre suivant, une application de l'approche proposée est présentée dans le cadre d'une étude des signaux EEG. L'application du quantile périodogramme pénalisé pour la reconnaissance d'activités humaines est introduite dans le chapitre 5. Finalement, des conclusions et futures travaux sont discutés dans le chapitre 6.

Chapitre 2

Préliminaires

“The book of nature is written in the language of Mathematics.”

– Galileo

Ce chapitre introduit des terminologies de base pour l’analyse spectrale. Une discussion plus détaillée peut être trouvée dans [13, 17, 78, 90]

2.1 Analyse spectrale

L’analyse spectrale est une des techniques les plus utilisées dans le domaine de l’analyse des séries temporelles. Elle est motivée par le fait que la périodicité est le comportement le plus régulier et prévisible d’une série temporelle. Cette approche procède ensuite à la détermination des composantes périodiques incorporées dans la série temporelle en calculant les périodes, les amplitudes et les phases associées[13].

Dans ce qui suit, une introduction de l’analyse spectrale est donnée en commençant par présenter quelques notions importantes d’algèbre linéaire.

Définition 1 ([17]) *Une base de \mathbb{C}^n est une famille d’éléments de \mathbb{C}^n tel que chaque*

élément de cette famille peut s'écrire comme une combinaison linéaire unique des éléments de cette base.

Le cas idéal pour cette définition est la base canonique de \mathbb{C}^n , B

$$B = (1, 0, \dots, 0)^T, (0, 1, \dots, 0)^T, \dots, (0, 0, \dots, 1)^T.$$

On peut écrire de façon équivalente, B est l'ensemble des vecteurs $\{e_i\}$ tels que

$$\forall k \in \{1, \dots, n\}, \forall j \in \{1, \dots, n\}, \quad e_k = 1_{k=j}.$$

Si x est un élément de \mathbb{C}^n , alors il existe une combinaison linéaire de base B tel que

$$x = \sum_{k=1}^n x_k e_k.$$

L'espace \mathbb{C}^n est muni du produit scalaire

$$\forall x, y \in \mathbb{C}, \quad \langle x, y \rangle = \sum_{k=1}^n \bar{x}_k y_k,$$

et de la norme correspondante

$$\forall x, y \in \mathbb{C}, \quad \|x\| = \sqrt{\sum_{k=1}^n |x_k|^2}.$$

On dit qu'une famille de vecteurs est orthogonale, si ses éléments sont deux à deux orthogonaux. Deux vecteurs sont dits orthogonaux si leur produit scalaire est nul. On appelle donc une base orthonormée, une base orthogonale et de plus ses éléments sont normés. Un vecteur est dit normé si sa norme vaut 1.

Proposition 1 ([17]) Une famille orthonormée de n vecteurs de \mathbb{C}^n (respectivement \mathbb{R}^n) est une base de \mathbb{C}^n (respectivement \mathbb{R}^n)

Base de Fourier

Considérons l'espace de Hilbert $L^2[-\pi, \pi] = L^2([-\pi, \pi], \mathbb{B}, U)$, où \mathbb{B} est construite de sous ensembles de $[-\pi, \pi]$, et U représente la probabilité uniforme. On définit la famille de fonctions suivante de l'espace $L^2[-\pi, \pi]$

$$\forall j \in \mathbb{Z}, \quad e_j(x) = e^{ijx},$$

avec $i^2 = -1$, e est la fonction exponentielle. Cette famille est une base orthonormée sur $L^2[-\pi, \pi]$.

Définition 2 ([17]) *Une approximation de Fourier d'une fonction f de $L^2[-\pi, \pi]$, n'est autre que la projection de cette fonction dans le sous espace engendré par la base $\{e_j, \quad j \in \mathbb{Z}\}$.*

On définit alors la transformée de Fourier de f par

$$S_n(f) = \sum_{j=-n}^n \langle f, e_j \rangle e_j,$$

tels que,

$$\langle f, e_j \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx,$$

sont appelés les coefficients de Fourier de la fonction f et $x \in [-\pi, \pi]$.

2.1.1 La périodicité et la représentation spectrale

La périodicité est un des phénomènes naturels les plus importants et les plus utiles. La terre tourne périodiquement autour de son propre axe, et autour du soleil en nous donnant les jours et les saisons, les mouvements du soleil, les mouvements de la lune, le rythme cardiaque, les ondes cérébrales. Tous ces phénomènes montrent un pattern périodique au fil du temps.

Par définition, une fonction périodique est une fonction qui se répète sur des intervalles

avec une longueur fixe appelée période [90].

Afin de définir le taux pour lequel la série oscille, on commence tout d'abord par définir un cycle comme une période complète d'une fonction sinus ou cosinus définie sur une unité de temps.

Considérons le processus périodique

$$x_t = A \cos(2\pi\omega t + \phi), \quad (2.1)$$

pour $t \in \mathbb{Z}$, avec ω est la fréquence, définie par cycle par unité de temps, A détermine l'amplitude de la fonction et ϕ est la phase, qui détermine le point de départ de la fonction cosinus.

Aux fins d'analyse de données, on peut utiliser l'identité trigonométrique et écrire l'équation (2.1) comme suit

$$x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t), \quad (2.2)$$

tels que, $U_1 = A \cos(\phi)$ et $U_2 = -A \sin(\phi)$ sont souvent prises comme des variables aléatoires normales.

Une généralisation de (2.2) est donnée selon le théorème de Dirichlet. Pour chaque série temporelle à valeurs réelles $\{Y_t\}$, il existe une représentation de série de Fourier

$$Y_t = \mu + \sum_{j=1}^n [A_j \cos(2\pi\omega_j t) + B_j \sin(2\pi\omega_j t)], \quad \omega_j = j/n, \quad (2.3)$$

avec, A_j et B_j sont non corrélés avec moyenne nulle et variance σ_j^2 , les ω_j sont des fréquences différentes [17]. Cette représentation permet des mélanges de fonctions périodiques avec différentes fréquences, amplitudes et phases. Une estimation de moindres carrés peut être donnée pour modéliser une série temporelle en utilisant la représentation (2.3). Notons que n'importe quelle série temporelle stationnaire, soit déterministe, stochastique, avec ou sans périodicité, elle peut être modélisée par le modèle (2.3) [89].

La représentation spectrale d'un processus aléatoire est donc définie comme une décomposition en une somme de composantes sinusoïdales avec des coefficients aléatoires non corrélés. Cette représentation est équivalente à l'analyse dans le domaine temporel, mais elle fournit une façon différente de traitement du processus, qui est dans certaines applications plus efficace.

Étant donné que (2.3) est vraie pour toute série temporelle stationnaire, il existe notamment une représentation spectrale pour la fonction d'autocovariance de la série. Si x_t est un processus stationnaire, avec une fonction d'autocovariance $\gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)]$, alors il existe une fonction unique monotone croissante $F(\omega)$, appelée la fonction de distribution spectrale du processus $\{Y_t\}$, qui est bornée, avec $F(-\infty) = F(-\pi) = 0$, et $F(+\infty) = F(\pi) = \gamma(0)$, tel que

$$\gamma(h) = \int_{-\pi}^{\pi} \exp^{2\pi i \omega h} dF(\omega), \quad (2.4)$$

ou F est définie par :

$$F(\omega) = \sum_{j:\omega_j \leq \omega} \sigma_j^2.$$

La fonction $F(\omega)$ se comporte comme une fonction de distribution cumulative pour une variable aléatoire discrète, sauf pour $F(\infty) = \sigma^2 = \text{var}(x_t)$ à la place de 1.

Un résultat très important qui suit la représentation (2.4), est celui qui a été présenté dans le théorème de Herglotz [17]. En supposant que la fonction d'autocovariance est absolument sommable, la fonction de distribution cumulative est absolument continue avec $dF(\omega) = f(\omega)d\omega$. La fonction f est appelée la densité spectrale. Ce résultat est la motivation pour introduire la fonction de densité spectrale.

2.1.2 La densité spectrale

La densité spectrale est un outil mathématique permettant de quantifier la distribution d'énergie d'une série temporelle, ou encore détecter la périodicité cachée d'un processus

périodique. La densité spectrale d'un processus contient des informations essentielles sur sa nature.

En utilisant la représentation (2.4), et le théorème de Herglotz, si :

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

alors, la fonction d'autocovariance peut être écrite comme suit :

$$\gamma(h) = \int_{-\pi}^{\pi} \exp^{2\pi i\omega h} f(\omega) d\omega,$$

comme la transformée de Fourier inverse de la densité spectrale, qui a la représentation :

$$f(\omega) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \gamma(n) e^{-in\omega} \quad \text{pour tout, } \omega \in [-\pi, \pi], \quad (2.5)$$

(voir [17] pour plus de détails sur les différentes représentations). La densité spectrale est analogue à la fonction de densité de probabilité, le fait que $\gamma(h)$ est non négative assure que :

$$f(\omega) \geq 0,$$

pour toute fréquence ω [89]. Ce qui implique d'après (2.5) que :

$$f(\omega) = f(-\omega) \quad \text{and} \quad f(\omega) = f(1 - \omega),$$

qui vérifie que la densité spectrale est une fonction paire avec période 1. Donc si x_t est un processus stationnaire avec une fonction d'autocovariance $\gamma(h)$ absolument sommable, alors $\gamma(h)$ et $f(\omega)$ contiennent la même information. Cette information, cependant, est exprimée de différentes manières.

2.1.3 L'estimation spectrale

Le problème d'estimation spectrale est un problème majeur dans le domaine d'analyse spectrale. En se basant sur la représentation (2.5), plusieurs points de départ peuvent être considérés. Dans la littérature, il existe deux méthodes de base pour estimer le spectre : la méthode paramétrique et la méthode non paramétrique.

L'estimation Paramétrique

L'estimation paramétrique de la densité spectrale suppose que le signal a une certaine structure qui peut être décrite selon un modèle paramétrique. Ensuite, l'estimation est donnée directement en utilisant les paramètres de ce modèle. Dans ce qui suit, quelques approches paramétriques pour l'estimation spectrale sont discutées.

Si la série temporelle peut être approximé par un processus autorégressif d'ordre p , AR(p), sous la forme :

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t,$$

ou $E(\epsilon_t^2) = \sigma^2$. En utilisant ce modèle, Parzen[71] a proposé l'estimation spectrale autorégressive.

La première étape de cette méthode est d'estimer les coefficients d'un processus AR(p) par les moindres carrés. Ensuite, étant donné qu'on connaît la forme de la fonction d'autocovariance d'un AR(p), l'estimation spectrale autoregressive peut être donnée comme :

$$f_Y(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{|\phi(\exp^{-i\omega})|^2},$$

tel que, $\phi(Z) = 1 - \phi_1 z - \dots - \phi_p z^p$. La difficulté dans ce cas est de choisir l'ordre p . Si le p est trop petit, le spectre estimé peut être biaisé. D'un autre côté, si p est grand, il peut augmenter la variance de l'estimateur. Harvey[38] a proposé de minimiser le critère AIC afin de trouver l'ordre approprié du modèle.

On peut suivre les mêmes étapes pour les modèles MA(q) et ARMA(p,q). Si Y_t est un processus ARMA(p,q) sous la forme :

$$Y_t = \epsilon_t - \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}, \quad \epsilon \sim \mathcal{N}(0, \sigma^2),$$

tels que, $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs, et $\phi(z)$ n'a pas de zéro sur le cercle unitaire. Alors, l'estimation de la densité spectrale peut être donnée comme suit :

$$f_Y(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(\exp^{-i\omega})|^2}{|\phi(\exp^{-i\omega})|^2}, \quad -\pi \leq \omega \leq \pi.$$

L'estimation non paramétrique

L'estimation non-paramétrique est la méthode qui se focalise sur l'estimation de la fonction d'autocovariance sans mettre des hypothèses sur la structure du signal. Généralement, le périodogramme est considéré comme l'approche la plus naturelle pour l'estimation non paramétrique.

Dans cette thèse, on va se concentrer sur l'estimation non paramétrique de la densité spectrale.

2.1.4 Le périodogramme

Dans le domaine du traitement de signal, le périodogramme est considéré comme un estimateur non paramétrique du spectre. Il est défini par le carré de la transformée de Fourier du signal comme suit

$$I_n(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^n Y_t \exp(-i\omega_j t) \right|^2, \quad (2.6)$$

tels que $\omega_j = 2\pi j/n$ sont des fréquences de Fourier. En pratique, le périodogramme est un outil très utile pour décrire une série temporelle. Si une série a un signal sinusoïdal fort pour une certaine fréquence, alors il y aura un pic dans le périodogramme à

cette fréquence. Un pic dans $I_n(\omega)$ à une fréquence $\omega = \omega'$ indique un phénomène de périodicité possible avec une période $p = 2\pi/\omega'$.

Exemple 1 *Figure 2.1* montre une réalisation du périodogramme avec un signal mixte de trois fréquences. Premièrement, pour $t = 1, \dots, 100$, on a généré les trois séries :

$$x1 = 2\cos(2\pi t6/100) + 3\sin(2\pi t6/100),$$

$$x2 = 4\cos(2\pi t10/100) + 5\sin(2\pi t10/100),$$

$$x3 = 6\cos(2\pi t40/100) + 7\sin(2\pi t40/100).$$

Après on a construit la série :

$$x = x1 + x2 + x3.$$

Ces quatre séries sont présentées dans la *Figure 2.1*.

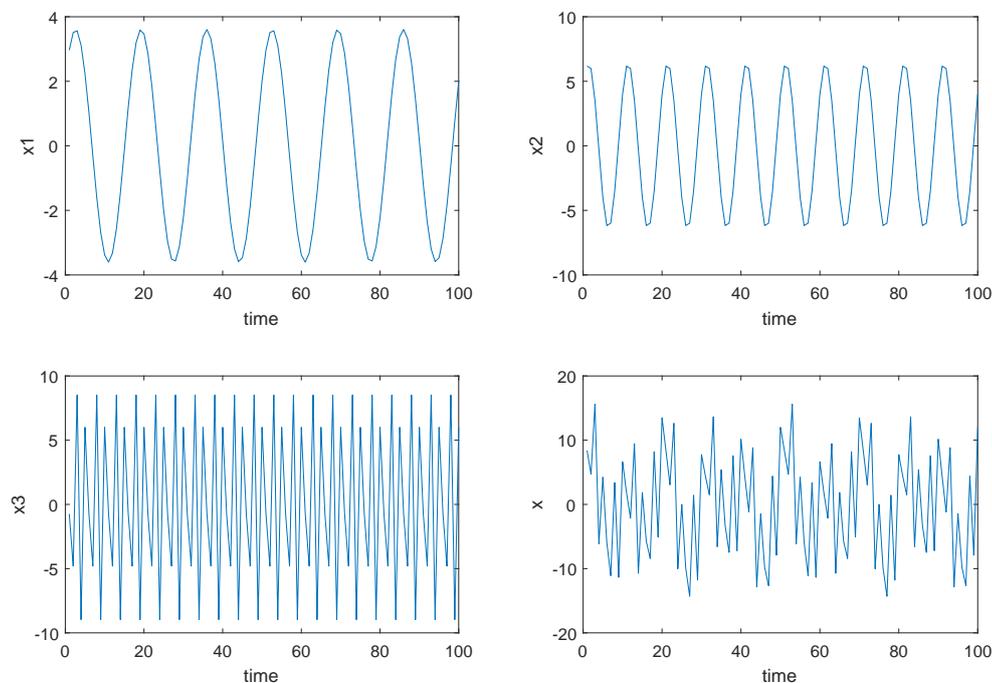


FIGURE 2.1: Signaux périodiques de l'exemple 1.

Dans la *Figure 2.2*, le périodogramme identifie clairement les trois fréquences.

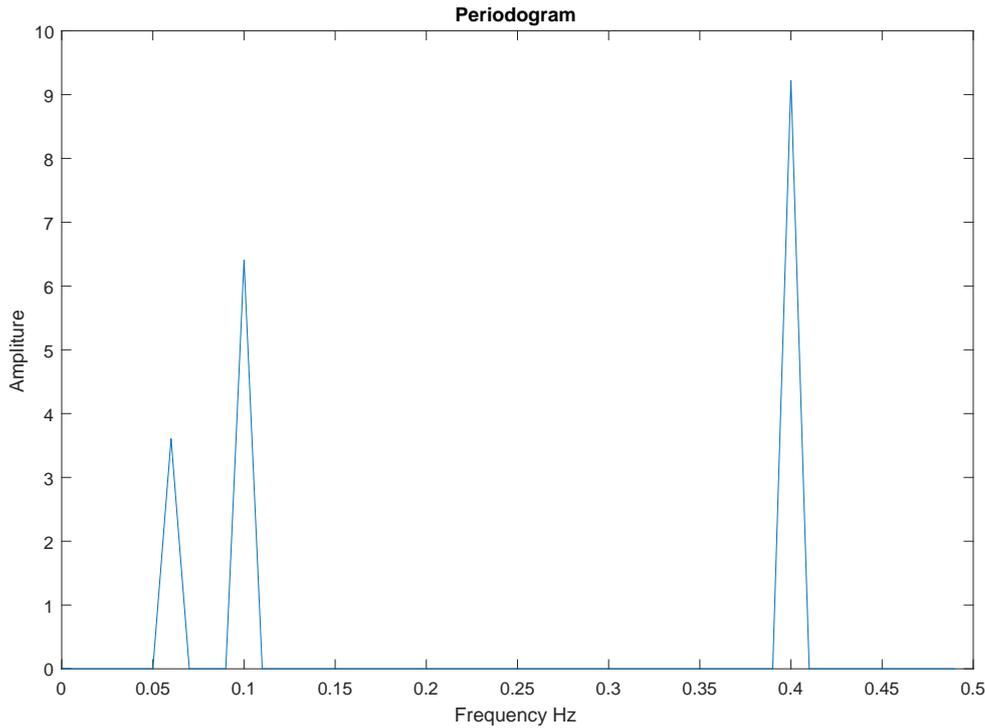


FIGURE 2.2: Périodogramme du signal à 3 période de l'exemple 1.

Une autre représentation équivalente du périodogramme est donnée par la propriété de la projection L^2 , en utilisant une régression harmonique (2.3), comme suit

$$I(\omega_j) = \frac{1}{2\pi} n(|A_j|^2 + |B_j|^2) = \frac{1}{2\pi} n \|\beta_j\|_2^2, \quad (2.7)$$

avec :

$$A_j = (1/n)^{-1/2} \sum_{t=1}^n Y_t \cos(\omega_j t),$$

et

$$B_j = (1/n)^{-1/2} \sum_{t=1}^n Y_t \sin(\omega_j t).$$

Le paramètre β peut donc être obtenu comme une solution du problème d'optimisation suivant :

$$\hat{\beta}(\omega_j) = \arg \min_{\beta \in \mathbb{R}^2} \|Y_t - x^T(\omega_j)\beta\|_2^2, \quad (2.8)$$

avec $x(\omega_j) = [\cos(\omega_j), \sin(\omega_j)]^T$. On peut aussi considérer le cas où le périodogramme est représenté avec la transformée de Fourier de l'autocovariance empirique. Soit la

fonction d'autocovariance empirique du signal $\{Y_t\}$:

$$\hat{\gamma}(k) = \frac{1}{n-k} \sum_{t=0}^{n-1-k} (Y_{t+k} - \bar{Y}_n)(Y_t - \bar{Y}_n),$$

si $k \geq 0$, $\bar{Y}_n = n^{-1} \sum_{t=0}^{n-1} Y_t$, alors, le périodogramme peut être représenté par :

$$I_n(\omega_j) = \frac{1}{2\pi} \sum_{|k| < n} \frac{n-k}{n} \hat{\gamma}_k e^{-ik\omega_j}. \quad (2.9)$$

Ces différentes représentations donnent différents points de départ pour l'estimation. On peut soit considérer d'autres fonctions de perte dans l'équation d'optimisation (2.8), soit remplacer la fonction d'autocovariance par d'autres mesures de dépendance comme la distribution empirique ou les copules.

Dans notre travail, on a considéré le premier cas en remplaçant la fonction perte par une approche de régression quantile. Ce cas a été introduit en premier temps par Li[51].

Propriétés asymptotiques du périodogramme

Considérons la série temporelle stationnaire $\{Y_t\}$, avec une moyenne μ et une fonction d'autocovariance absolument sommable $\gamma(\cdot)$. Si ω_j est une fréquence de Fourier non nulle, alors on peut écrire :

$$I_n(\omega_j) = \sum_{|k| < 1}^n \hat{\gamma}(k) e^{-ik\omega_j}, \quad (2.10)$$

telle que,

$$\hat{\gamma}(k) = n^{-1} \sum_{t=1}^{n-k} (y_{t+k} - \mu)(y_t - \mu),$$

est la fonction d'autocovariance asymptotique, $k \geq 0$ et $\mu = n^{-1} \sum_{t=1}^n y_t$. Si $\mu = 0$, alors l'équation (2.10) est valide pour toutes les fréquences de Fourier, $\omega_j \in [-\pi, \pi]$ [17].

A partir de l'équation (2.10), un estimateur naturel du spectre $f(\omega_j)$ pour $\omega_j \neq 0$ est $(2\pi)^{-2} I(\omega_j)$.

Pour étendre le périodogramme à couvrir tout l'intervalle $[-\pi, \pi]$, on doit passer par la définition de Fuller[30], qui l'a défini avec une fonction par morceaux coïncide avec (2.10) dans les fréquences de Fourier.

Pour chaque $\omega_j \in [-\pi, \pi]$, le périodogramme peut être défini par :

$$I(\omega) = \begin{cases} I(\omega_k) & \text{si } \omega_k - \pi/n < \omega \leq \omega_k + \pi/n \text{ et } 0 \leq \omega \leq \pi \\ I(-\omega_k) & \text{si } \omega \in [-\pi, 0) \end{cases}$$

La proposition suivante montre que $(2\pi)^{-1}I_n(\omega)$ est un estimateur non biaisé du spectre $f(\omega)$.

Proposition 2 ([17]) *Si $\{Y_t\}$ est une série temporelle stationnaire, avec une moyenne μ , et une fonction d'autocovariance absolument sommable, alors :*

$$(i) \quad E(I_n(0)) - n\mu^2 \rightarrow 2\pi f(0),$$

et

$$(ii) \quad E(I_n(\omega)) \rightarrow 2\pi f(\omega) \quad \text{if } \omega \neq 0$$

Si $\mu = 0$, alors $E(I_n(\omega))$ converge uniformément vers $2\pi f(\omega)$ sur $[-\pi, \pi]$.

En utilisant la représentation donnée dans (2.7) on peut directement calculer la distribution asymptotique du périodogramme.

Comme $\{\cos(\omega_j), \sin(\omega_j)\}$ forme une base orthonormé de $L^2[-\pi, \pi]$, si le processus $\{Y_t\}$ est un processus iid avec une loi $\mathcal{N}(0, \sigma^2)$, alors les variables A_j et B_j sont indépendantes avec une distribution $\mathcal{N}(0, \sigma^2)$.

Par conséquent $I_n(\omega_j)$ définis par (2.7) sont indépendants et distribués comme :

$$\frac{1}{2\pi f(\omega_j)} I_n(\omega_j) \sim \chi_2^2.$$

Dans le cas ou le processus est construit à partir d'une combinaison linéaire, ou dans

le cas du processus non-Gaussien ou non-stationnaire, le périodogramme ordinaire ne peut pas donner une bonne performance. Dans la littérature, beaucoup d'extensions et de modifications ont été proposées pour surmonter ce problème, telles que les techniques de lissage (smoothing).

2.2 La régression quantile

2.2.1 La régression linéaire et la régression quantile

En statistiques, l'une des tâches les plus fondamentales est de chercher l'existence d'une relation (linéaire ou non linéaire) entre une variable expliquée (de sortie) et une série de variables explicatives (d'entrée). Cette tâche qui apparaît dans plusieurs domaines d'application (Économie, Physique, Machine learning, etc...) a une grande importance pour comprendre certains phénomènes, extraire des informations utiles et ainsi, pouvoir faire des prédictions et prévisions. Par ailleurs, la régression est le cas idéal pour la résolution de ce problème. [89]

La première forme de la régression a été publiée dans [48], et dans [32] dans le cadre de la détermination des orbites des corps sur le soleil, à partir des observations astronomiques. Le terme régression a été inventé par Galton[31]. Dans un papier connu, Galton a trouvé que la taille moyenne des nouveaux nés tend vers (régresse) la taille moyenne de la population quelque soit la taille de leurs parents. Mais son travail a été étendu plus tard dans [105] à un contexte statistique plus général.

L'interprétation moderne de la régression est cependant très différente. En gros, on peut dire qu'un modèle de régression linéaire revient à supposer que la moyenne conditionnelle de la variable expliquée Y sachant les variables explicatives X , est une fonction affine de celle des variables explicatives [34], c'est-à-dire :

$$E(Y/X) = X^T \beta,$$

avec β est le paramètre de régression à estimer. Ainsi, le modèle de régression linéaire peut s'écrire comme suit :

$$Y = X^T \beta + \epsilon, \quad (2.11)$$

ou ϵ représente le bruit ou l'erreur de la régression. Pour l'estimation du paramètre, la méthode des moindres carrés est admirablement adaptée à ce modèle avec certaines conditions. Mais, il existe encore d'autres méthodes d'estimation comme le maximum de vraisemblance, ou encore l'inférence bayésienne.

La méthode des moindres carrés se concentre sur la minimisation de l'erreur quadratique entre la variable de sortie et la moyenne conditionnelle des variables explicatives :

$$\hat{\beta} = \arg \min_{\beta} \|Y - X^T \beta\|_2^2 = \arg \min_{\beta} \sum_{i=1}^n (\epsilon_i)^2.$$

Un exemple de l'estimation des moindres carrés ordinaires (MCO) est donné dans l'exemple suivant. Considérons l'ensemble de données des accidents mortels de la route aux États-Unis. Ce Dataset contient des données sur le nombre des accidents mortels de la route par rapport à la population.

Un modèle linéaire simple peut être présenté pour modéliser ce phénomène :

$$y = \beta_0 + \beta_1 x + \epsilon.$$

Figure (2.3) montre une estimation des moindres carrés ordinaires de la droite de régression par rapport au nuage de points présenté.

La méthode de moindres carrés souvent utilisée dans les études statistiques, apporte une information essentielle, néanmoins limitée. Dans le cas où les données contiennent des valeurs aberrantes, ou même encore bruitées, l'estimation devient peu fiable.

Dans plusieurs études récentes la moyenne a montré ses limites au sens de la représentation d'information et de la robustesse. Dans des études des revenus dans les dernières années, Landais[46] et Solard[88] ont observé une très faible augmentation de moyenne, alors qu'il y a une augmentation remarquable dans le dernier décile.

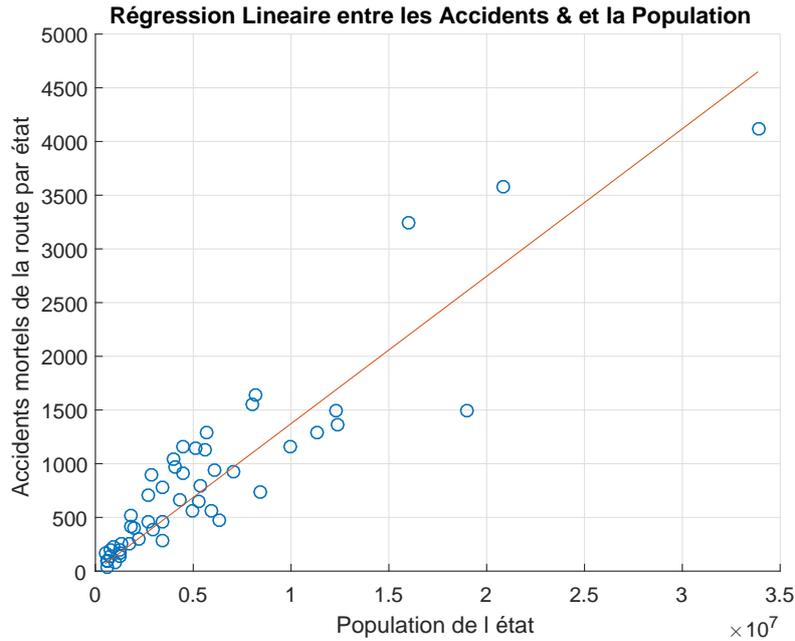


FIGURE 2.3: Régression linéaire de l'ensemble de données des accidents aux États-Unis (Matlab)

Pour résoudre ce genre de problème, Koenker et Basset [45] ont développé le modèle de régression quantile. Dans leur approche, ils ont supposé que le quantile conditionnel $\xi_\tau(Y/X)$ est une transformation affine de X .

Par définition, si Y a une fonction de distribution de probabilités connue $F_Y(y) = P(Y \leq y)$, le quantile d'ordre τ n'est autre que l'inverse de cette fonction,

$$\xi_\tau(Y) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\},$$

avec, τ est bien dans l'intervalle $(0,1)$. Les quantiles les plus utilisés sont la médiane ($\tau = 0.5$) et les premier et dernier quartiles ($\tau = 0.25$, $\tau = 0.75$ et $\tau = 0.95$). D'autre part, si Y a une fonction de distribution conditionnelle de probabilité $F_{Y/X}$, alors le quantile conditionnel d'ordre τ est donné par :

$$\xi_\tau(Y/X) = F_{Y/X}^{-1}(\tau) = \inf\{y : F_{Y/X}(y) \geq \tau\}.$$

D'après [45], le τ -ème quantile conditionnel de Y sachant X , est donné par :

$$\xi_\tau(Y/X) = X^T \beta_\tau.$$

Le modèle de régression quantile peut donc s'écrire de manière suivante :

$$Y = X^T \beta_\tau + \epsilon,$$

tel que $\xi_\tau(\epsilon_\tau/X) = 0$.

Ainsi, en utilisant plusieurs valeurs de τ , la méthode de régression quantile nous permet de voir l'effet des variables explicatives sur toute la distribution de la variable d'intérêt. Les paramètres de la régression quantile β_τ se modifient pour chaque valeur de τ en nous donnant une régression différente.

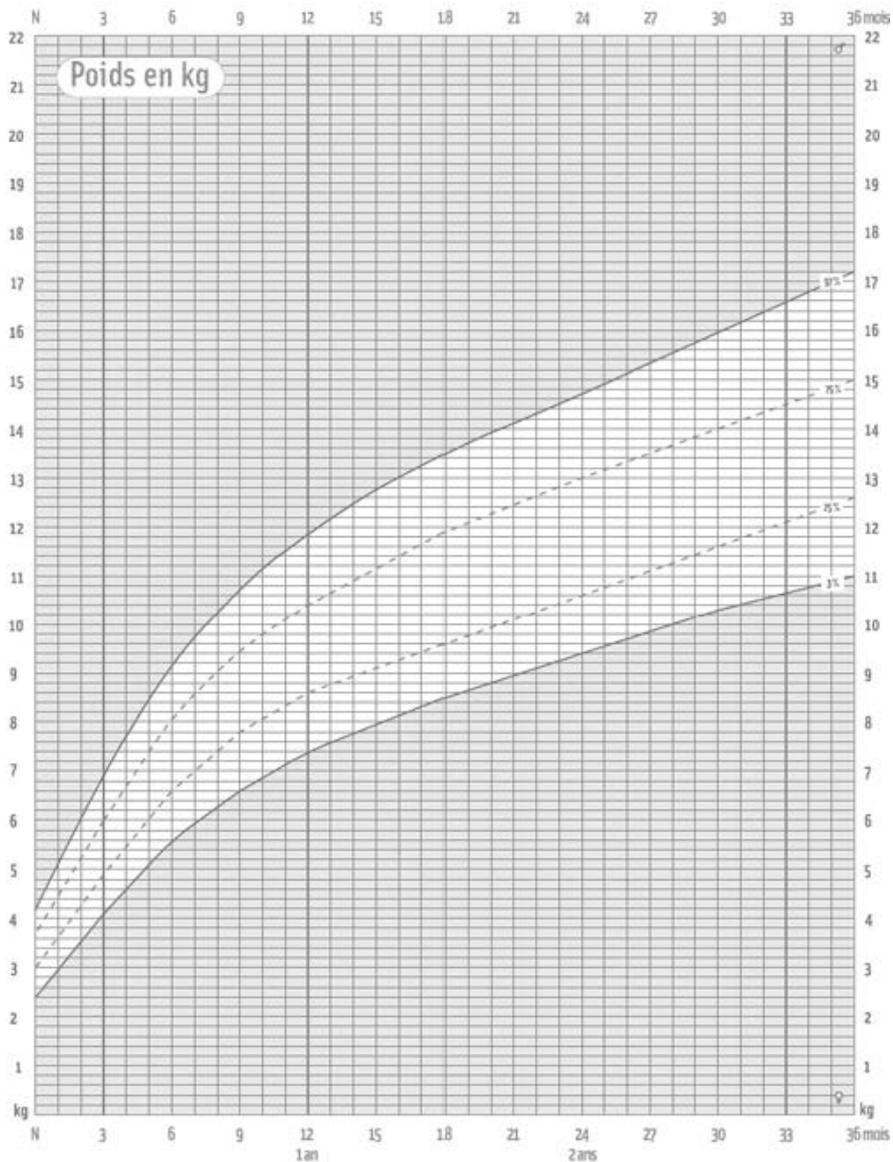


FIGURE 2.4: Croissance de poids selon l'âge ([20]).

Considérons l'exemple connu dans le domaine de santé, qui montre comment la distribution du poids des nouveaux nés varie en fonction de l'âge [19]. Figure(2.4) montre une illustration des quartiles (3ème, 25ème, 75ème et 95ème) pour ce dataset. A partir de 15 mois, on peut remarquer que les différents quantiles des poids augmentent d'une manière linéaire. Mais, les droites de régression ne sont pas parallèles. Ce qui signifie que l'information ne peut pas être présentée avec une simple régression linéaire.

Pour démontrer la robustesse de la régression quantile par rapport aux valeurs aberrantes, considérons l'exemple de simulation suivant. Supposons qu'on a un dataset généré à partir d'un modèle linéaire simple avec une hétéroscédasticité (la variance varie avec x).

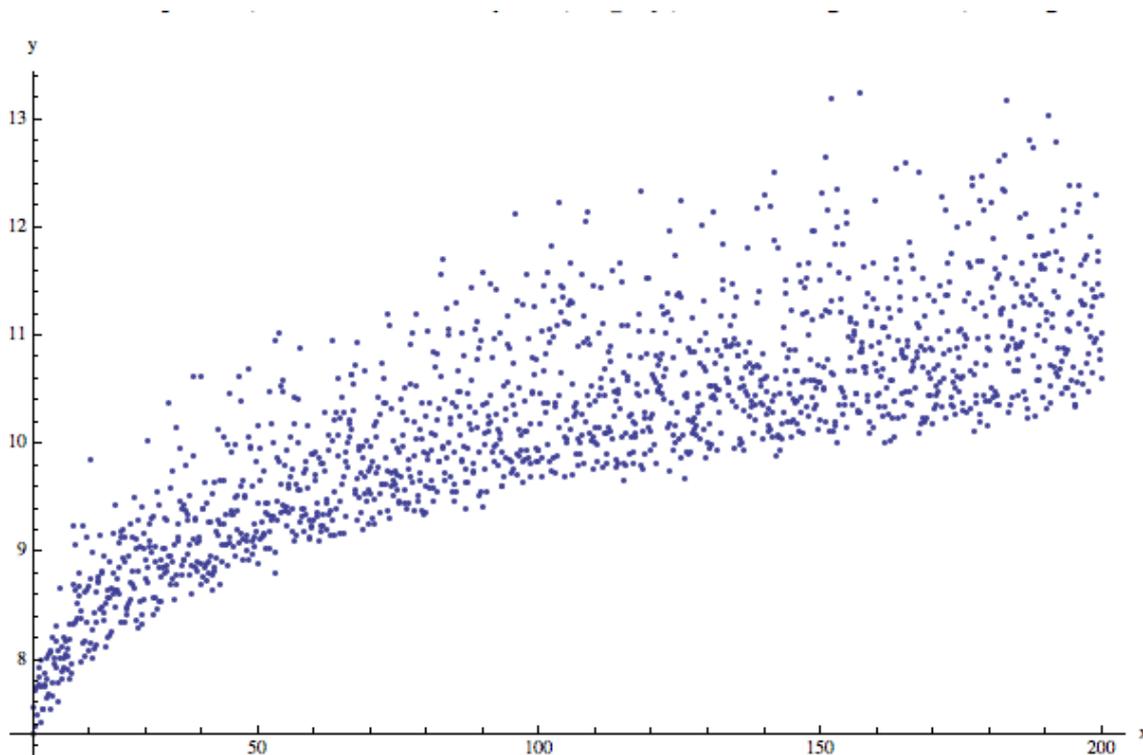


FIGURE 2.5: Dataset généré a partir d'un modèle linéaire hétéroscédastique ([7]).

Figure 2.5 montre une illustration du dataset. En regardant le graphe, on peut supposer que le modèle est de la forme :

$$y = \beta_0 + \beta_1 * x + \beta_2 * \log(x)$$

Une estimation des régressions MCO et quantile est donnée dans la Figure 2.6.

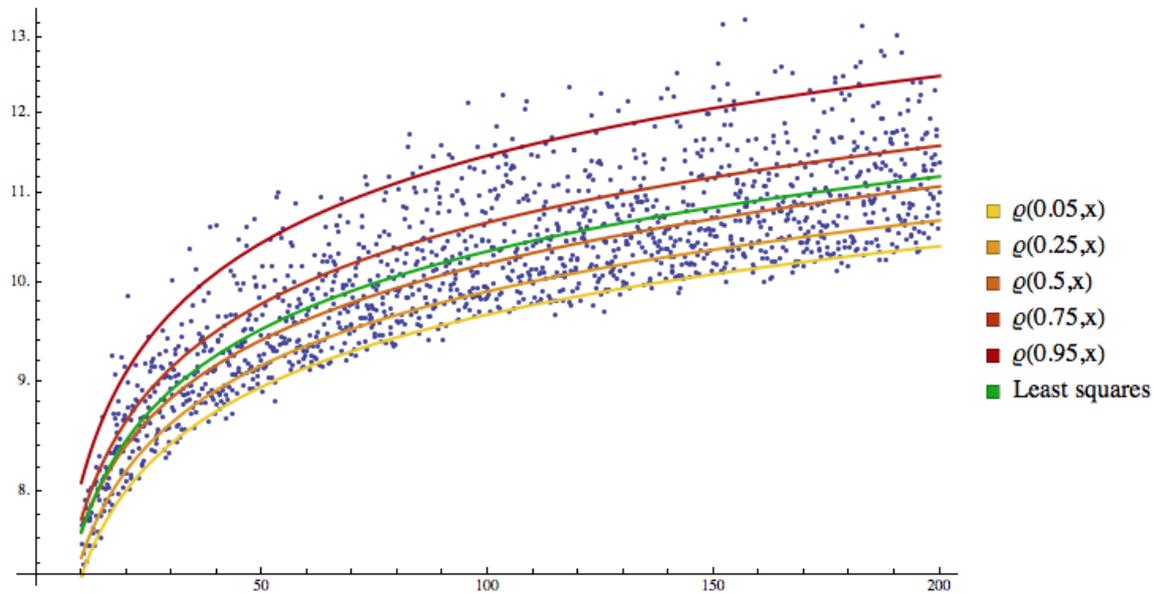


FIGURE 2.6: Régression MCO et quantiles pour un modèle linéaire hétéroscédastique [7].

Maintenant, on multiplie 50% des valeurs de y qui sont supérieures à 11.25 par un facteur disant $\alpha = 1.8$, pour avoir des valeurs aberrantes. Figure 2.7 montre une illustration du nouveau dataset.

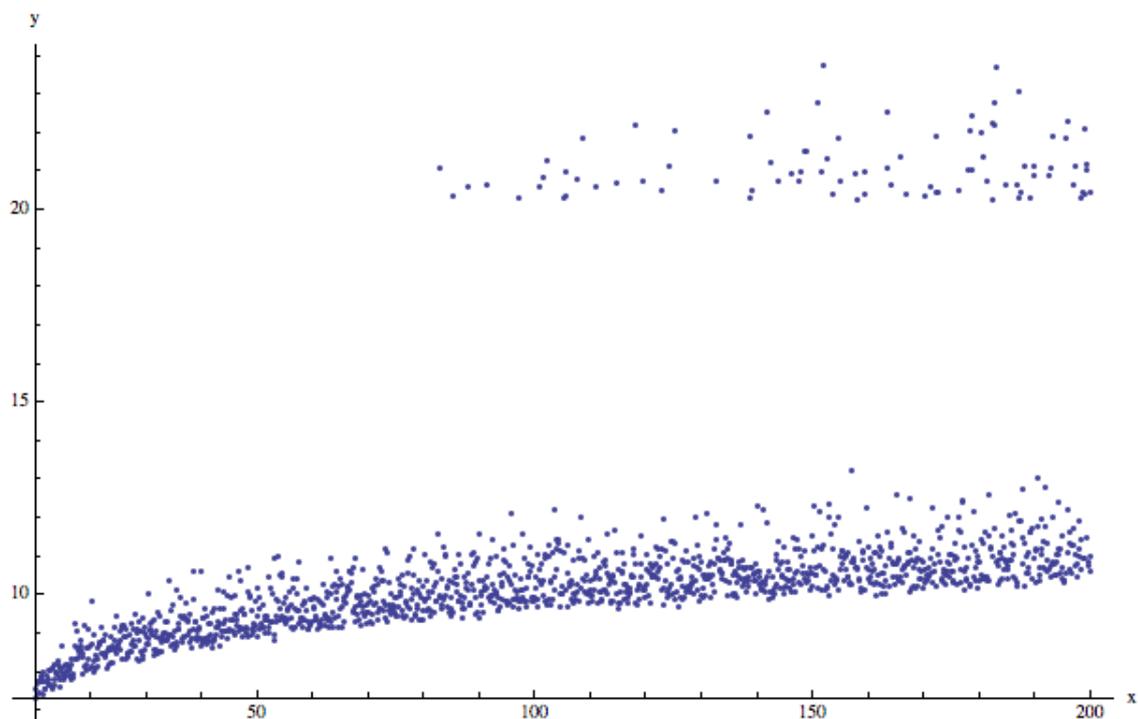


FIGURE 2.7: Données avec des Valeurs aberrantes [7].

En estimant les régressions MCO et quantiles du nouveau dataset Figure (2.8), on

peut remarquer clairement que les MCO ont eu un changement significatif, alors que les régressions quantiles n'ont pas eu la même influence des valeurs aberrantes.

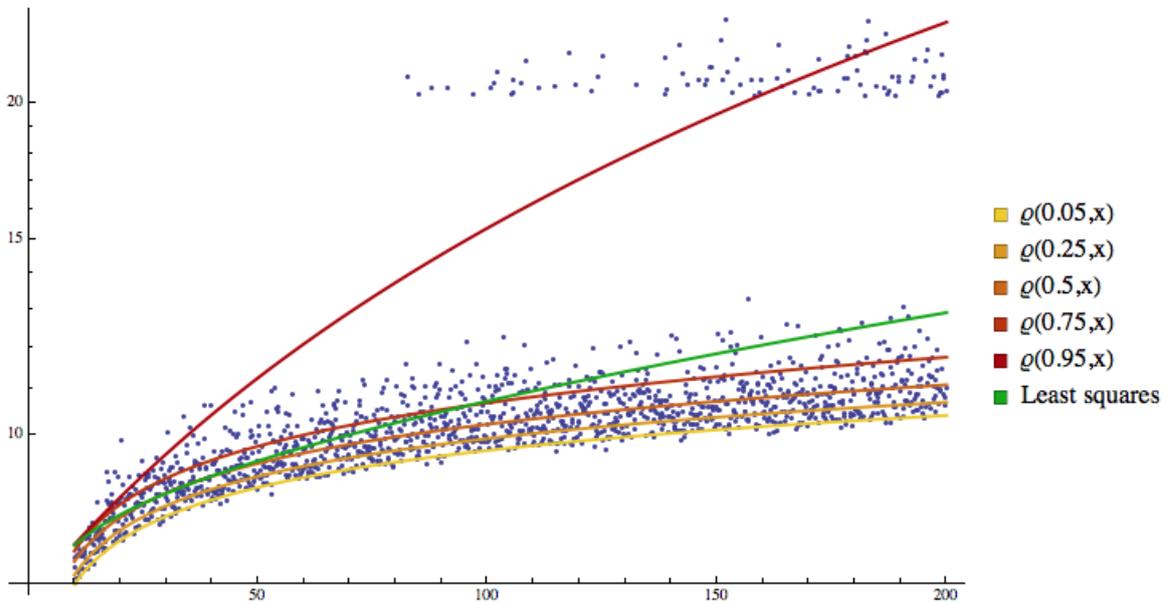


FIGURE 2.8: Régression MCO et quantiles pour un dataset avec des valeurs aberrantes [7].

2.2.2 Estimation des paramètres

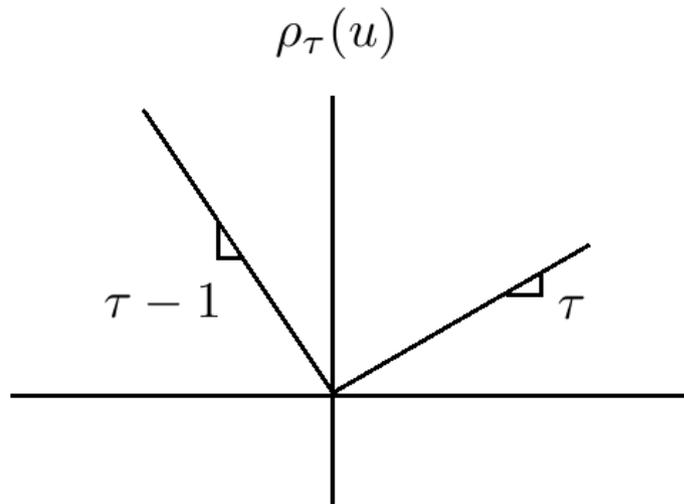
Pour pouvoir estimer les coefficients de la régression quantile, il est nécessaire de commencer par l'estimation des quantiles d'une variable d'intérêt Y .

Fox and Rubin[28] ont montré que la définition des quantiles peut être obtenue par un simple problème d'optimisation. Si la fonction de perte $\rho_\tau(\cdot)$ est définie par la fonction linéaire par morceaux :

$$\rho_\tau(u) = u(\tau - I(u < 0)),$$

pour $\tau \in (0, 1)$, et I est la fonction indicatrice. En minimisant le problème :

$$E(\rho_\tau(X - \hat{x})) = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x}) dF(x) + \tau \int_{\hat{x}}^{+\infty} (x - \hat{x}) dF(x).$$

FIGURE 2.9: La fonction de perte ρ_τ

Différenciant par rapport à \hat{x} , on a

$$0 = (1 - \tau) \int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{\infty} dF(x) = F(\hat{x}) - \tau.$$

Puisque F est monotone, tout élément de $\{x : F(x) = \tau\}$ minimise l'espérance de la fonction perte. Si la solution est unique, $\hat{x} = F^{-1}(\tau)$; sinon, on a un intervalle de quantiles pour lesquels la plus petite doit être choisie (voir [44] p5-p7 pour plus de détails). On peut alors définir les quantiles d'une variable d'intérêt Y par :

$$\widehat{\xi}_\tau = \arg \min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(y_i - \xi).$$

Ainsi, l'estimateur des paramètres de la régression quantile est simplement définie par :

$$\widehat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta). \quad (2.12)$$

Figure (2.9) montre une illustration de la fonction perte ρ_τ .

2.3 Régression régularisée

Dans les domaines des statistiques, apprentissage automatique, et autres domaines, la régularisation est le fait de rajouter une fonction appelée fonction de régularisation (pénalité) à la fonction objective du problème d'estimation, dont l'objectif est de résoudre un problème mal-posé, ou d'assurer la convergence en ajoutant des contraintes.

On dit qu'un problème est bien posé au sens de Hadamard [36], si la solution existe, elle est unique et le comportement de la solution se change continument avec les conditions initiales. On dit qu'un problème est mal posé, s'il ne vérifie pas ces conditions [22].

Le rôle principal des méthodes de régularisation, est de trouver le meilleur sous ensemble parmi les variables explicatives, qui donne la meilleure prédiction. La régression régularisée a été utilisée beaucoup dans la littérature comme une méthode de sélection de variables et réduction de dimension [92].

Il existe plusieurs types de fonctions de régularisation dans la littérature. En utilisant la norme L_1 , Tibshirani[92] a défini ce qu'on appelle le LASSO (Least Absolute Shrinkage and Selection Operator), cette fonction force certains prédicteurs pour être exactement des zéros.

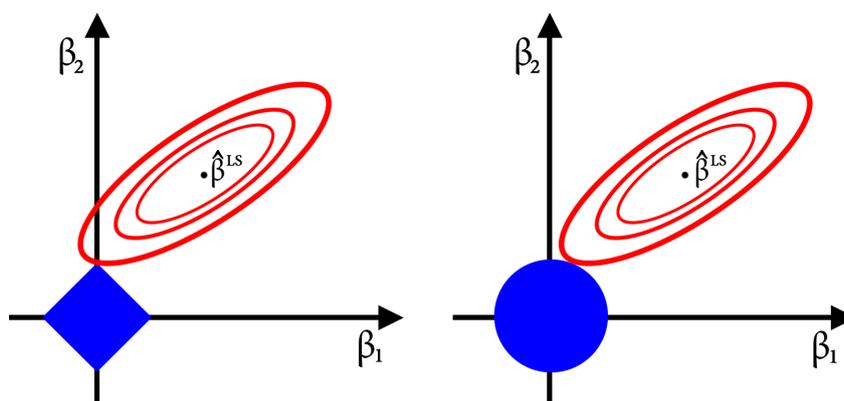


FIGURE 2.10: les formes géométriques des fonctions de régularisation, Lasso à gauche et Ridge à droite.

Cela est due à la forme de la région des contraintes sous la norme L_1 , qui est géométriquement représentée par un carré tourné (une polyédrique dans un espace plus de

deux dimensions) Figure (2.10). Donc, lorsque la solution est unique, la fonction objective touche la région des contraintes aux coins, ce qui explique la nullité de certaines variables.

la fonction Ridge (norme L_2), offre aussi la propriété de sélection de variables. Mais, elle réduit juste l'effet de certaines variables, et ne les force pas vers zéro. Ce résultat peut être conclu via la forme géométrique de la fonction, qui forme un cercle (boule dans un espace plus de deux dimensions) Figure (2.10).

On définit l'estimateur de régression linéaire régularisé comme suite :

$$\hat{\beta}^{reg} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i^T \beta) + \alpha J(\beta), \quad (2.13)$$

avec, $\alpha \geq 0$ est le coefficient de régularisation, et $J(\beta)$ représente la fonction de régularisation de β . L'estimateur en utilisant la fonction de LASSO est alors :

$$\hat{\beta}^{LASSO} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i^T \beta) + \alpha \sum_{j=1}^p |\beta_j|. \quad (2.14)$$

2.3.1 Inférence

De point de vue inférence, les propriétés oracles [109] d'un estimateur sont très importantes. elles assurent deux propriétés asymptotiques importantes. La première est la consistance, c'est à dire que lorsque la taille d'échantillon approche l'infini, le sous ensemble sélectionné converge vers le vrai ensemble des variables explicatives avec probabilité 1. La deuxième est que les estimateurs sont asymptotiquement distribués avec une loi normale.

Dans [25], les auteurs ont démontré que la pénalisation en utilisant la fonction LASSO ne vérifie pas les propriétés oracles. Cela a été confirmé après par Zou[109].

Soit $\hat{\beta}^{LASSO}$ défini par l'équation (2.14) un estimateur du modèle linéaire (2.11), avec les conditions suivantes :

- (a) $\{\epsilon_i\}$ sont des variables aléatoires iid avec moyenne nulle et une variance σ^2 .
- (b) $\frac{1}{n}x^tx \rightarrow C$, ou C est une matrice définie positive.

Supposons que $A = \{1, 2, \dots, p_0\}$, et soit

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

telle que C_{11} est une matrice de taille $p_0 \times p_0$. Soit l'ensemble $A_n = \{j : \widehat{\beta}_j \neq 0\}$. L'estimateur (2.14) est consistant si et seulement si $P(A_n = A) = 1$.

Lemme 1 ([109]) *Si $\alpha_n/n \rightarrow \alpha_0 \geq 0$, alors $\sqrt{n}(\widehat{\beta} - \beta^*) \rightarrow_d \arg \min V_2$, avec*

$$V_2(u) = -2u^T W + u^T C u + \alpha_0 \sum_{j=1}^p [u_j \operatorname{sgn}(\beta_j^*) I(\beta_j^* \neq 0) + |u_j| I(\beta_j^* = 0)].$$

et W a une distribution $\mathbf{N}(0, \sigma^2 C)$.

Lemme (1) montre que l'estimateur du Lasso n'est pas consistant, car pour $\alpha_n = O(\sqrt{n})$, A_n ne peut pas être A avec une probabilité positive.

Zou[109] a proposé de faire une modification sur la fonction de régularisation en rajoutant des fonctions de poids.

$$\widehat{\beta}^{ALASSO} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i^T \beta) + \alpha \sum_{j=1}^p \nu_j |\beta_j|, \quad (2.15)$$

avec, $\nu_j = 1/|\widehat{\beta}^{reg}|^d$ tel que $d > 0$.

Cet estimateur est appelé le LASSO adaptatif (adaptive LASSO). Il est important de noter que (2.15) est un problème d'optimisation convexe. Ainsi, il ne souffre pas de plusieurs minimums locaux.

Théorème 1 ([109]) *Supposons que $\alpha_n/\sqrt{n} \rightarrow 0$ et $\alpha_n n^{(d-1)/2} \rightarrow \infty$. Alors, le LASSO adaptatif doit satisfaire le suivant :*

(1) *La consistance dans la sélection de variables* : $\lim_n P(A_n^* = A) = 1$

(2) *La normalité asymptotique* : $\sqrt{n}(\hat{\beta}_A - \beta^*) \rightarrow_d N(0, \sigma^2 \times C_{11}^{-1})$

Le théorème (1) montre que le LASSO adaptatif a les propriétés d'oracle. La fonction $\hat{\nu}_j$ est la clé dans le théorème. Quand la taille de l'échantillon augmente, les pondérations pour les prédicteurs à coefficients zéros tendent vers l'infini, tandis que les pondérations pour les prédicteurs à coefficients non nuls convergent vers une constante finie.

2.3.2 La régression quantile régularisée

Aujourd'hui l'appariement de la régression quantile et les méthodes de régularisation ont attiré beaucoup d'attention dans la littérature. Koenker[44] a appliqué le LASSO pénalité sur le modèle de régression quantile à effet mixe pour les données longitudinales, afin d'estimer l'effet aléatoire. Li and Zhu[53] ont développé le chemin de solution pour l'estimateur de L_1 régression quantile. Les auteurs dans [103] ont étudié les propriétés oracles pour le SCAD et Adaptive lasso pénalités pour le contexte de régression quantile pénalisé.

Un estimateur de régression quantile régularisée peut s'écrire comme suit :

$$\hat{\beta}^{AL} = \arg \min_{\beta \in \mathbb{R}^p} \rho_\tau \sum_{i=1}^n (y_i - x_i^T \beta) + \alpha J(\beta), \quad (2.16)$$

Comme pour la régression standard régularisée, plusieurs types de fonction de régularisation ont été proposé pour la régression quantile régularisée. Les propriétés d'oracle ont été démontré pour l'estimateur de régression quantile avec une pénalité de LASSO adaptatif [103]. Ils ont proposé de prendre l'estimateur de régression quantile non régularisée comme une fonction de poids.

$$\hat{\beta}^{ALQ} = \arg \min_{\beta \in \mathbb{R}^p} \rho_\tau \sum_{i=1}^n (y_i - x_i^T \beta) + \alpha \sum_{j=1}^p \nu_j |\beta_j|, \quad (2.17)$$

avec, $\nu_j = 1/|\hat{\beta}_\tau|^d$ tel que $d > 0$. $\hat{\beta}_\tau$ défini par l'équation (2.12).

Supposons que les conditions suivantes sont satisfaites :

(i) Les termes $\{\epsilon_i\}$ sont des variables aléatoires iid, avec une densité de probabilité $f(\cdot)$ continue et positive au voisinage de zéro.

(ii) $\lim_{n \rightarrow \infty} \sum_{i=1}^n x_i x_i^T = C$, avec C est une matrice définie positive.

Alors, le théorème suivant peut être démontré.

Théorème 2 ([109]) *Soit $\{(x_i, Y_i), i = 1, \dots, n\}$ du modèle linéaire (2.11), satisfait les conditions (i) et (ii). Si $\sqrt{n}\alpha_n \rightarrow 0$ et $n^{(d+1)/2}\alpha_n \rightarrow \infty$, alors on a*

1. *Sparsité : $\widehat{\beta}_{A_n} = 0$.*

2. *La normalité asymptotique : $\sqrt{n}(\widehat{\beta}_{A_n}^{ALQ} - \beta_{A_n}) \rightarrow^L N(0, \tau(1 - \tau)C_{11}^{-1}/f(0)^2)$.*

Théorème 2 montre que l'estimateur du régression quantile avec une fonction de régularisation LASSO adaptatif, a les propriétés d'oracle (voir [109] pour plus de détails).

Chapitre 3

Quantile périodogramme pénalisé pour l'estimation spectrale

“Pure mathematicians just love to try unsolved problems - they love a challenge.”

– Andrew Wiles

Ce chapitre discute les propriétés asymptotiques du quantile périodogramme ainsi que le nouvel estimateur le quantile périodogramme pénalisé.

3.1 Le quantile périodogramme

Les techniques des moindres carrés permettent d'estimer la moyenne conditionnelle de la variable réponse en fonction des prédicteurs. Les techniques de régression quantile, cependant, se concentrent sur l'estimation des quantiles de la variable réponse.

Pour toute série temporelle bornée à valeurs réelles $\{Y_1, Y_2, \dots, Y_n\}$, la représentation de Fourier [17] peut être donnée comme suit :

$$Y_t = \lambda_0 + \sum_{j=1}^q \beta_{1j} \cos(\omega_j t) + \beta_{2j} \sin(\omega_j t), \quad (3.1)$$

tels que $\omega_j = 2\pi j/n$, $j = 1, \dots, q$, sont les fréquences de Fourier, et $(\lambda_0, \beta_{1j}, \beta_{2j}) \in \mathbb{R}^3$ sont les paramètres à estimer.

En utilisant la représentation spectrale de la régression harmonique (3.1), [51] a proposé dans un premier temps de remplacer la fonction de minimisation de moindres carrés ordinaires, par une minimisation de valeurs absolues :

$$\tilde{\beta}(\omega_j) = \arg \min_{\beta \in \mathbb{R}^2} |Y_t - x^T(\omega_j)\beta|, \quad (3.2)$$

tels que, $x_t(\omega_j) = [\cos(\omega_j), \sin(\omega_j)]^T$. Le périodogramme de Laplace (Laplace periodogram) comme appelé par [51] peut être donc défini à partir de l'équation (3.2), par :

$$L(\omega_j) = \frac{1}{4}n \|\tilde{\beta}_j\|_2^2, \quad (3.3)$$

Le périodogramme de Laplace bénéficie des avantages de la régression LDA (least absolute deviation), comprenant la robustesse contre les valeurs aberrantes et la distorsion non-linéaire. Cela peut être aussi interpréter comme un cas particulier de la régression quantile, dont le quantile d'intérêt est la médiane.

Plus tard en 2012, [52] a généralisé sa définition pour un quantile quelconque. Utilisons la régression quantile pour estimer les coefficients dans l'équation (3.1) :

$$\hat{\lambda}_0 = \arg \min_{\lambda} \sum_{t=1}^n \rho_{\tau}(Y_t - \lambda)$$

et

$$\hat{\beta}_{\tau}(\omega_j) = \arg \min_{\beta \in \mathbb{R}^2} \sum_{t=1}^n \rho_{\tau}(Y_t - \lambda_0 - x_t^T(\omega_j)\beta) \quad (3.4)$$

avec, $\rho_{\tau}(u)$ est la fonction de perte définie par :

$$\rho_{\tau}(u) = u\{\tau - I(u < 0)\} = \begin{cases} -1(1 - \tau)u, & \text{if } u < 0. \\ \tau u, & \text{if } u \geq 0. \end{cases}$$

ou, $I(\cdot)$ est la fonction indicatrice, et τ est le quantile d'intérêt. Alors, on a :

$$\tilde{Y}_t = \hat{\lambda}_0 + \sum_{j=1}^q \hat{\beta}_{1j}^{\tau} \cos(\omega_j t) + \hat{\beta}_{2j}^{\tau} \sin(\omega_j t). \quad (3.5)$$

Le quantile périodogramme de Y_t est défini par :

$$Q_{\tau}(\omega_j) = \frac{1}{4} n \left\| \hat{\beta}_{\tau}(\omega_j) \right\|_2^2 \quad (3.6)$$

Le quantile périodogramme est une généralisation du périodogramme de Laplace pour un quantile quelconque. Il peut donner une vue complète sur l'information contenu dans la série en prenant en compte toute la distribution du processus, et non seulement sa moyenne. Il a la capacité de détecter des périodicités dans différentes parties de la distribution. De plus, il est robuste contre les valeurs aberrantes et les distorsions non linéaires [52].

3.1.1 Propriétés asymptotiques

Il est bien connu que le périodogramme lissé (smoothed periodogram) constitue un estimateur de la densité spectrale de puissance. Le Laplace périodogramme est trouvé avoir une relation similaire au spectre de passage à zéro [51]. De manière analogue, le quantile périodogramme défini par [52] est démontré avoir une relation au spectre de passage à niveau.

Soit $S(\omega)$ le spectre de passage à niveau défini par :

$$S(\omega) = \sum_{h=-\infty}^{\infty} \left\{ 1 - \frac{1}{2\tau(1-\tau)} \gamma_h \right\} \exp(i\omega h) \quad (3.7)$$

$$= \sum_{h=-\infty}^{\infty} \left\{ 1 - \frac{1}{2\tau(1-\tau)} \gamma_h \right\} \cos(i\omega h). \quad (3.8)$$

ou, γ_h est le taux de passage à niveau de décalage h de $\{Y_t\}$ défini par :

$$\gamma_{t-s} = P\{(Y_t - \lambda)(Y_s - \lambda) < 0\}$$

Considérons le cas du calcul de distribution du quantile périodogramme à q fréquences différentes. Soit $\{Y_t\}$ un processus aléatoire, et soient $F_t(u)$ et $F_{t,s}(u, v)$ dénotent les fonctions de distribution marginale univariée et bivariée de $\{Y_t\}$. Supposons que les conditions suivantes sont satisfaites :

(1) $f_t(u) := F'_t(u)$ existe pour tout u , et $F_t(u + \lambda) - F_t(\lambda) = f_t(\lambda)u + O(u^{d+1})$ uniformément pour $|u| \geq u_0$.

(2) $F_t(\lambda) = \tau$ et $f_t(\lambda) = \kappa > 0$ pour tout t

(3) $\{Y_t\}$ est stationnaire en λ -passages à niveau, c'est-à-dire $P\{(Y_t - \lambda)(Y_s - \lambda) < 0\} = \gamma_{t-s}$ pour tout t et s .

(4) $\{Y_t\}$ est un processus linéaire de la forme $\lambda + \sum_{l=-\infty}^{+\infty} \phi_l e_l$, ou, $\{e_t\}$ est un processus aléatoire iid avec, $E(|e_t|) < \infty$ et $\{\phi_l\}$ est absolument sommable tels que :

$$\sum_{|l| > n^r} \phi_l = o(n^{-1}) \quad \text{as } n \rightarrow \infty,$$

pour certain constant $r \in [0, \frac{1}{4}]$. Alors, le théorème suivant peut être établi.

Théorème 3 Soit $\{Y_t\}$ satisfait les conditions (1)-(4). Soient $Q_n(\omega_j)$ défini par (3.6) avec $\hat{\beta}_\tau(\omega_j)$ est donnée par (3.4). Supposons que $S(\omega_j)$ est fini et borné loin de zéro. Alors, si $n \rightarrow \infty$,

$$\{Q_n(\omega_j)\} \sim \{(1/2)\eta^2 S(\omega_j)\xi_j\}$$

tels que " \sim " représente "asymptotiquement distribué comme", la séquence $\{\xi_j\}$ est une variable aléatoire iid centré avec une lois chi-deux de deux degrés de liberté, et $\eta^2 = \tau(1 - \tau)/\kappa$. Le terme $1/\kappa = 1/f_t(\lambda)$ représente ce qu'on appelle la fonction sparsité, (regardez [44] pour plus de détails pour l'estimation de cette quantité).

Le théorème (3) montre que le quantile périodogramme expose une distribution Chi-deux de deux degrés de liberté, avec échelle. Cela est similaire au comportement du

périodogramme classique [13, 17, 78]. La seule différence est la fonction échelle η^2 . La preuve du théorème peut être trouvée dans [52].

Le spectre de passage à niveau peut être décrit comme une représentation alternative de dépendance dans le domaine fréquentiel. Dans [52] l'auteur a démontré qu'il existe une transformation non triviale liant le taux de passage à niveau à la fonction d'auto-corrélation en supposant que le bruit est un bruit blanc Gaussien. Il est donc possible de trouver une relation directe entre le spectre de passage à niveau et la densité spectrale de puissance ordinaire.

La transformée de Fourier de la fonction d'auto-corrélation est considéré comme la représentation traditionnelle du dépendance dans le domaine fréquentiel. Dans le cas Gaussien, une relation simple entre le taux de passage à niveau et les coefficients d'auto-corrélations peut être obtenu. D'après les conditions (2) et (3), on peut écrire :

$$\gamma_h = 2(\tau - p_h),$$

telle que $p_h = P(Y_{t+h} \leq \lambda, Y_t \leq \lambda)$ est une probabilité non centrée. Si $\{Y_t\}$ est un processus Gaussien stationnaire de moyenne nulle et de variance σ^2 , et une fonction d'auto-corrélation $\{r_h\}$, alors il peut être montré que :

$$p_h = \int_{-\infty}^{\lambda/\sigma} \Phi\left(\frac{\lambda/\sigma - r_h x}{\sqrt{1 - r_h^2}}\right) \phi(x) dx,$$

ou, $\phi(x)$ et $\Phi(x)$ dénotent la densité de probabilité et la distribution cumulative d'une variable aléatoire normale standard. Cette relation qui relie la fonction d'auto-corrélation avec le taux de passage à niveau, nous donne la possibilité de trouver une relation directe entre le spectre de passage à niveau et la densité spectrale.

Le théorème (2) de [52] indique qu'une fuite spectrale se produit comme des petits pics, peuvent apparaître à d'autres fréquences (par exemple, les multiples entiers de ω_0), en particulier lorsque la périodicité est forte. Cette fuite est encore plus forte lorsque multiples périodicités sont présentes dans les données, ce qui affecte l'appartenance des

vraies fréquences. Pour surmonter ce problème une version régularisée du Quantile périodogramme est proposée.

3.2 Le quantile périodogramme pénalisé

Les méthodes de pénalisation (régularisation) sont une classe d'algorithmes qui transforment un problème d'optimisation sous contraintes en un problème sans contraintes en ajoutant un terme appelé pénalité, afin de garantir la convergence vers la solution.

Dans notre développement, nous avons utilisé la fonction de pénalisation L_1 . Cette approche a la capacité de réduire certains des prédicteurs vers zéro, et certains d'entre eux pourraient même être réduits à zéro exactement, due à la forme géométrique de la fonction L_1 .

Dans le contexte de l'estimation des paramètres en régression, la notion de propriétés oracles concerne la consistance de l'estimation des paramètres et de la sélection des variables. Ainsi, une procédure qui ne possède pas les propriétés oracles peut entraîner une consistance dans la sélection des variables, mais pas nécessairement dans l'estimation des paramètres, et vice versa, ou ni l'une ni l'autre ([25], [109]). Il a été démontré que le LASSO ne possède pas les propriétés oracles (non-biaisé, sparsité et continuité).

En fait, nulle de la famille L_q ne possède les propriétés d'oracle. Dans [109] l'auteur a donné une définition généralisée de la fonction LASSO en introduisant une fonction de poids. Il a proposé de prendre la réciproque des estimations ordinaires des moindres carrés avec une certaine puissance. Dans le cas de la régression quantile, les estimations de régression quantile ont été considérés comme des poids [103].

Dans cette partie, Le Quantile périodogramme pénalisé est défini en ajoutant un terme de pénalisation à la définition (3.4). Ainsi, Nous introduisons le quantile périodogramme du LASSO (LQP) comme suit :

$$Q_{\tau, \alpha_n}^{(L)}(\omega_j) = \frac{n}{4} \left\| \widehat{\beta}_{\tau, \alpha_n}^{(L)}(\omega_j) \right\|_2^2, \quad (3.9)$$

ou

$$\widehat{\beta}_{\tau, \alpha_n}^{(L)}(\omega_j) = \arg \min_{\beta \in \mathbb{R}^2} \sum_{t=1}^n \rho_{\tau}(Y_t - \lambda - x_t^T(\omega_j)\beta) + \alpha_n \|\beta\|_1. \quad (3.10)$$

tels que, $\|\cdot\|_1$ représente la norme L_1 , $\alpha \geq 0$ est le paramètre de régularisation et $\|\beta\|_1$ dénote la fonction de pénalité de β . Par conséquent, le *quantile périodogramme du LASSO adaptatif* (ALQP) est défini comme suit :

$$Q_{\tau, \alpha_n}^{(AL)}(\omega_j) = \frac{n}{4} \left\| \widehat{\beta}_{\tau, \alpha_n}^{(AL)}(\omega_j) \right\|_2^2, \quad (3.11)$$

avec,

$$\widehat{\beta}_{\tau, \alpha_n}^{(AL)}(\omega_j) = \arg \min_{\beta \in \mathbb{R}^2} \sum_{t=1}^n \rho_{\tau}(Y_t - \lambda - x_t^T(\omega_j)\beta) + \alpha_n \sum_{i=1}^2 \vartheta_i |\beta_i|, \quad (3.12)$$

ou, $\vartheta_i = 1/|\widehat{\beta}_{\tau}^{(i)}|^d$ pour certains $d > 0$.

Soit $A_j^{(n)} = \{i : \widehat{\beta}_{\tau, \alpha_n}(\omega_j) \neq 0\}$ le sous-ensemble de paramètres non nuls pour chaque fréquence ω_j . La sélection de variable LASSO est consistante si et seulement si $\lim_n P(A_j^{(n)} = A_j) = 1$, où A_j est le sous-ensemble des vrais paramètres. Les propriétés d'oracle de (3.12) sont développées en annexe. Les propriétés asymptotiques de l'estimateur ALQP sont détaillées ci-dessous.

3.3 Propriétés asymptotiques

Dans cette section, des résultats théoriques sur la distribution asymptotique du quantile périodogramme pénalisé défini dans l'équation (3.11) sont discutés.

Il est bien connu que le périodogramme ordinaire est lié au spectre d'autocorrélation. Le quantile périodogramme a une relation similaire avec le spectre dit de passage à niveau (level-crossing spectrum [52]). Dans le théorème suivant, nous montrons que le quantile périodogramme pénalisé préserve la même relation lorsque nous appliquons la pénalisation du lasso adaptatif.

Théorème 4 Soient $F_t(u)$ et $F_{t,s}(u, v)$ dénotent les fonctions de distribution marginale

univariée et bivariée de $\{Y_t\}$. Supposons que les conditions suivantes sont satisfaites pour τ et λ :

(C1) $f_t(u) := F_t'(u)$ existe pour tout u , et $F_t(u + \lambda) - F_t(\lambda) = f_t(\lambda)u + O(u^{d+1})$ uniformément pour $|u| \leq u_0$.

(C2) $F_t(\lambda) = \tau$ et $f_t(\lambda) = \kappa > 0$ pour tout t

(C3) $\{Y_t\}$ est stationnaire en λ passage à niveau, c'est à dire $P\{(Y_t - \lambda)(Y_s - \lambda) < 0\} = \gamma_{t-s}$ pour tout t et s .

(C4) $\{Y_t\}$ est un processus m -dépendent ou un processus linéaire de la forme $\lambda + \sum_{l=-\infty}^{+\infty} \phi_l e_t$, ou $\{e_t\}$ est une suite iid aléatoire avec $E(|e_t|) < \infty$ et $\{\phi_l\}$ est une suite déterministe absolument sommable pour laquelle

$$\sum_{|l| > n^r} \phi_l = o(n^{-1}) \quad \text{as } n \rightarrow \infty$$

pour une constante $r \in [0, \frac{1}{4}]$.

Étant donné $Q_{\tau, \alpha_n}^{(AL)}$ définie par (3.11), et supposons que $S(\omega)$ est fini et borné loin de zéro. Supposons que A_j soit le vrai modèle de sous ensemble pour chaque fréquence, soit p_j la longueur de A_j . Alors, comme $\alpha_n/\sqrt{n} \rightarrow 0$ et $\alpha_n n^{(d-1)/2} \rightarrow \infty$ on a :

$$\{Q_{\tau, \alpha}^{(AL)}(\omega_j)\} \stackrel{A}{\sim} \left\{ \frac{1}{2} \eta^2 S(\omega_j)(\xi_{p_j}) \right\},$$

avec $\eta^2 = \tau(1 - \tau)/\kappa^2$, et $\xi_{p_j} \sim \chi_{p_j}^2$, $p_j \in \{1, 2\}$.

Le théorème (4) indique que l'ALQP a un comportement asymptotique similaire au QP, dans le sens qu'ils sont tous les deux proportionnels au spectre du quantile ($\eta^2 S(\omega_j)$) tel que défini dans [52]). La différence provient de la distribution limite, qui est une khi-deux à deux degrés de liberté pour toutes les fréquences dans le cas du QP, alors que, dans le cas du ALQP, elle pourrait être une chi-deux à 1 degré de liberté pour certaines fréquences.

Le théorème 4 montre également que le biais de modélisation introduit par le terme de régularisation de la fonction LASSO adaptative doit être nul lorsque les coefficients sont suffisamment grands. Ainsi, l'ajout d'un terme de pénalité au quantile périodogramme ne devrait produire aucun biais.

L'association des estimateurs quantiles au spectre de passage à niveau est bénéfique. Ils partagent également l'invariance à la distorsion non linéaire. Cela signifie que toute transformation non linéaire sans mémoire des données affecte uniquement la constante d'échelle dans la distribution asymptotique [52]. Alors que la non-linéarité des données déforme entièrement le spectre de puissance [101]. Les processus de passages à niveau sont connus pour avoir des informations riches sur la structure de dépendance des séries chronologiques et sont largement utilisés dans les applications de traitement de signal.

[37] et [23] ont proposé une forme équivalente du quantile périodogramme de [52]. Leurs approches sont basées sur les processus de passages à niveau, ce qui est équivalent aux spectres de quantiles définis dans (3.8). Ils ont fourni des estimateurs consistants en évitant l'estimation de la fonction de dispersion ($f^{-1}(\lambda)$). Ces approches pourraient également bénéficier de la simplicité du calcul en utilisant la transformée de Fourier rapide (FFT). Cependant, la perte d'informations lors de la transformation des séries chronologiques en processus binaire peut entraîner une dégradation de l'efficacité statistique. [52] a montré que le périodogramme de passage à niveau perd la puissance de détection de 11 % - 28 % en fonction de la probabilité de fausse alarme.

3.4 Simulation

Dans cette section, nous fournissons quelques résultats numériques pour démontrer l'efficacité du quantile périodogramme pénalisé, et pour confirmer les conclusions théoriques discutées dans la section précédente. Nous montrons que le quantile périodogramme pénalisé est un outil puissant pour la détection des périodicités cachées dans des parties différentes de la distribution, avec plusieurs périodicités et lorsque nous nous écartons des hypothèses standard sur le bruit.

Le choix du paramètre de régularisation α est très important car, il a une forte relation avec la sparsité du modèle. Il existe de nombreux outils utilisés dans la littérature pour choisir un paramètre optimale, comme la cross validation généralisé, critère d'information Akaike et critère d'information bayésien [25]. Dans [95] les auteurs montrent que l'utilisation de la pénalité lasso adaptatif avec le critère de sélection BIC peut être aussi efficace qu'une procédure d'oracle.

Nous avons choisi le paramètre α en utilisant la minimisation du critère (BIC) comme suit :

$$BIC = \log\left(\frac{1}{n} \sum_{t=1}^n \rho_{\tau}(Y_t - \beta^T x_t)\right) + \frac{\log(n)}{n} edf$$

Où, edf est le nombre de coefficients non nuls dans $\hat{\beta}$. Des techniques de validation croisée pourraient également être utilisées dans ce contexte. Néanmoins, une procédure automatique de l'estimation du paramètre de régularisation minimum est plus utile. Une discussion de la sensibilité du quantile périodogramme pénalisé au choix de α est également fournie dans cette section.

Dans nos simulations, nous avons utilisé le modèle suivant (utilisé dans [52]). Soit le processus $\{Y_t\}$:

$$Y_t = a_t X_t \tag{3.13}$$

$$a_t = b_0 + b_1 \cos(\omega_0 t) + b_2 \sin(\omega_1 t),$$

avec, $(b_0 = 1, b_1 = 0.9, b_2 = 1)$ et $\{X_t\}$ est un processus AR(2), satisfait

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \xi_t \quad (t = 1, \dots, n) \tag{3.14}$$

avec, $\phi_1 = 2r \cos(\omega_c)$ et $\phi_2 = -r^2$ ($r = 0.6$, et $\omega_c = 2\pi * 0.22$), et $\{\xi_t\}$ est le bruit du modèle.

Dans notre étude, nous avons considéré le cas $\omega_0 = 2\pi \times 0.1$ et $\omega_1 = 2\pi \times 0.12$ afin de tester la capacité des différents estimateurs à détecter plusieurs fréquences voisines. Nous allons simuler en utilisant différents types de bruits, pour comparer la robustesse des estimateurs contre la déviation des hypothèses standards sur le bruit.

Comme le périodogramme ordinaire et le quantile périodogramme, le ALQP pourrait également servir comme une représentation de la dépendance dans le domaine des fréquences pour les séries chronologiques sans périodicités cachées. La figure (3.1) repré-

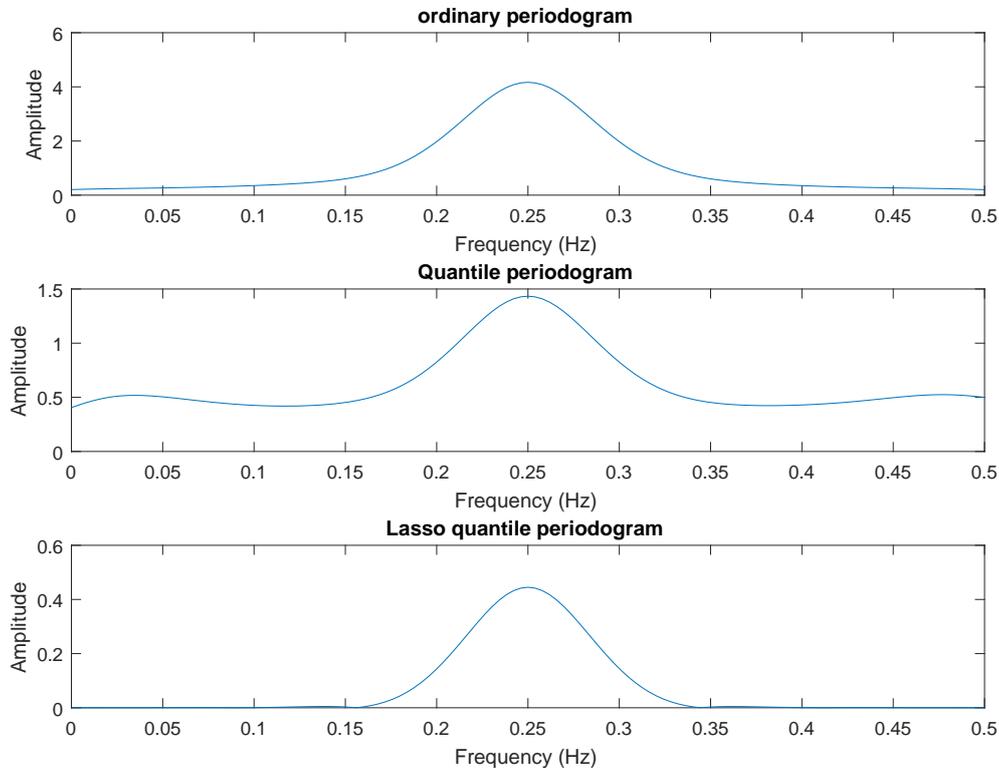


FIGURE 3.1: Moyenne de l'ensemble de 1000 périodogrammes lissés d'un processus AR(2) définie dans (3.14) avec un bruit blanc. (a) montre le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La taille de l'échantillon utilisée est $n = 200$. Le quantile $\tau = 0.8$.

sente la moyenne d'un ensemble de 1000 réalisations des périodogrammes lissés d'une série temporelle AR(2) définie dans l'équation (3.14). Les périodogrammes lissés ont été obtenus à l'aide de la fonction Smooth.spline de Matlab. Les figures (3.1) montrent que le quantile périodogramme pénalisé se comporte de la même manière que le quantile périodogramme et le périodogramme ordinaire présentant une forme de "cloche" attendu pour une densité spectral d'un processus AR(2).

Dans une première expérimentation, les estimateurs considérés sont effectués pour une seule réalisation de la série 3.13 avec différents types de bruits. La figure 3.2 présente les résultats de l'expérimentation. Le quantile $\tau = 0.8$ a été utilisé pour les estima-

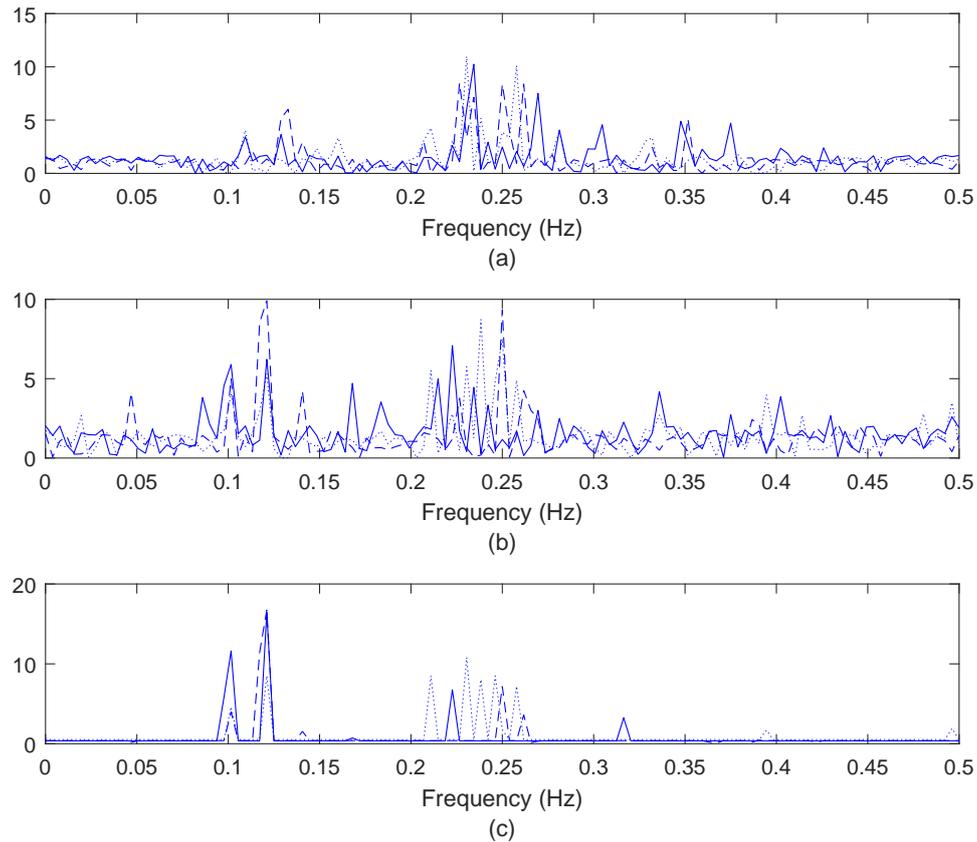


FIGURE 3.2: Estimateurs de la densité spectrale pour trois types de bruit pour une seule réalisation pour le modèle défini par (3.13). (a) montre le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La ligne pointillée (.) est pour le bruit gaussien standard, la ligne (-) est pour le bruit gaussien de grande variance et la ligne solide est pour le bruit de student-t. La taille de l'échantillon utilisé est $n = 200$. Le quantile utilisé est $\tau = 0.8$.

teurs quantiles. On note que seulement notre nouvel estimateur détecte clairement les périodicités avec tous les types de bruit. La présence de plusieurs périodicités produit d'autre pics dans le quantile périodogramme ce qui affecte l'appartenance des vraies périodicités. Alors que le périodogramme ordinaire ne détecte aucune périodicité.

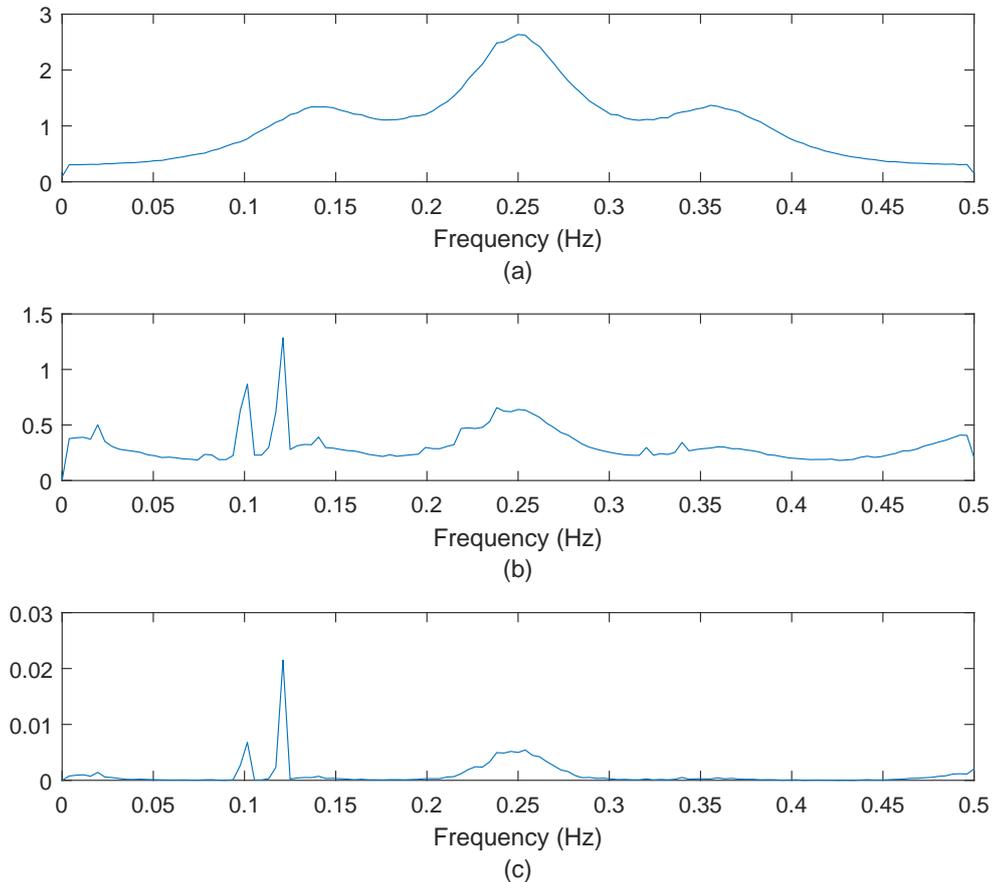


FIGURE 3.3: Moyenne de 1000 réalisations de densités spectrales estimées en utilisant le bruit de student-t pour un processus AR (2) simple (équation (3.14)) (a) représente le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé. La taille de l'échantillon utilisé est $n = 200$. Le quantile utilisé $\tau = 0.8$

Dans la figure 3.3, une moyenne d'un ensemble de 1000 réalisations des estimateurs spectraux appliqués à la série (3.13) avec un bruit de student-t, et avec $\tau = 0.8$ pour les estimateurs quantiles.

On note bien l'échec du périodogramme ordinaire, auquel il faut s'attendre, car il n'est pas efficace quand nous nous écartons des conditions normales du bruit. Le quantile périodogramme et le quantile périodogramme pénalisé détectent les périodicités correc-

tement. Le quantile périodogramme, cependant, souffre de plusieurs pics supplémentaires qui apparaissent à coté des vraies périodicité et ailleurs. Cela peut être expliquer par la fuite spectrale suggéré par le théorème (2) du [52].

En utilisant le périodogramme ordinaire, un test d'hypothèse pour la détection des périodicités cachées pourrait être proposé. Le test de Fisher est l'une des procédures les plus courantes. considérons la statistique :

$$g = \frac{\max_{1 \leq k \leq m} (I_n(\omega_k))}{\sum_{k=1}^m I_n(\omega_k)} \quad (3.15)$$

L'idée est de rejeter l'hypothèse nulle selon laquelle la série temporelle est un bruit blanc Gaussien contre l'hypothèse alternative selon laquelle la série chronologique contient une composante périodique cachée si le périodogramme contient une valeur substantiellement supérieure à la valeur moyenne, c.-à-d si la statistique g est suffisamment grande. Ce test a été appliqué dans [52] en remplaçant le périodogramme ordinaire dans 3.15 par le quantile périodogramme. La même idée pourrait être appliquée ici également pour l'ALQP en supposant que l'ALQP du bruit blanc a une distribution χ_2^2 pondérée indépendante de la fréquence ω_j (voir [50]). Ceci est clairement visible dans la figure (3.4) où les distributions des estimateurs QP et ALQP sont comparées pour les données de bruit blanc.

Nous rejetons l'hypothèse nulle si la valeur P est inférieure à $\tilde{\alpha}$ au niveau $\tilde{\alpha}$.

Probabilité de fausse alarme	Périodogramme	Q p			PQP		
		$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
0.01	0	0.2380	0.5100	0.2120	0.9010	0.9830	0.8480
0.05	0.01	0.3950	0.6870	0.4150	0.9300	0.9870	0.8480
0.1	0.02	0.4950	0.7680	0.5080	0.9410	0.9890	0.8480

TABLE 3.1: Puissance de test de Fisher pour 1000 simulations de Monte Carlo pour la série temporelle définie dans le premier exemple en utilisant une périodicité ($\omega_0 = 0.1 \times 2\pi$).

Le tableau 3.1 montre la puissance de test de Fisher calculée à l'aide de la statistique g définie dans (3.15). 1000 simulations de Monte Carlo ont été effectuées pour les trois estimateurs spectraux appliqués aux série temporelle définie dans (3.13) avec $b_2 = 0$. Comme on peut le voir, au niveau 0.01, le quantile périodogramme pénalisé a 98%

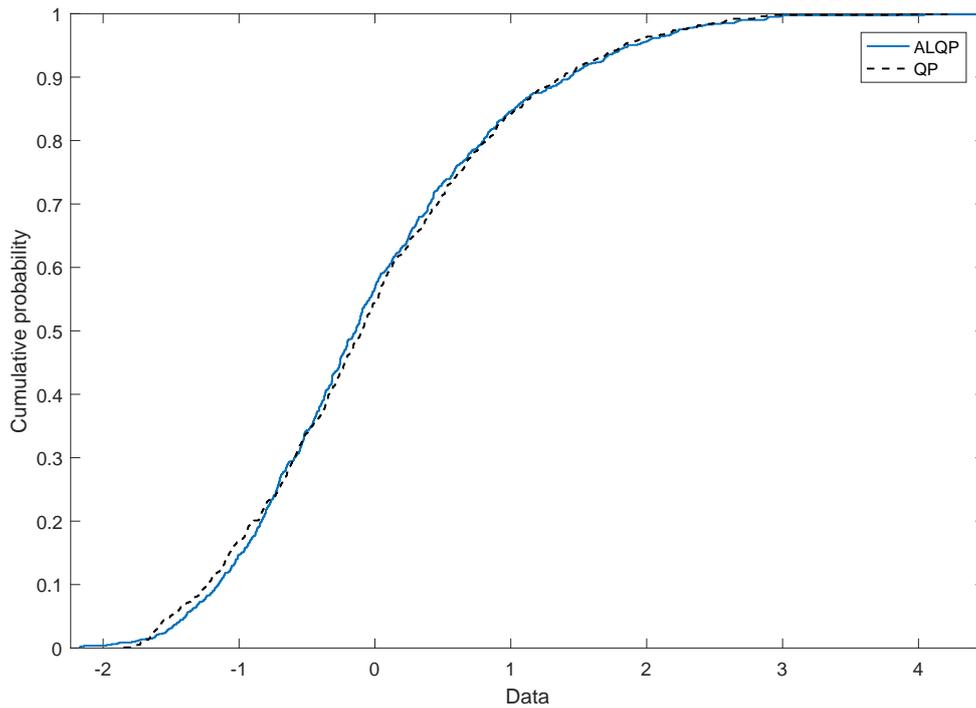


FIGURE 3.4: La distribution cumulative de l'ALQP et du QP calculés pour 1000 simulations de bruit blanc standard à une fréquence fixe.

chance de détecter la périodicité en utilisant $\tau = 0,8$ et $\alpha = 10$, alors que le quantile périodogramme n'a que 51% de chance. Nous notons que le quantile périodogramme pénalisé n'est pas aussi sensible que le quantile périodogramme pour le choix du quantile d'intérêt. Néanmoins, un choix correct de τ est nécessaire pour maximiser la puissance du test.

Pour la sensibilité du quantile périodogramme pénalisé au choix du paramètre de régularisation, on a effectué la même expérimentation avec différentes valeurs de α . Le tableau 3.2 affiche la puissance du test de Fisher pour 1000 simulations du quantile périodogramme pénalisé avec $\tau = 0.8$ et avec des valeurs de régularisation différentes. Notons que la puissance de détection dépend également de α . Cependant, un grand intervalle de valeurs fournit des performances similaires. Par conséquent, pour optimiser la détection de puissance, une gamme appropriée de α est nécessaire.

Le test de Fisher ne traite que la plus grande valeur du périodogramme ordinaire, bien que nous puissions le remplacer par la vraie périodicité. Une version étendue du test de

Probabilité de fausse alarme	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
0.01	0	0.0030	0.8990	0.9830
0.05	0.02	0.0150	0.9630	0.9870
0.1	0.02	0.0410	0.9780	0.9890

TABLE 3.2: Puissance de test de Fisher pour le quantile périodogramme pénalisé avec différentes valeurs de régularisation. 1000 simulations utilisées avec le quantile $\tau = 0.8$.

Probabilité de FA	Périodogramme		QP		ALQP	
	1ère per	2ème per	1ère per	2ème per	1ère per	2ème per
0.01	0.0000	0.0000	0.0440	0.0500	0.6600	0.6290
0.05	0.0000	0.0010	0.1160	0.1120	0.6930	0.6480
0.1	0.0010	0.0010	0.1580	0.1630	0.7130	0.6970

TABLE 3.3: Puissance de test de Fisher pour 1000 simulations de Monte Carlo pour la série temporelle définie dans le premier exemple en utilisant deux périodicités ($\omega_0 = 0.1 \times 2\pi$, $\omega_1 = 0.12 \times 2\pi$).

Fisher proposée par [100] pour la détection du second pic en omettant le terme du plus grand pic du dénominateur de la statistique g et en substituant $(n - 1)$ pour n dans la distribution de cette dernière [78]. Ainsi, si le plus grand pic devient significatif, nous pouvons tester un deuxième pic en utilisant :

$$g' = \frac{I_n(\omega_1)}{\{\sum_{k=1}^m I_n(\omega_k)\} - \max_{1 \leq k \leq m} (I_n(\omega_k))} \quad (3.16)$$

Si le deuxième pic se révèle significatif, on peut continuer à utiliser la même procédure jusqu'à ce que nous ne parvenions pas à obtenir un résultat significatif.

Le tableau 3.3 représente les résultats de 1000 simulations de Monte Carlo de la statistique de test définie dans (3.16) pour les trois estimateurs spectraux appliqués à la série temporelle définie dans (3.13) avec deux périodicités. Comme nous pouvons le voir, au niveau 0.1, le quantile périodogramme pénalisé a 71,3% de chance pour détecter la première périodicité et 69,7% pour détecter la deuxième, tandis que le quantile périodogramme n'a que 15,8% de chance pour la première et 16,3% de chance pour la détection de la deuxième périodicité. Cela pourrait s'expliquer par la fuite spectrale produite par le quantile périodogramme, qui est encore plus forte lorsque les données contiennent plusieurs périodicités.

Données réelles

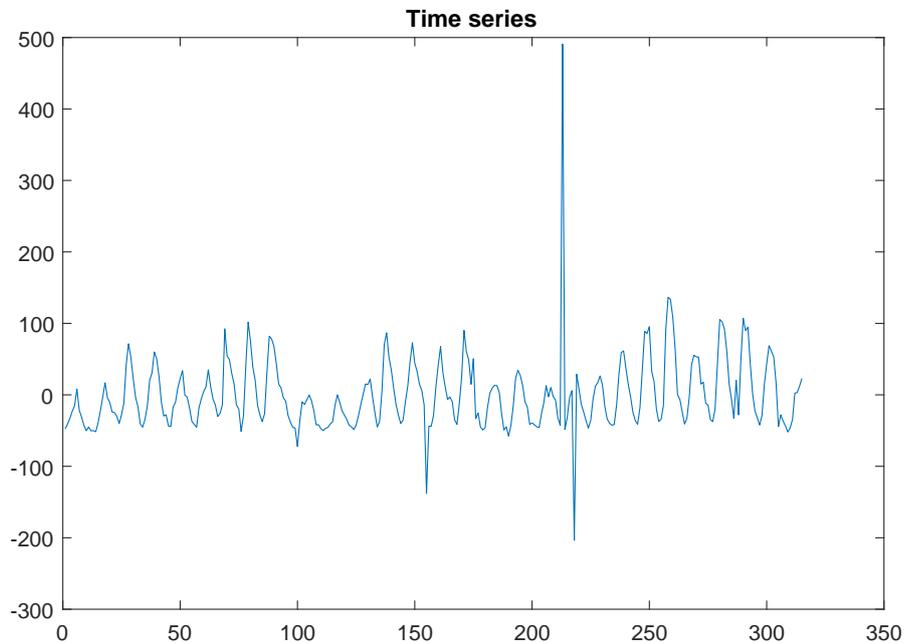


FIGURE 3.5: séries chronologiques de nombre annuel de taches solaires avec bruit de student-t supplémentaire

Le jeu des données annuel sur les taches solaires est largement utilisé dans le domaine de l'analyse spectrale. il est connu d'avoir un cycle de 11 ans. Le quantile périodogramme a montré sa capacité pour détecter un pic clair à $0,09 = 1/11$, avec des quantiles inférieurs. Le périodogramme ordinaire montre des pics supplémentaires des deux côtés de la vraie périodicité. Nous avons ajouté un bruit à l'ensemble de données à partir d'une distribution de student- t figure (3.5).

La figure (3.6) représente les différents estimateurs. Nous avons choisi le quantile $\tau = 0.25$ car il présente les meilleurs résultats pour les estimateurs quantiles. On peut voir la capacité du quantile périodogramme pénalisé à détecter clairement la seule et vraie périodicité dans ce cas.

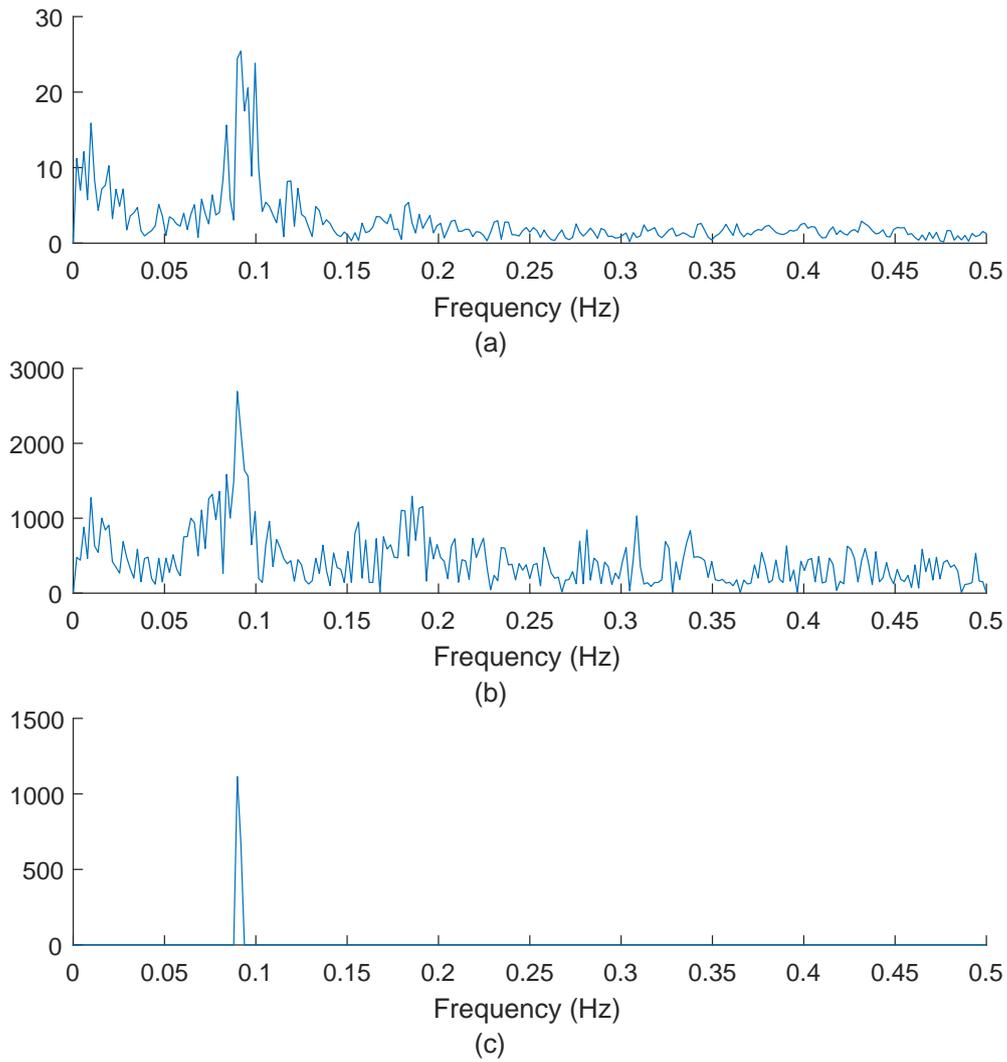


FIGURE 3.6: Estimateurs spectraux pour les nombres annuels de taches solaires avec ajout de bruit student-t. (a) représente le périodogramme ordinaire, (b) le quantile périodogramme et (c) le quantile périodogramme pénalisé

3.5 Conclusion

Dans ce travail, nous avons proposé un nouvel estimateur de domaine fréquentiel en ajoutant un terme de pénalisation au quantile périodogramme. Nous avons utilisé la norme L_1 du type lasso adaptatif, ce choix a été motivé par le fait que les propriétés d'oracle sont satisfaites. Nous avons démontré que cet estimateur a plus de chance de détecter la périodicité cachée que le périodogramme ordinaire et le quantile périodogramme avec une ou plusieurs périodicités. Nos simulations montrent que la performance du quantile périodogramme pénalisé est supérieure aux autres estimateurs spectraux, et est robuste quelque soit le bruit utilisé. Des recherches futures sont en cours pour développer des méthodes permettant de déterminer les valeurs appropriées du quantile et du paramètre de régularisation.

3.6 Appendice

3.6.1 Théorème de régression quantile régularisé

Pour la démonstration du théorème principal, nous devons développer la théorie asymptotique du problème de régression quantile pénalisé. Concernant les travaux effectués dans [52] et [50], un problème plus général de régression quantile pénalisé est considéré. Pour chaque $j = 1, \dots, q$, soit $\{x_{jt}\}$ une suite bornée de vecteurs déterministes dans \mathbb{R}^p pour un certain $p \in \mathbb{N}$.

Soit $Y_t = u_t + \varepsilon_t$ ou, $\{u_t\}$ est une séquence déterministe et $\{\varepsilon_t\}$ est un processus aléatoire m -dépendant ayant la fonction de distribution marginale $G_t(u)$, densité marginale $g_t(x) = G'_t(x)$, et $G_{t,s}(u, v)$ dénote la fonction de distribution marginale bivariée.

Soit $r_{ts}(u, v) = G_{ts}(u, v) - G_t(u)G_s(v)$. pour chaque $\beta_j \in \mathbb{R}^2$, soit $w_{jt} = x_{jt}^T \beta_j - u_t$. Supposons qu'il existe des nombres positifs d, c et N_0 et des matrices définies positives C_0 et Λ_0 , et soit $A = \{j : \hat{\beta}_j \neq 0\}$, le bon sous-ensemble et supposons que $A = \{1, 2, \dots, p\}$ tels que :

- (i) $g_t(w_{jt}) = O(1)$ uniformément en $t \in \mathbb{N}$ pour tout j .
- (ii) $G_t(u_t + w_{jt}) - G_t(w_{jt}) = g_t(w_{jt})u + O(u^{d+1})$ uniformément en $t \in \mathbb{N}$ pour $|u| \leq c$ et tout j .
- (iii) $C_{jkn} = n^{-1} \sum_{t=1}^n \sum_{s=1}^n r_{ts}(w_{jt}, w_{ks}) x_{jt} x_{ks}^T \geq C_0$ pour tout j, k et $n \geq N_0$
- (iv) $\Lambda_{jn} = n^{-1} \sum_{t=1}^n g_t(w_{jt}) x_{jt} x_{jt}^T \geq \Lambda_0$ pour tout j et $n \geq N_0$.

Sous ces conditions, les résultats suivants peuvent être établis.

Lemme 2 Soit $\widehat{\beta}_{j,n}^{(L)}$ définie par

$$\widehat{\beta}_{j,n}^{(L)} = \arg \min_{\beta \in \mathbb{R}^2} \sum_{t=1}^n \rho_\tau(Y_t - \lambda - x_{jt}^T \beta) + \alpha \sum_{i=1}^2 |\beta_i|.$$

Si $\alpha_n/\sqrt{n} \rightarrow 0$ et $\alpha_n \rightarrow \infty$ Alors,

$$\frac{n}{\alpha_n} \text{vec}[\widehat{\beta}_{j,n}^{(L)} - \beta_j]_{j=1}^q \xrightarrow{p} \arg \min(V_n(\delta)),$$

ou,

$$V_n(\delta) = -\delta^T \zeta_n + \frac{1}{2} \delta^T \Lambda_n \delta + \alpha \sum_{i=1}^2 [\delta_i \text{sgn}(\beta_i) I(\beta_i \neq 0) + |\delta_i| I(\beta_i = 0)].$$

ζ_{jn} a une distribution $\mathcal{N}(h_{jn}, C_{jkn})$.

Remarque 1 Dans de nombreux travaux récents, la pénalisation au LASSO s'est révélée être un bon estimateur uniquement dans certaines conditions restrictives. Pour résoudre ce problème, nous avons choisi d'utiliser la pénalité du LASSO adaptatif proposée dans [109].

Ensuite, nous établissons les propriétés d'oracle du problème ALQP dans le cas particulier présenté dans cet article. Soit $\widehat{\beta}_{j,\alpha_n}^{(AL)}$ définie par (3.12). Supposons que $\widehat{\beta}_{j,\alpha_n}^{(AL)} = [\widehat{\beta}_{j,\alpha_n 1}^{(AL)}, \widehat{\beta}_{j,\alpha_n 2}^{(AL)}]$ ou $\widehat{\beta}_{j,\alpha_n 1}^{(AL)} \in \mathbb{R}^{p_j}$ et $\widehat{\beta}_{j,\alpha_n 2}^{(AL)} \in \mathbb{R}^{p-p_j}$.

Théorème 5 Supposons que (i) à (iv) sont satisfaites et que $\alpha_n/\sqrt{n} \rightarrow 0$ et $\alpha_n^{(d-1)/2} \rightarrow \infty$, alors on a

$$\sqrt{n} \text{vec}[\widehat{\beta}_{j,1}^{(AL)} - \beta_{j,1}]_{j=1}^q \stackrel{A}{\sim} \mathcal{N}(\theta_{A_j}, \Sigma_{A_j}),$$

ou $\Sigma_{A_j} = [\Lambda_{j,A_j}^{-1} C_{j,A_j} \Lambda_{j,A_j}^{-1}]_{j=1}^q$, $\theta_{A_j} = \text{vec}\{\Lambda_{j,A_j}^{-1} h_{j,A_j}\}$, et $h_{j,A_j} = n^{-1/2} \sum_{t=1}^n \{\tau - G_t(w_{jt})\} x_{jt,A_j}$. $\stackrel{A}{\sim}$ dénote asymptotiquement distribué comme.

Preuve du théorème 5

Commençant par le cas d'un seul régresseur, on laisse tombé j et k pour la simplicité. Pour chaque vecteur $\delta \in \mathbb{R}^p$, supposons A soit le vrai modèle de sous ensemble avec une longueur p_0 , et soit :

$$Z_n(\delta) = \frac{1}{2} \sum_{t=1}^n \{\rho_\tau(u_t - v_t) - \rho_\tau(u_t)\} + \sum_{i=1}^p \frac{\alpha_n}{\sqrt{n} |\widehat{\beta}_{\tau,i}|^d} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right], \quad (3.17)$$

ou $u_t = Y_t - x_t^T \beta_0 = \varepsilon_t - w_t$ et $v_t(\delta) = n^{-1/2} x_t^T \delta$.

En prenant $Z_n(\delta) = Z_{1n}(\delta) + Z_{2n}(\delta)$, ou

$$Z_{1n}(\delta) = \frac{1}{2} \sum_{t=1}^n \{\rho_\tau(u_t - v_t) - \rho_\tau(u_t)\} \quad (3.18)$$

,et

$$Z_{2n}(\delta) = \sum_{i=1}^p \frac{\alpha_n}{\sqrt{n} |\widehat{\beta}_{\tau,i}|^d} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right]. \quad (3.19)$$

En suivant les étapes dans [51], Le résultat suivant pourraient être établis par la délimitation de $\{x_t\}$:

$$Z_{1n}(\delta) = \widetilde{Z}_{1n}(\delta) + o_p(1),$$

pour chaque $\delta \in \mathbb{R}^p$ fixe, avec $\widetilde{Z}_{1n}(\delta) = -\delta^T \zeta_n + \frac{1}{2} \delta^T \Lambda_n \delta$ et $\zeta_n \stackrel{A}{\sim} \mathcal{N}(h_n, C_n)$.

Pour la deuxième partie Z_{2n} , si $\beta_i \neq 0$, alors $|\widehat{\beta}_{\tau,i}|^d \xrightarrow{p} |\beta_i|^d$ et

$$\sqrt{n} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right] \rightarrow \delta_i \operatorname{sgn}(\beta_i).$$

Alors par le théorème de Slutsky,

$$\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right] \xrightarrow{p} 0,$$

et si $\beta_i = 0$, alors

$$\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right] = |\delta_i|,$$

et $\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \xrightarrow{p} \infty$. Par conséquent, on a

$$Z_n(\delta) = \begin{cases} \delta_A^T \zeta_n^A + \frac{1}{2} \delta_A^T \Lambda_A \delta_A & \text{si } \delta_i = 0, \quad \forall i \notin A \\ \infty & \text{autrement} \end{cases}$$

δ_A^T et $\zeta_n^A \in \mathbb{R}^{p_0}$, Λ_A est une matrice $p_0 \times p_0$. De plus, sous l'hypothèse (iii), $\tilde{Z}_{1n}(\delta)$ a un minimum unique $\tilde{\delta}_{1,n} = \Lambda_n^{-1} \zeta_n$ avec $\Lambda_n \geq \Lambda_0$ for $n \geq N_0$. Alors, par la convexité de $Z_{1n}(\delta)$, on peut montré que $\widehat{\delta}_{1,n} - \tilde{\delta}_{1,n} \xrightarrow{p} 0$, ou $\widehat{\delta}_{1,n}$ dénote le minimum de $Z_{1n}(\delta)$. Par conséquent, $\widehat{\delta}_{1,n}$ a la même distribution asymptotique comme $\tilde{\delta}_{1,n}$.

La fonction $Z_n(\delta)$ pourrait être reparamétré en fonction de $\beta = \beta_0 + \delta n^{-1/2}$ comme :

$$Z_n(\delta) = \frac{1}{2} \sum \{ \rho_\tau(Y_t - x_t^T \beta) - \rho_\tau(u_t) \} + \sum_{i=1}^p \frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \left[\left| \beta_i + \frac{\delta_i}{\sqrt{n}} \right| - |\beta_i| \right],$$

Comme cette fonction est minimisée à $\beta = \widehat{\beta}_{\tau,\alpha_n}^{(AL)}$, il s'ensuit que

$$\widehat{\delta}_n = \frac{1}{\sqrt{n}} (\widehat{\beta}_{\tau,\alpha_n}^{(AL)} - \beta_0).$$

En utilisant les mêmes arguments dans [43] et [44], et supposant $\delta = (\delta_1^T, \delta_2^T)^T$ ou δ_1

contient les premiers p_0 éléments de δ . Alors, on a

$$\widehat{\delta}_{2,n} \xrightarrow{d} 0$$

$$\text{and } \widehat{\delta}_{1,n} \xrightarrow{d} \mathcal{N}(\theta_A, \Sigma_A).$$

Avec $\Sigma_A = \Lambda_A^{-1} C_A \Lambda_A^{-1}$, $\theta_A = \Lambda_A^{-1} h_{p_0}$. Maintenant, considérons le cas de plusieurs régresseurs. Comme indiqué dans [51], soit $q = 2$, pour prouver la normalité asymptotique de $\widehat{\beta}_{1n}$ et $\widehat{\beta}_{2n}$, il suffit de considérer $Z_n(\delta_1, \delta_2) = Z_n^{(1)}(\delta_1) + Z_n^{(2)}(\delta_2)$, ou $Z_n^{(j)}$ est défini de même que $Z_n(\delta)$ dans l'équation (3.17) en remplaçant x_t par le multirégresseur x_{jt} . En prenant $Z_{1n}^{(j)}$ et $Z_{2n}^{(j)}$ tels que définies dans les équations (3.18) et (3.19) respectivement. Il a été démontré que $Z_{1n}^{(j)}(\delta_j) = -\delta_j^T \zeta_{jn} + \frac{1}{2} \delta_j^T \Lambda_{jn} \delta_j + o_p(1)$ uniformément dans $\delta_j \in \Delta$, tel que $\zeta_{jn} \stackrel{A}{\sim} \mathcal{N}(h_{jn}, C_{jjn})$. Par conséquent, on a besoin seulement de calculé $Cov(\zeta_{1n}, \zeta_{2n})$ qui est par définition $Cov(\zeta_{1n}, \zeta_{2n}) = \Lambda_{12n}$. L'argument restant est le même que pour un seul régresseur.

3.6.2 Preuve du théorème 4

Pour établir les résultats du théorème (4) il suffit d'appliquer les résultats du théorème (5) sur $Y_t - \lambda$ avec $x_{jt} = x_t(\omega_j)$, $\beta_0 = 0$, et $u_t = 0$ pour lesquels $\widehat{\beta}_{jn} = \widehat{\beta}(\omega_j)$, $\epsilon_t = Y_t - \lambda$, et $w_{jt} = 0$. Alors on a : $G_t(u) = F_t(\lambda + u)$ et $G_{ts}(u, v) = F_{ts}(\lambda + u, \lambda + v)$, cela implique $g_t(w_{jt}) = f_t(\lambda) = \kappa$ par (C1), alors (i) est satisfaite. De même, (ii) est satisfaite car d'après (C2) :

$$G_t(u + w_{jt}) - G_t(w_{jt}) = F_t(\lambda + u) - F_t(\lambda) = g_t(w_{jt})u + O(u^{a+1})$$

Maintenant, soit $\{\delta_j\}$ dénote la suite delta de Kronecker. Alors, on a par les identités trigonométriques

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega_j) x_t^T(\omega_k) = \frac{1}{2} \delta_{j-k} I + o(1) \quad (3.20)$$

On note que Λ_{j,A_j} est directement implicite de l'hypothèse (iv) et la forme de θ_{A_j} dans le théorème 5. Alors, on a d'après (C2)

$$\Lambda_{j,A_j} = n^{-1} \sum_{t=1}^n f_t(\lambda) x_{t,A_j}(\omega_j) x_{t,A_j}^T(\omega_j) = \frac{1}{2} \kappa I_{p_j} + o(1)$$

ou I_{p_j} est une matrice d'identité avec p_j éléments, alors (iii) est satisfaite. On a

$$x_t x_{t-h}^T = \begin{bmatrix} c_h c_t^2 + s_h c_t s_t & c_h c_t s_t - s_h c_t^2 \\ c_h c_t s_t + s_h s_t^2 & c_h s_t^2 - s_h c_t s_t \end{bmatrix}$$

tels que $c_t = \cos(\omega t)$ et $s_t = \sin(\omega t)$. Il a été démontré dans [51] que $\sum c_t^2 = \frac{1}{2}n + o(1)$, $\sum s_t^2 = \frac{1}{2}n + o(1)$, et $\sum c_t s_t = O(1)$. Par conséquent, pour chaque h fixe,

$$n^{-1} \sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_t(\omega_j) x_{t-h}^T(\omega_k) \rightarrow M_{j,h} = \frac{1}{2} \begin{bmatrix} c_{j,h} & -s_{j,h} \\ s_{j,h} & c_{j,h} \end{bmatrix}$$

Alors on a

$$n^{-1} \sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_{t,A_j}(\omega_j) x_{t-h,A_j}^T(\omega_k) \rightarrow \begin{cases} c_{j,h}, & \text{if } p_j = 1 \\ M_{j,h}, & \text{if } p_j = 2 \end{cases}$$

Par (C3), on a

$$r_{ts}(w_{jt}, w_{ks}) = F_{ts}(\lambda, \lambda) - F_t(\lambda)F_s(\lambda) = F_{ts}(\lambda, \lambda) - \tau^2 = \tau(1 - \tau) - (1/2)\gamma_{t-s}$$

Maintenant, comme $\{r_h\}$ est absolument sommable, on obtient $C_{j_n} \rightarrow \sum r_h M_{j,h}$. Parce que r_h est une fonction paire de h , on a $\sum r_h s_h = 0$, et $\sum r_h c_{j,h} = \frac{1}{2}\tau(1 - \tau)S(\omega_j)$.

Nous avons d'après les équations ci-dessus et en utilisant les résultats du théorème 5,

C_{jkn} de (iii) réduit à

$$\begin{aligned} C_{j,A_j} &= n^{-1} \sum_{|h|<n} \{\tau(1-\tau) - (1/2)\gamma_h\} \sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_{t,A_j}(\omega_j) x_{t-h,A_j}^T(\omega_k) \\ &= \frac{1}{2} \tau(1-\tau) S(\omega_j) \delta_{j-k} I_{p_j} + o(1), \end{aligned}$$

donc, (iv) est satisfaite. Enfin, sous (C2), nous avons

$$h_{jp_j} = n^{-1/2} \sum_{t=1}^n \{\tau - F_t(\lambda)\} x_{t,A_j}(\omega_j) = 0.$$

Par conséquent, on obtient

$$\sqrt{n} \text{vec}[\widehat{\beta}_{j,1}^{(AL)}(\omega_j)]_{j=1}^q \stackrel{A}{\approx} \mathcal{N}(0, (2/\kappa^2)\tau(1-\tau)\mathbf{S}),$$

ou $\mathbf{S} = \text{diag}\{S_{p_j}(\omega_1), \dots, S_{p_j}(\omega_q)\}$, pour lesquels $S_{p_j}(\omega_j) = S(\omega_j)$ si $p_j = 1$ et $S_{p_j}(\omega_j) = \{S(\omega_j), S(\omega_j)\}$ si $p_j = 2$. Cela implique que $Q_{\tau,\alpha}^{(AL)}(\omega_j)$ sont asymptotiquement indépendants et

$$\{Q_{\tau,\alpha}^{(AL)}(\omega_j)\} \stackrel{A}{\approx} \{(1/2)\eta^2 S(\omega_j)(\xi_{p_j})\},$$

pour chaque j avec $\eta^2 = \tau(1-\tau)/\kappa^2$ et $\xi_{p_j} \sim \chi_{p_j}^2$, $p_j \in \{1, 2\}$.

Chapitre 4

Le quantile périodogramme pénalisé pour la détection des patterns dans les signaux EEG

“Art is the imposing of a pattern on experience, and our aesthetic enjoyment is recognition of the pattern.”

– Alfred North Whitehead

4.1 Introduction

Les interfaces cerveau-ordinateur (Brain computer interfaces BCI) sont des systèmes de communication et de contrôle qui aident les gens à contrôler des appareils en utilisant l'activité électrique du cerveau. Les systèmes basés sur le BCI sont très importants dans de nombreux domaines, tels que la réhabilitation, la technologie d'assistance et la robotique [62, 21]. Les systèmes de BCI utilisent des signaux extraits de l'activité électrique du cerveau mesurés par un électroencéphalogramme (EEG) [69].

Les signaux EEG ont été largement utilisés dans l'étude des phénomènes cérébraux de

l'imagerie motrice (IM). Sur la base de la synchronisation liée à l'événement (ERS) et de la désynchronisation liée à l'événement (ERD), l'EEG associé au IM peut être caractérisé et reconnu. L'examen du contenu spectral du signal EEG a été étudié comme un moyen efficace pour avoir des informations discriminantes [3].

Il existe des régions spécifiques dans le cerveau qui sont responsables des mouvements corporels volontaires, ils sont situés dans le cortex précentral. Dans le système à 32 électrodes qui pourrait être utilisé pour extraire les informations de ces régions, les électrodes suivantes sont généralement utilisées (C3, C4, Cz, CP5, CP1, CP2, CP6)(Figure (4.1)). Les changements dans les mouvements corporels réels et imaginaires peuvent être observés dans des bandes de fréquences spécifiques connues sous le nom de rythmes sensorimoteurs (sensorimotor rhythms). Ils sont appelés le rythme μ (8-12Hz) et le rythme β (18-25Hz) [73, 74].

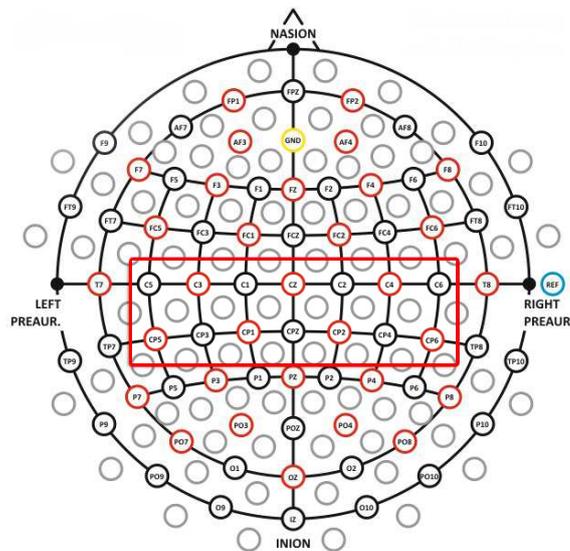


FIGURE 4.1: La région d'intérêt dans un système à 32 électrodes ([59]).

L'extraction de caractéristiques est l'une des tâches les plus difficiles dans l'imagerie motrice basée sur le BCI, car elle vise à extraire des informations distinctives à partir des signaux EEG. La nature des signaux EEG est difficile pour l'analyse spectrale classique, compte tenu de la complexité de l'activité cérébrale qui est une source de signaux non stationnaires et non linéaires [69, 75, 63].

Dans cette étude, on propose de nouvelles approches d'estimation spectrale pour sur-

monter ces problèmes.

Des techniques d'extraction de caractéristiques ont été développées dans la littérature dans le domaine temporel ainsi que dans le domaine fréquentiel [104].

Les modèles spatiaux communs (Common spatial patterns CSP) ont attiré beaucoup d'attention comme étant un outil d'extraction de caractéristiques pour les systèmes BCI [82, 106]. En utilisant l'estimation du spectre, les modèles autorégressifs (AR) [86, 39], les modèles autorégressifs multivariés (MVAR) [6] et le périodogramme ont été utilisés avec succès dans les systèmes de BCI [39, 58].

Afin de caractériser l'information non gaussienne contenue dans les signaux EEG, l'approche d'extraction de caractéristiques basée sur le bispectre (bispectrum) a été proposée [107]. L'analyse temps-fréquence a été reconnue comme un outil utile pour la composante oscillante des signaux EEG ainsi que pour l'analyse de temps-échelle dans l'imagerie motrice basée sur le BCI. Par conséquent, des méthodes de distribution-t quadratique et des méthodes d'ondelettes (wavelets) ont été étudiées [97, 26].

Dans [39], les auteurs ont effectué une analyse comparative des différentes méthodes d'extraction des caractéristiques, et ont conclu que la densité spectrale est l'approche la plus efficace et la plus robuste pour extraire des informations distinctives dans l'imagerie motrice. Dans cette thèse, on analyse différents estimateurs de la densité spectrale en tant que des méthodes d'extraction de caractéristiques pour les signaux d'imagerie motrice à base d'EEG.

En raison des propriétés de non-stationnarité et de non-gaussianité des signaux EEG, les techniques classiques d'analyse spectrale ne conviennent pas. Cette étude introduit un nouveau type d'estimateurs spectraux basés sur la régression quantile et la régularisation du lasso. La performance de ces nouvelles approches est comparée aux estimateurs classiques comme la FFT et le périodogramme de Welsh.

4.2 Activité cérébrale et les systèmes BCI

Le cerveau humain est un système complexe qui contient des interconnexions de milliards de neurones. Les neurones se connectent entre eux avec des impulsions électriques. Pour mesurer les amplitudes de ces impulsions, beaucoup de techniques ont été proposées dans la littérature comme : EEG (Électroencéphalogramme), fMRI (Functional Magnetic Resonance Imaging), MEG (Magneto Encephalography), NIRS (Near-infrared Spectroscopy), PET (Positron Emission Tomography), EROS (Eventrelated optical signal).

Motivé par le fait de pouvoir mesurer et analyser l'activité cérébrale, les systèmes BCI (Brain-Computer interface) ont attiré beaucoup d'attention dans les dernières années. Les systèmes BCI comme défini dans la première conférence internationale du BCI, sont des systèmes de communication qui ne dépendent pas des voies de sortie normales des nerfs et des muscles périphériques du cerveau [102]. Cette définition limite les BCI seulement sur les signaux du cerveau comprenant les signaux obtenus des muscles ou des nerfs périphériques.

Il existe deux types de BCI : endogène et exogène. Un système BCI exogène est un système qui a besoin d'un stimulus externe qui déclenche le phénomène neuronal de l'utilisateur. Figure (4.2) montre un exemple d'un système BCI exogène. Alors que dans le système endogène, il n'y a aucun stimulus affiché et les utilisateurs peuvent contrôler le système comme ils veulent.

Parmi les techniques d'enregistrement de l'activité cérébrale, le système EEG est un des systèmes les plus utilisés.

4.2.1 Electroencephalography

Le EEG est une technique d'enregistrement de l'activité cérébrale en utilisant des électrodes placés sur le scalp. Le système EEG peut capturer des activités corticales avec une résolution temporelle qui est inférieure d'une milliseconde [69]. Par contre,

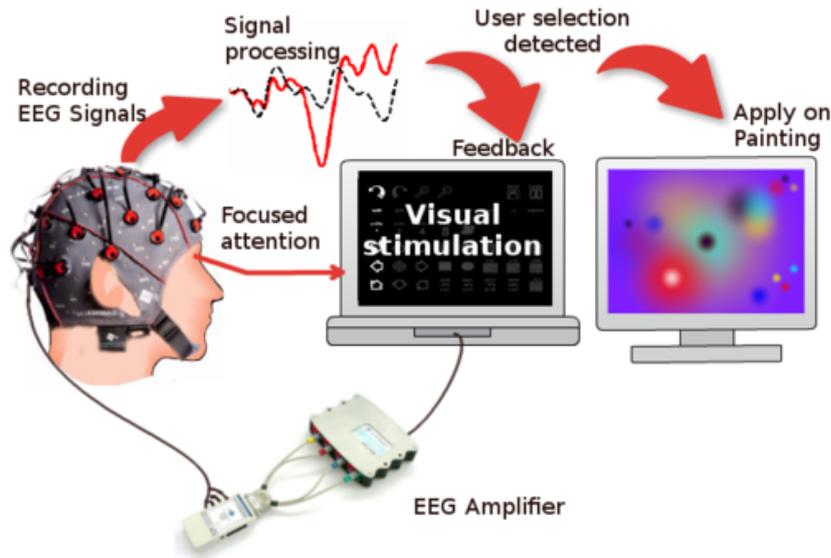


FIGURE 4.2: Un exemple illustratif d'un système BCI exogène ([59]).

l'EEG est connu pour avoir une faible précision spatiale puisqu'elle ne permet que de mesurer l'activité des neurones qui sont en surface. Le système EEG peut contenir un certain nombre d'électrodes qui peut aller de 1 à 256 électrodes.

Les positions des électrodes sont déterminées par rapport à des repères anatomiques spécifiques sur la tête et sont séparées par des distances également déterminées [69].

Chaque électrode a un nom composé d'une lettre et d'un chiffre. La première partie du nom provient de la partie du cerveau sur laquelle l'électrode est positionnée. F, T, P et O désignent respectivement les lobes frontal, temporal, pariétal et occipital. Les électrodes avec des nombres impairs sont placées sur l'hémisphère gauche et, les nombres pairs, sur l'hémisphère droit. Les électrodes numérotées avec 0 ou avec la lettre "z" sont positionnées sur l'axe central de la tête.

Chaque partie du cerveau est responsable sur une ou plusieurs tâches, la mémoire et la parole, le mouvement des bras et des jambes et la fonction de nombreux organes du corps. Ce qui nous intéresse dans ce travail, est la partie responsable sur les mouvements corporels volontaires y compris l'imagerie motrice. Cette partie est bien évidemment le cortex précentral (figure 4.3).

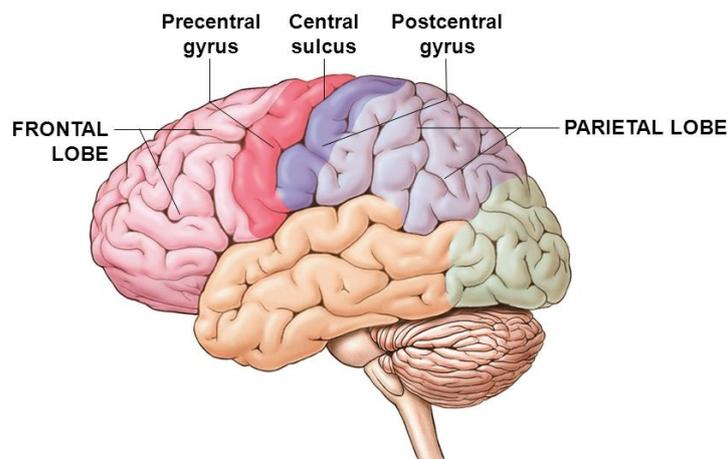


FIGURE 4.3: les différentes parties du cerveau humain.

Les signaux EEG

Les signaux EEG provient de l'activité synaptique synchronisée dans les neurones corticaux. Un signal EEG est une somme des tensions à différents dipôles neuronaux dans une région cérébrale donnée [59].

Les signaux EEG sont de nature hautement non-linéaire, non Gaussienne et non-stationnaire.

Les signaux EEG sont de nature oscillatoire. Ils peuvent être classés selon la bande de fréquence. Les différents rythmes de l'activité cérébrale sont présentés dans la figure (4.4). Cette classification a un rôle très important dans la recherche relié à l'EEG comme chaque oscillation est connectée à un état cognitif [12, 91].

- **delta (0 :5–4 Hz)** le rythme cérébral le plus lent, pendant le sommeil ce rythme est le plus fort
- **theta (4–8 Hz)** ne survient pas fréquemment chez les adultes.
- **alpha (8–12 Hz)** onde prédominante durant l'éveil, elle se présente sur la zone du cortex visuel, la plus forte lorsque les yeux sont fermés ou lorsque la personne est en état de relaxation
- **mu (8–12 Hz)** elle a le même gamme de fréquences que alpha, mais elle se

présente sur le cortex motrice

- **beta (18–25 Hz)** pour les signaux qui sont connectés à des états de vigilance et d'attention
- **gamma (30 >Hz)** pour les signaux qui sont reliés au traitement de l'information.

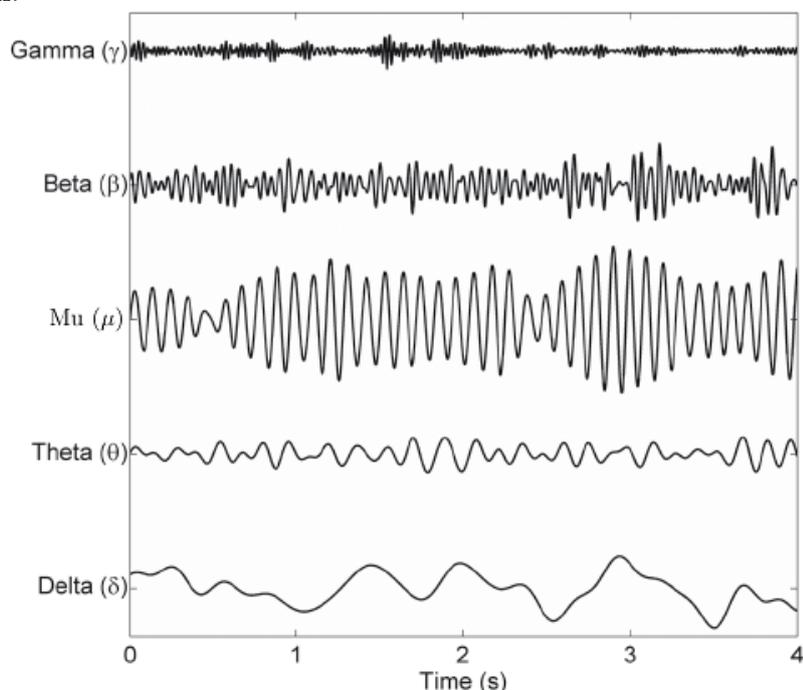


FIGURE 4.4: Les rythmes de l'activité cérébrale ([59])

4.2.2 L'imagerie motrice et la désynchronisation liée à un événement (ERD)

Les oscillations EEG qui sont reliées à l'imagerie motrice sont appelées les SMR (sensory motor rhythms). elles sont présentées par les bandes de fréquences μ et β (8-12Hz et 18-25Hz respectivement). Ces oscillations peuvent capturer les mouvements volontaires réels ou d'imagerie.

Une diminution de l'amplitude des SMR est appelée "event-related desynchronisation (ERD)" tandis qu'une augmentation de l'amplitude est appelée "event related synchronisation (ERS)".

Le phénomène ERD a été introduit dans les années 1970 par Gert Pfurtscheller pour

désigner le blocage des ondes de la bande mu en réaction à divers types de stimuli [49], par exemple : les mouvements volontaires ou de l'imagerie motrice. Dans le cas d'imagerie motrice, le phénomène ERD commence lorsque le sujet commence à imaginer un mouvement avec une partie du corps et se manifeste par une diminution de la puissance dans les bandes mu et bêta. Après cela, un phénomène différent se produit, la synchronisation événementielle (ERS), qui correspond à une augmentation de la puissance dans les bandes mu et bêta lorsque le sujet cesse d'imaginer un mouvement (Figure (4.5)).

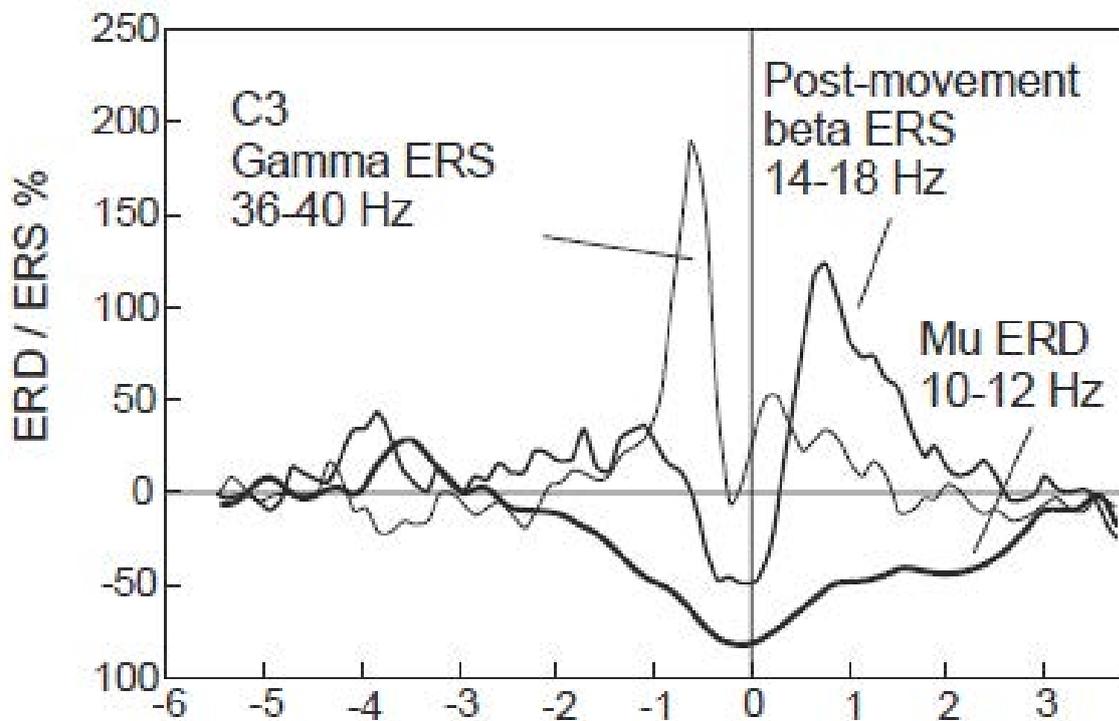


FIGURE 4.5: ERD et ERS durant l'imagerie motrice [68]

Le phénomène ERD / ERS lié aux actions d'imagerie motrice a différentes caractéristiques spatiales et force pour différents parties du corps : il est plus fort pour l'hémisphère controlatéral et plus faible pour l'hémisphère ipsilatéral. L'hémisphère controlatéral est l'hémisphère situé du côté du corps opposé à la partie du corps pour laquelle la tâche d'imagerie motrice est exécutée, alors que l'hémisphère ipsilatéral est du même côté du corps. Par exemple : si le sujet imagine un mouvement de la main droite, la désynchronisation et la synchronisation liées à un événement se produisent

principalement dans l'hémisphère gauche. De même, si le sujet imagine un mouvement de la main gauche, le ERD / ERS se produit principalement dans l'hémisphère droit. Les mouvements des pieds et des bras sont plus difficiles à séparer, car dans ce cas, de plus grandes surfaces du cortex moteur sont atteintes.

En enregistrant le signal, divers artefacts et interférences peuvent être enregistrer avec l'information. Cligner les yeux pendant la procédure d'acquisition, les activités musculaires et les signaux électriques du système peuvent être combinés avec le signal. Par conséquent, des techniques de débruitage sont nécessaires pour éliminer tout type d'information inutile.

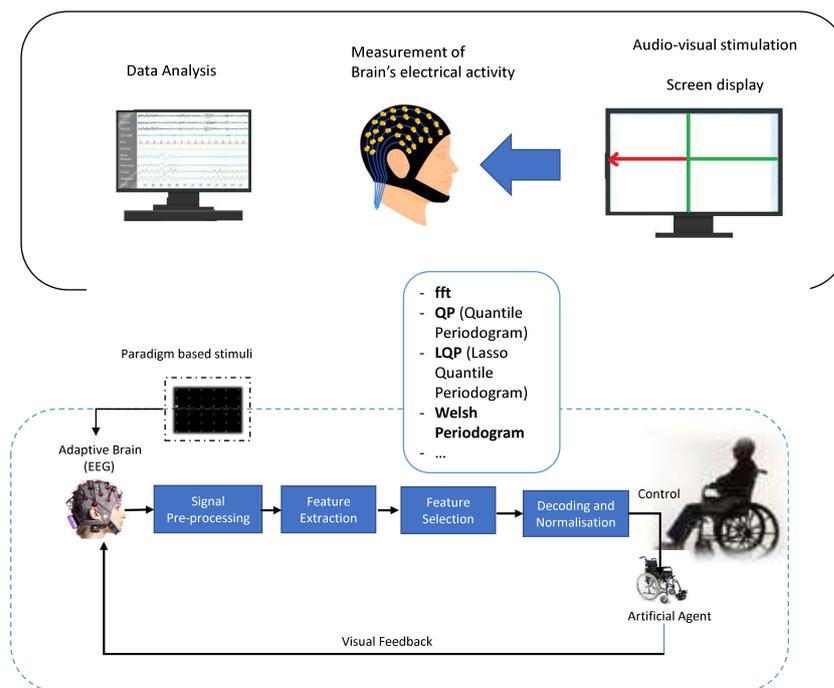


FIGURE 4.6: Résumé graphique de l'expérimentation.

4.3 Description du dataset

L'ensemble de données a été collecté lors d'une étude sur un projet d'imagerie motrice de la BCI conduit à l'Université de Pretoria en Afrique du Sud [59]. 5 sujets ont

participé, tous n'avaient aucune expérience préalable sauf le sujet 3 qui a participé à une session BCI quelques mois avant l'expérience. Les sujets doivent faire des tâches d'IM (main droite Vs. main gauche, main droite où repos) sur 5 séances, à l'exception du premier sujet qui n'en a fait que quatre. Chaque session contient 20 essais. L'essai consiste en 5 secondes, en commençant par un bip pour attirer l'attention du sujet, puis une flèche apparaît sur l'écran indiquant la tâche. L'enregistrement de l'essai commence après 3 secondes Figure [4.7].

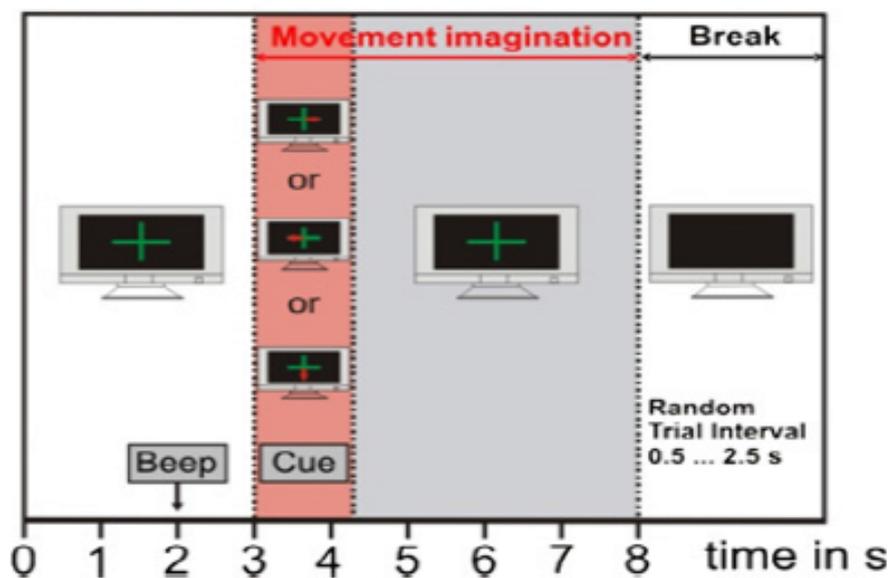


FIGURE 4.7: Scénario Off-ligne

Les enregistrements ont été effectués à l'aide d'un kit sans fil de 32 électrodes sèches actives, avec une fréquence d'échantillonnage de 250 Hz, bande passante entre 2 Hz et 60 Hz. avec un rejet à la fréquence de bruit Powerline (48Hz-52Hz) et filtrage en utilisant la moyenne commune (CAR).

4.4 Analyse des signaux EEG

Dans cette partie on présente quelques visualisations et une analyse des données présentées dans la section précédente.

Comme mentionné dans la partie 4.2.2 le phénomène ERD apparaît quand le sujet commence à imaginer le mouvement, ce qui mène à une décroissance dans les bandes de

fréquences mu et bêta. Une illustration de ce phénomène est présentée dans la figure 4.8 pour les données du sujet 2 (électrode C3). La figure représente la médiane sur tous les essais des différents estimateurs spectraux, en considérant l'imagerie de la main gauche. Les données ont été filtrés avec une bande passante entre 2Hz-60Hz, ainsi qu'une bande de rejet à la fréquence de bruit de courant (48Hz-52Hz). On peut voir d'après la figure 4.8, que tous les estimateurs spectraux présentent une forme générale similaire, ce qui montre la capture du phénomène ERD sous la forme d'une diminution des bandes de fréquences mu et bêta. Il est à noter que la bande de réjection (48Hz-52Hz) n'a pas d'effet important sur le quantile périodogramme. La figure 4.9 illustre la médiane de tous les essais du quantile périodogramme du LASSO appliqué aux données des sujets 2 et 5 avec l'électrode C3 lors du mouvement d'imagerie de la main gauche. L'estimation a été réalisée avec différentes valeurs de quantiles ($\tau = [0.25, 0.5, 0.75, 0.95]$). Il est à noter qu'il existe une différence notable dans le comportement des estimateurs du quantile lors du changement du quantile d'intérêt. Cela conforte l'idée que l'information la plus discriminante est associée à des parties spécifiques de la distribution. La différence du comportement est observée sous une forme de décalage et d'une modification des amplitudes des différents estimateurs pour certaines fréquences dans les bandes d'intérêt. On note, par exemple, la différence entre les trois premiers quartiles et le 95-quantile, autour de la fréquence 12Hz pour le sujet 5. Un choix approprié du quantile d'intérêt est donc nécessaire. Cela pourrait également être montré dans le cas du quantile périodogramme. Il est important de noter que le comportement des estimateurs spectraux change également d'un sujet à l'autre. Pour le sujet 2, l'estimateur se manifeste par un pic important autour de 3Hz, puis ça continue à diminuer jusqu'à qu'il atteint 60Hz. Alors que pour le sujet 5, l'estimateur montre un deuxième pic autour de 12Hz. Cela suggère que les bandes de fréquences pourraient être ajustées pour chaque sujet afin d'améliorer le taux de classification.

La figure 4.10 représente la médiane de tous les essais des estimateurs spectraux pour les tâches de l'imagerie de la main gauche et droite. Les données du sujet 2 sur l'électrode C3 ont été utilisées dans cette expérience. On peut noter qu'il existe une différence

entre les tâches dans le cas des estimateurs quantiles et en particulier pour le quantile périodogramme du LASSO. Dans les cas de la FFT et du périodogramme de Welch, il est difficile de voir la différence, les classes sont presque superposées.

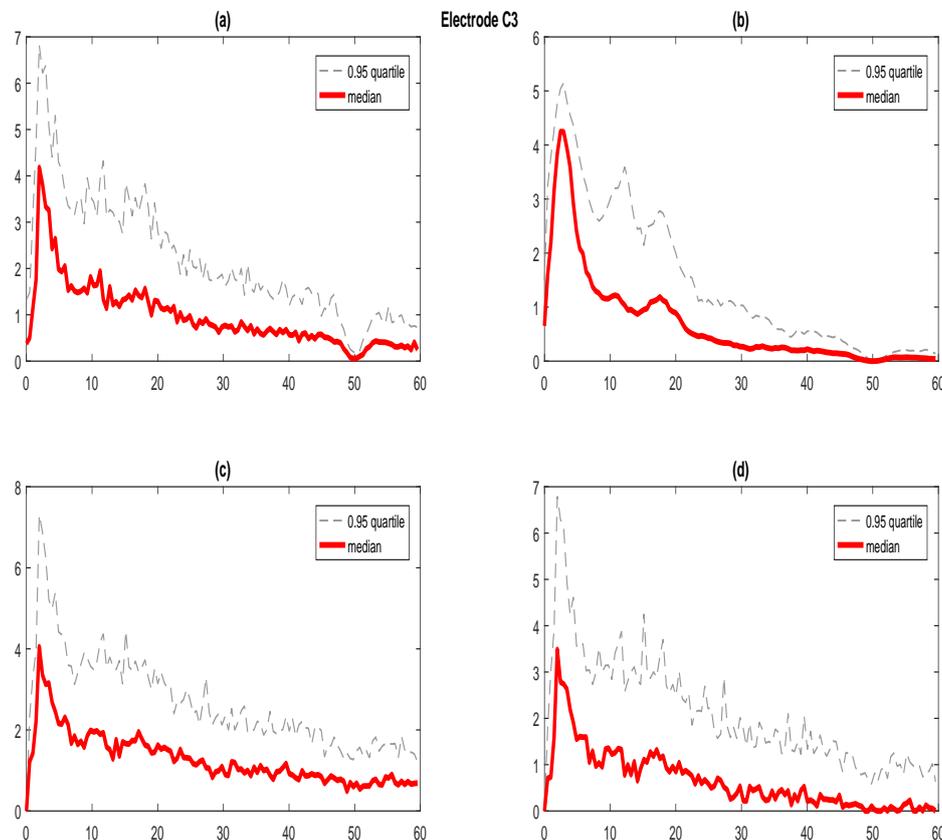


FIGURE 4.8: La médiane et le 0.95-quantile des différents estimateurs spectraux pour chaque tâche (sujet2, électrode C3). Pour les paramètres, (a) représente la FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO. Les paramètres ($\tau = 0.5, \alpha = 5$) ont été utilisés respectivement comme le quantile d'intérêt et le paramètre de régularisation pour les estimateurs quantiles.

4.5 La chaîne de traitement de signal

L'analyse des tâches d'IM dans cette étude est basée sur un protocole de classification statistique. La figure (4.11) résume la chaîne de traitement du signal.

Le protocole d'apprentissage commence par une simple technique de pré-traitement basée sur un filtre (CAR), en mettant l'accent sur l'utilisation des signaux EEG obtenus

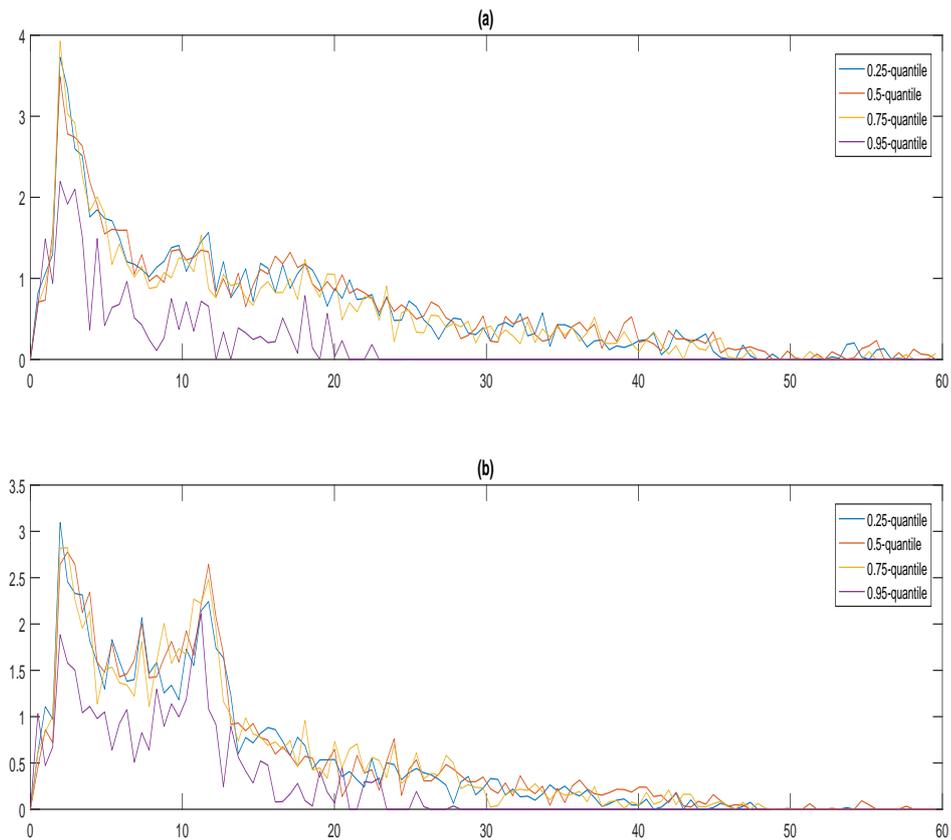


FIGURE 4.9: La médiane des Quantiles périodogrammes du LASSO pour la classe d'imagerie motrice de la main droite (électrode C3) pour tous les essais. (a) représente la médiane du Quantile périodogramme du LASSO appliqué sur les données du sujet 2 avec les quantiles (0,25, 0,5, 0,75 et 0,95). (b) représente le Quantile périodogramme du LASSO appliqué sur les données du sujet 5 avec les quantiles (0,25, 0,5, 0,75 et 0,95). $\alpha = 5$ a été utilisé comme le paramètre de régularisation.

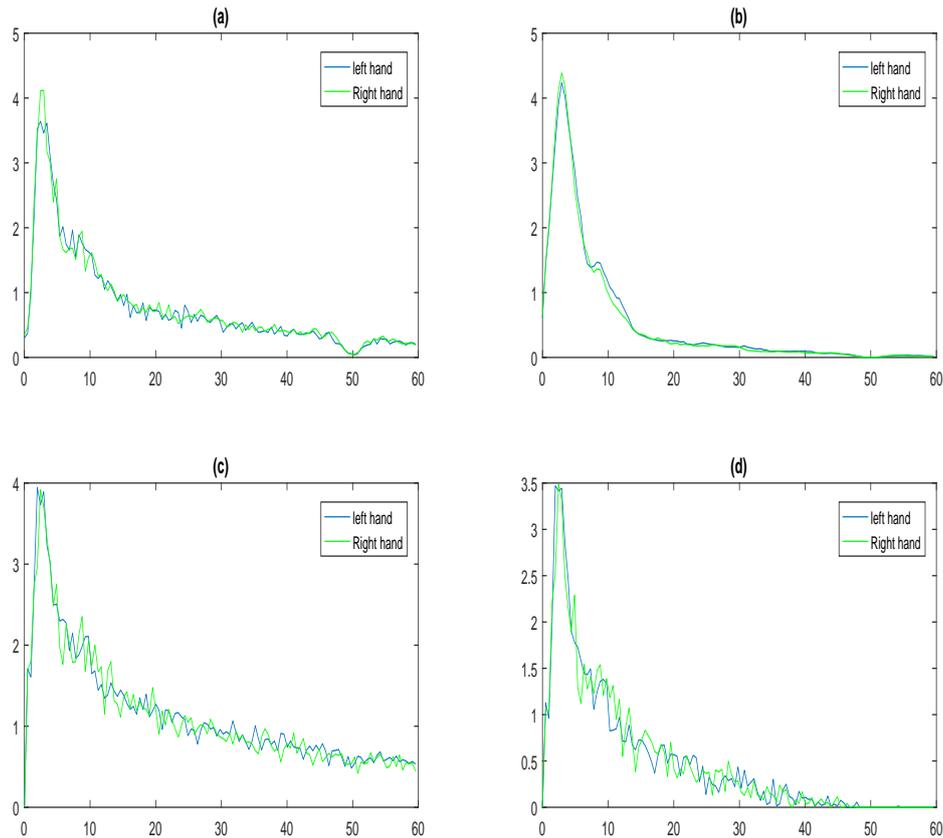


FIGURE 4.10: La médiane des différents estimateurs spectraux appliqués pour les classes d'imagerie motrice de la main droite et la main gauche (sujet 2, électrode C3) pour tous les essais. (a) représente la FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO. Les paramètres ($\tau = 0.5, \alpha = 5$) ont été utilisés respectivement comme le quantile d'intérêt et le paramètre de régularisation pour les estimateurs Quantile.

à partir des canaux C3, C4 et Cz comme couramment utilisé dans la littérature [39, 58].

Les approches spectrales utilisées dans la littérature pour l'imagerie motrice à base de BCI ont une information discriminante dans ce qu'on appelle les rythmes sensori-moteurs, qui sont respectivement les Bandes de fréquence μ (8-12Hz) et β (18-25Hz).

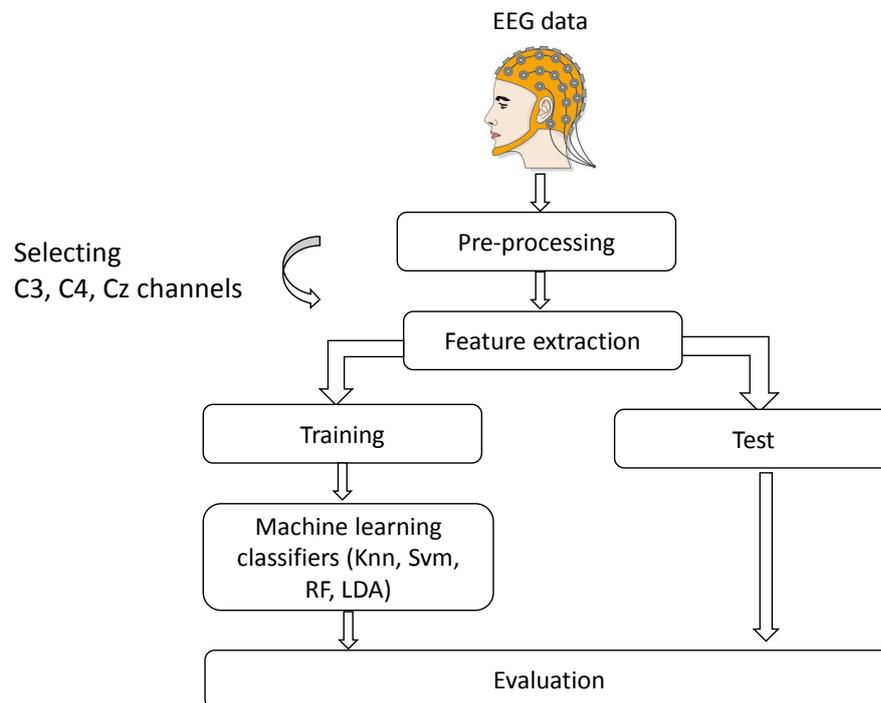


FIGURE 4.11: La chaîne de traitement de Signal

Dans cette étude, la qualité des caractéristiques extraites est examinée en utilisant le quantile périodogramme, le quantile périodogramme du Lasso, la FFT et le périodogramme de Welch, dans les bandes de fréquence μ et β , en mettant l'accent sur l'utilisation des signaux EEG à partir des canaux C3, C4 et Cz tels qu'adoptés couramment dans la littérature [39, 58].

Le périodogramme de Welch est une version modifiée du périodogramme ordinaire basé sur la méthode de Bartlett. Il vise à réduire le bruit du périodogramme en faisant la moyenne des petites fenêtres du périodogramme du signal à chaque fréquence [99].

Dans le cas du FFT et de périodogramme de Welch, le vecteur de caractéristiques peut être écrit comme :

$$F = (r_{\mu}^{C3}, r_{\beta}^{C3}, r_{\mu}^{C4}, r_{\beta}^{C4}, r_{\mu}^{Cz}, r_{\beta}^{Cz})$$

où $r_k^{C_i}$ représente l'amplitude sur les bandes de fréquences $k \in \{\mu, \beta\}$, de la représentation spectral des chaînes $C_i, i \in \{3, 4, z\}$.

Pour le périodogramme quantile, le vecteur de caractéristiques dépend du quantile d'intérêt τ , est donné par ce qui suit :

$$F_{\tau} = (r_{\mu, \hat{\tau}}^{C3}, r_{\beta, \hat{\tau}}^{C3}, r_{\mu, \hat{\tau}}^{C4}, r_{\beta, \hat{\tau}}^{C4}, r_{\mu, \hat{\tau}}^{Cz}, r_{\beta, \hat{\tau}}^{Cz})$$

Le quantile d'intérêt τ est estimé en utilisant la validation croisée (Cross Validation CV) sur les quantiles souvent utilisés ($\tau = 0,25, 0,5, 0,75$ et $0,95$).

Finalement, le quantile périodogramme du lasso dépend d'une variable supplémentaire qui est le paramètre de régularisation α . Par conséquent, le vecteur de caractéristiques est défini en utilisant $\hat{\tau}$ et $\hat{\alpha}$ comme suit :

$$F_{\tau, \alpha} = (r_{\mu, \hat{\tau}, \hat{\alpha}}^{C3}, r_{\beta, \hat{\tau}, \hat{\alpha}}^{C3}, r_{\mu, \hat{\tau}, \hat{\alpha}}^{C4}, r_{\beta, \hat{\tau}, \hat{\alpha}}^{C4}, r_{\mu, \hat{\tau}, \hat{\alpha}}^{Cz}, r_{\beta, \hat{\tau}, \hat{\alpha}}^{Cz})$$

Afin d'estimer le paramètre de régularisation $\hat{\alpha}$ on a utilisé une approche de CV dans l'intervalle $[10^{-2}, 10]$. Pour le quantile $\hat{\tau}$ on a utilisé la quantilté précédemment estimé dans le quantile périodogramme.

Enfin, la matrice de caractéristiques aura une taille de 100×6 pour chaque sujet, sauf le premier pour lequel la taille est de 80×6 .

4.6 Les algorithmes de classification

Les algorithmes de classification statistique visent à créer un modèle qui prédit si une nouvelle observation appartient à une catégorie ou une autre. Il crée une fonction

de décision en utilisant l'espace des caractéristiques pour produire une prédiction de nouvelles données.

Cette étude compare la performance de 4 algorithmes de classification d'apprentissage automatique. On considère le k -plus proches voisins (KNN), Machine à vecteurs de support (SVM), l'analyse discriminante linéaire (LDA) et les algorithmes de forêt aléatoire (RF). La mesure du score de précision a été utilisée comme un quantificateur de la performance des résultats prédits.

Le k -plus proches voisins (KNN)

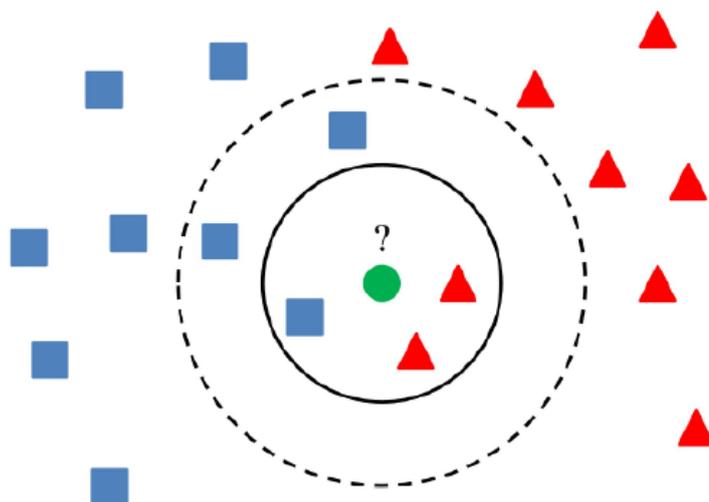


FIGURE 4.12: K plus proches voisins

Le k -plus proches voisins est un algorithme de classification non paramétrique qui prédit un nouveau vecteur en se basant sur un vote de k voisins de l'espace de caractéristiques ($k \in \mathbb{N}^+$) voir figure (4.12). les voisins peuvent être déterminés en utilisant une fonction de distance. Par exemple la distance Euclidienne :

$$d(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Les algorithmes KNN ne sont pas très populaires dans la communauté BCI, probablement parce qu'ils sont connus pour être très sensibles à la dimensionnalité [29], ce qui les a fait échouer dans plusieurs expériences BCI [[11], [65]]. Cependant, Lorsqu'ils sont

utilisés dans des systèmes BCI avec des vecteurs de caractéristiques dimensionnelles petites, les KNN peuvent être plus efficaces [15].

La machine à vecteurs de support (SVM)

L'algorithme de machine à vecteurs de support (Support vector machine) est un classifieur d'apprentissage supervisé qui peut assigner un nouveau vecteur de caractéristiques à une classe ou à une autre, en séparant l'espace d'entraînement à l'aide d'un hyperplan. L'hyperplan est estimé par une maximisation de la marge de sorte que la distance entre celui-ci et les deux côtés des données est maximisée (Figure(4.13)).

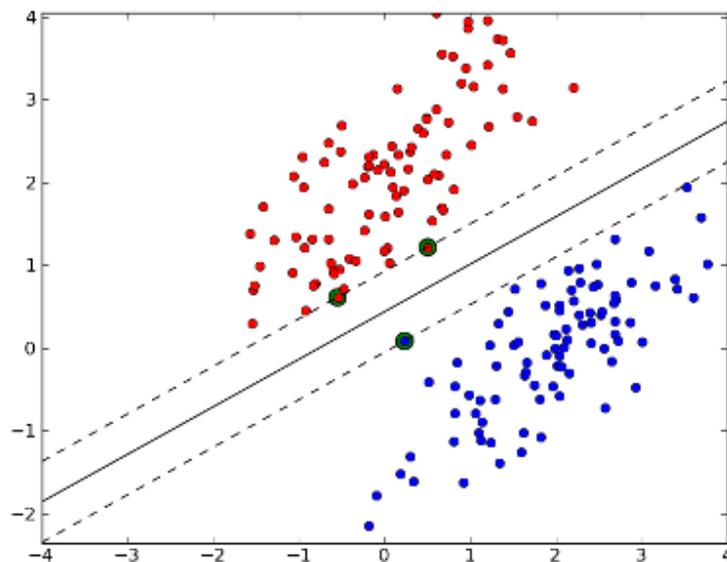


FIGURE 4.13: Support vector machine optimal hyperplane

Plus précisément, un SVM construit un hyperplan ou un ensemble d'hyperplans dans un espace de dimension élevée ou infinie, qui peut être utilisé pour la classification, la régression ou d'autres tâches, telles que la détection des valeurs aberrantes. une bonne séparation est obtenue par l'hyperplan qui a la plus grande distance au point d'apprentissage le plus proche de chaque classe (dite marge fonctionnelle), car en général, plus la marge est grande, plus l'erreur de généralisation du classifieur est faible [40].

Pour un problème de deux dimension, on veut trouver l'hyperplan à marge maximale

qui divise le groupe des point x_i pour les quels $y_i = 1$ du groupe de points pour les quels $y_i = -1$, ce qui est défini pour que la distance entre l'hyperplan et le plus proche point x_i des deux groupes est maximisé.

Chaque hyperplan peut être écrit comme l'ensemble de points satisfont :

$$\vec{w} \cdot \vec{x} - b = 0,$$

tel que \vec{w} le vecteur normal de l'hyperplan. Le paramètre $\frac{b}{\|\vec{w}\|}$ détermine le décalage de l'hyperplan de l'origine par rapport au vecteur normal \vec{w} .

L'analyse discriminante linéaire (LDA)

L'analyse discriminante linéaire est un algorithme de classification linéaire qui assigne une nouvelle observation à une classe en utilisant la séparation des classes avec les hyperplans (voir figure 4.14). Pour un problème de deux classes, L'algorithme LDA assume que les densités de probabilités $p(x|y = 0)$ et $p(x|y = 1)$ sont de distribution normal avec moyennes et même matrice de covariance (μ_0, Σ) et (μ_1, Σ) , tels que x représente l'ensemble d'entraînement (Training set) et y est une classe connue [94]. Sous ces conditions, La solution optimale de Bayes est de prédire les points appartenant à la seconde classe si le log des rapports de vraisemblance est supérieur à un seuil T , de sorte que :

$$(x - \mu_0)^T \Sigma^{-1} (x - \mu_0) - (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) > T$$

ou plus précisément,

$$\vec{w} \cdot \vec{x} > c$$

pour une certaine constante c et :

$$\vec{w} = \Sigma^{-1} (\mu_1 - \mu_0)$$

$$c = \frac{1}{2} (T - \mu_0^T \Sigma^{-1} \mu_0 + \mu_1^T \Sigma^{-1} \mu_1)$$

Alors le critère de décision est purement une fonction linéaire des termes connus.

Pour résoudre un problème de classe N , ($N > 2$), plusieurs hyperplans sont utilisés. La stratégie généralement utilisée pour une problème multiclassés est la stratégie "un contre le reste" (OVR one over all) qui consiste à séparer chaque classe de toutes les autres. Cette technique nécessite très peu de calculs et simple à utiliser [57]. En termes géométriques : le critère d'une entrée \vec{x} faisant partie à une classe y est purement une fonction de projection du point d'espace multidimensionnel \vec{x} sur le vecteur \vec{w} (ainsi, on ne considère que sa direction). Alors, l'observation appartient à y si le \vec{x} correspondant est situé sur un certain côté d'un hyperplan perpendiculaire à \vec{w} .

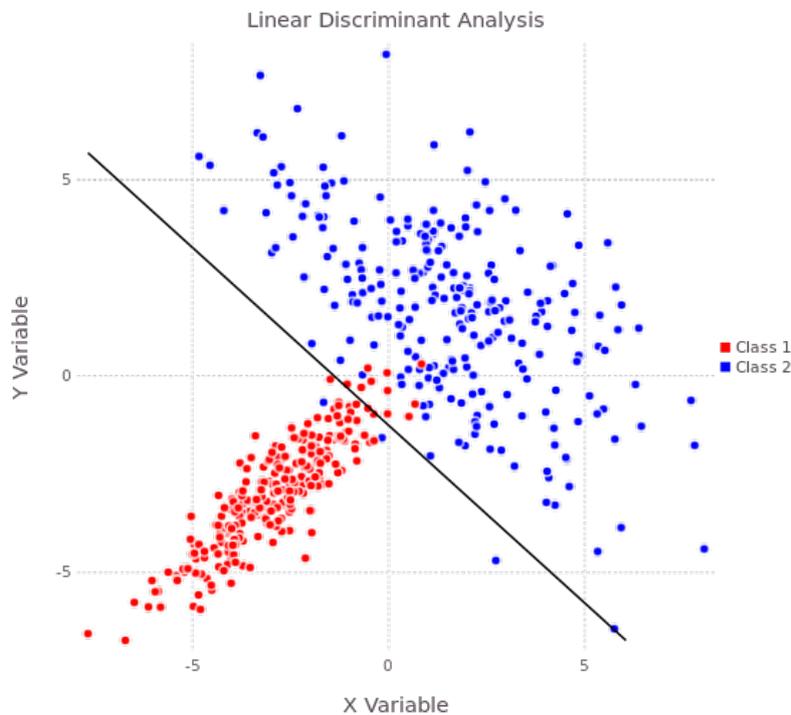


FIGURE 4.14: L'algorithme de l'analyse discriminante linéaire (LDA)

La forêt aléatoire (RF)

L'approche de la forêt aléatoire [16] qui peut être considérée comme une forme de l'algorithme de k plus proche voisins, est un ensemble de classifieurs d'arbres de décision non corrélés. Un classifieur d'arbre de décision est construit à partir d'un ensemble de règles binaires dans une structure hiérarchique contenant des nœuds et des bords dirigés

(voir figure 4.15). Les arbres de décision sont connus pour avoir une variance élevée, car ils ont une tendance à surajuster les données d'entraînement. En utilisant des méthodes d'agrégation bootstrap, la forêt aléatoire peut générer des modèles qui peuvent réduire la variance des arbres, ce qui peut fournir de meilleures performances.

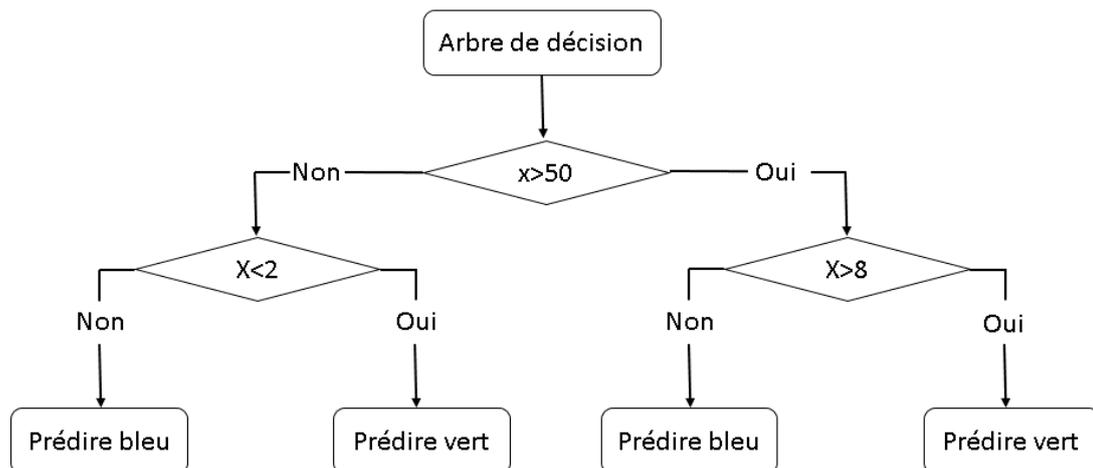


FIGURE 4.15: Exemple d'une arbre de décision simple

Finalement, les algorithmes d'apprentissage statistique sont évalués à l'aide de la fonction score de précision (AS), sur les données de chaque sujet.

4.7 Expérimentations et étude comparative

Cette étude a deux objectifs généraux ; (a) L'évaluation de la qualité de l'information des approches de périodogrammes quantiles dans les systèmes BCI, et (b) l'évaluation comparative d'un ensemble de classifieurs d'apprentissage automatique en utilisant les quantiles périodogrammes introduits. Nous nous référons au K-plus proches voisins par Knn, Support vector machine par SVM, analyse discriminante linéaire par LDA et forêt aléatoire par RF.

Dans cette étude, les caractéristiques sont extraites des signaux EEG à travers les estimateurs spectraux et injectées dans les classifieurs d'apprentissage automatique

pour décider si elles correspondent à des tâches d'imagerie motrice de la main gauche-droite ou de repos.

Deux expérimentations ont été réalisées pour comparer les performances des algorithmes de classification. Dans la première expérimentation, l'objectif est la mesure de la séparabilité de chaque approche d'estimateurs spectraux, en utilisant un schéma de cross validation (CV) interne-externe (CV de 10 fois avec plusieurs exécutions) avec tous les classifieurs à l'étude. Le taux de précision (AS) a été obtenu par la moyenne sur les exécutions plus/moins l'écart-type.

L'objectif de la seconde expérimentation est de mesurer la performance dite de session à session, puisque les systèmes basés sur BCI sont connus pour avoir des variations considérables de la dynamique EEG à un temps éloigné (entre les sessions) [39]. Ces variations dégradent la performance d'intersession de BCI [87]. Dans notre cas, la dernière session est utilisée comme une session de test et les restants sont utilisés comme sessions d'entraînement du modèle.

La table 4.1 résume les résultats du schéma de CV de la première expérimentation. Elle représente la moyenne des résultats de performance de classification associés à différentes techniques spectrales : FFT, périodogramme de Welch, quantile périodogramme, et quantile périodogramme du lasso. Les paramètres estimés sont : Le premier sujet, le quantile $\hat{\tau}$ estimé est égale à 0.5, et le paramètre de régularisation $\hat{\alpha} = 1$. Pour le deuxième sujet, $\hat{\tau} = 0.95$ et $\hat{\alpha} = 1$. Pour le troisième sujet, $\hat{\tau} = 0.25$ et $\hat{\alpha} = 2$. Le quatrième sujet $\hat{\tau} = 0.5$ et $\hat{\alpha} = 2$, et finalement pour le cinquième sujet $\hat{\tau} = 0.5$ et $\hat{\alpha} = 2$. On a mis les meilleurs résultats dans chaque ligne en gras.

Comme on le voit dans les résultats, les nouvelles approches surpassent les techniques classiques avec tous les sujets et presque tous les algorithmes d'apprentissage automatique. Il est clair à partir des résultats obtenus que l'extraction de caractéristiques basée sur le quantile périodogramme du lasso associée au classifieur de forêt aléatoire présente les meilleures performances. Le quantile périodogramme du lasso montre une amélioration significative avec tous les classifieurs, ce qui pourrait s'expli-

Sujet	N. essai	Classes	FFT	pér. de Welsh	Q Pér.	LQ pér.		
1	80	M D vs. Rep.	knn	67.5 +/- 0.5%	knn	73.25 +/- 1%	knn	75 +/- 1%
			DLA	47.5 +/- 1%	DLA	62.5 +/- 3%	DLA	62.5%
			SVM	42.5 +/- 4%	SVM	60.75 +/- 2%	SVM	61.75 +/- 2%
			RF	70 +/- 1%	RF	78 +/- 2%	RF	84 +/- 1%
2	100	M D vs. M G	knn	64 +/- 2%	knn	65 +/- 2.5%	knn	65 +/- 0.25%
			DLA	47 +/- 1.5%	DLA	51 +/- 1%	DLA	52 +/- 1%
			SVM	47 +/- 3%	SVM	51 +/- 1.4%	SVM	56 +/- 1%
			RF	69 +/- 1%	RF	74 +/- 0.5%	RF	77 +/- 2.5%
3	100	M D vs. M G	knn	66.5 +/- 2%	knn	70 +/- 2%	knn	72 +/- 1%
			DLA	47 +/- 3%	DLA	54 +/- 2%	DLA	58 +/- 2%
			SVM	50 +/- 2%	SVM	58 +/- 2%	SVM	59 +/- 1%
			RF	75 +/- 1%	RF	76.75 +/- 1%	RF	77 +/- 2%
4	100	M D vs. M G	knn	67.5 +/- 2%	knn	70 +/- 1.5%	knn	71 +/- 1%
			DLA	55 +/- 1%	DLA	55 +/- 2%	DLA	57 +/- 2%
			SVM	56 +/- 0.5%	SVM	54 +/- 2%	SVM	55 +/- 2%
			RF	70 +/- 1%	RF	73 +/- 1%	RF	74.5 +/- 1%
5	100	M D vs. M G	knn	68.5%	knn	68 +/- 1%	knn	68.5 +/- 0.5%
			DLA	47 +/- 3%	DLA	52 +/- 2%	DLA	59 +/- 2%
			SVM	43 +/- 2%	SVM	56 +/- 2%	SVM	59 +/- 2%
			RF	78 +/- 0.5%	RF	78.5 +/- 1%	RF	78 +/- 1%

TABLE 4.1: Résultats de 4 estimateurs spectraux différents, avec 4 classifieurs

Quantile	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$	$\tau=0.95$				
Classifieurs	km	60+/-0%	km	73.25 +/-1%	km	65+/-0%	km	50+/-0%
	DLA	45+/-0%	DLA	62.5 +/-3%	DLA	50+/-0%	DLA	50+/-0%
	SVM	50+/-0%	SVM	60.75 +/-2%	SVM	55+/-0%	SVM	50+/-0%
	RF	70+/-3%	RF	78+/-2%	RF	60+/-2%	RF	65+/-5%

TABLE 4.2: Quantile périodogramme avec différents quantiles en utilisant les données du premier sujet

quer par le fait que l'approche est robuste contre la distorsion non linéaire, les valeurs aberrantes et les signaux bruyants [64].

Afin de voir l'impact des différents quantiles et de changement du paramètre de régularisation sur les résultats des approches basées sur les quantiles, on a effectué une expérimentation de grille avec les données du premier sujet. La table 4.2 montre les résultats des algorithmes de classification en utilisant le quantile périodogramme comme méthode d'extraction de caractéristiques, avec des quantiles différents. On peut voir à travers ces résultats que le choix du quantile est très important. Ainsi, des techniques d'optimisation pourraient être proposées pour surmonter ce problème. Dans ce cas, le quantile minimum est un minimum global pour tous les classifieurs. Avec cela étant dit, on note que les autres quantiles pourraient également avoir des informations utiles.

La table 4.3 résume les résultats des classifieurs en utilisant le quantile périodogramme du LASSO comme méthode d'extraction de caractéristiques, avec différentes valeurs de paramètres de régularisation dans l'intervalle $[10^{-2}, 10]$. Comme le montrent les résultats, le choix du paramètre α compte aussi. Des techniques d'optimisation pourraient alors être proposées pour le paramètre de régularisation. On note dans ce cas que le minimum α n'est pas global pour tous les classifieurs. Ainsi, des techniques de minimisation pourraient être appliquées à chaque classifieur seul.

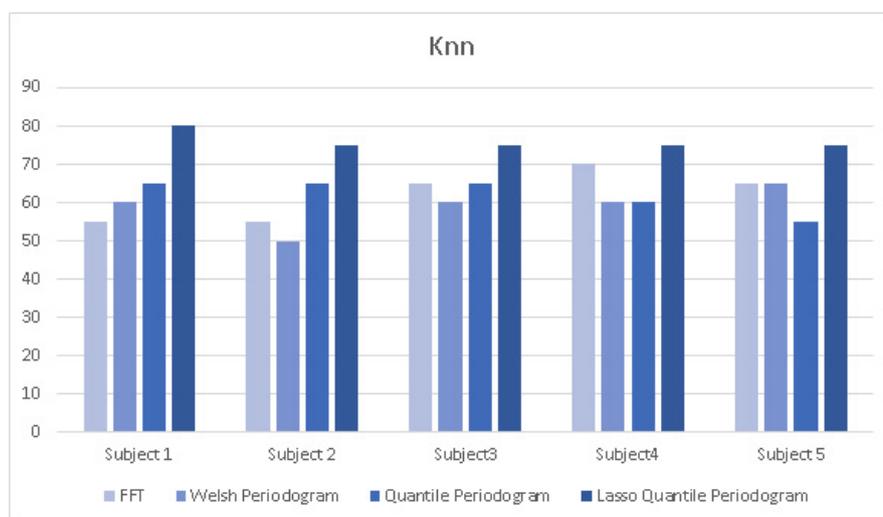


FIGURE 4.16: Taux de précision pour l'algorithme de Knn avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.

Les figures 4.16, 4.17, 4.18 et 4.19 décrivent les résultats des performances de session

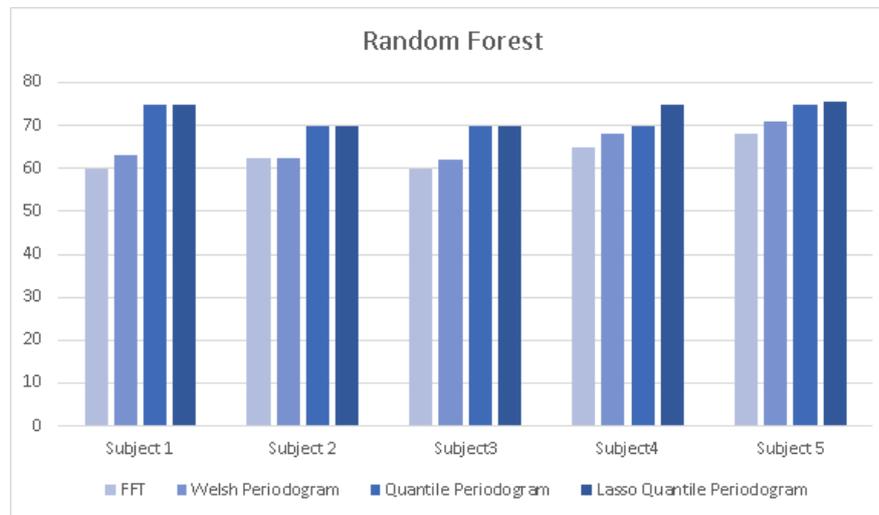


FIGURE 4.17: Taux de précision pour l'algorithme de Forêt aléatoire avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.

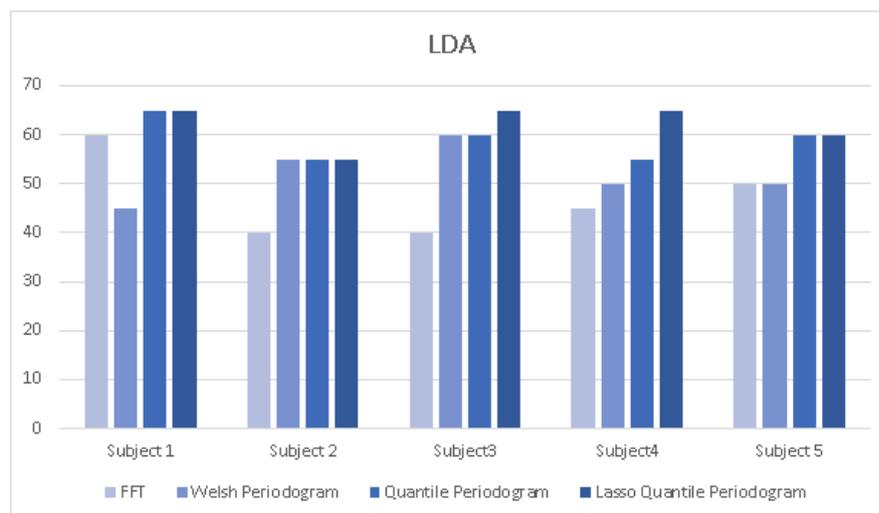


FIGURE 4.18: Taux de précision pour l'algorithme d'analyse discriminante linéaire avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.

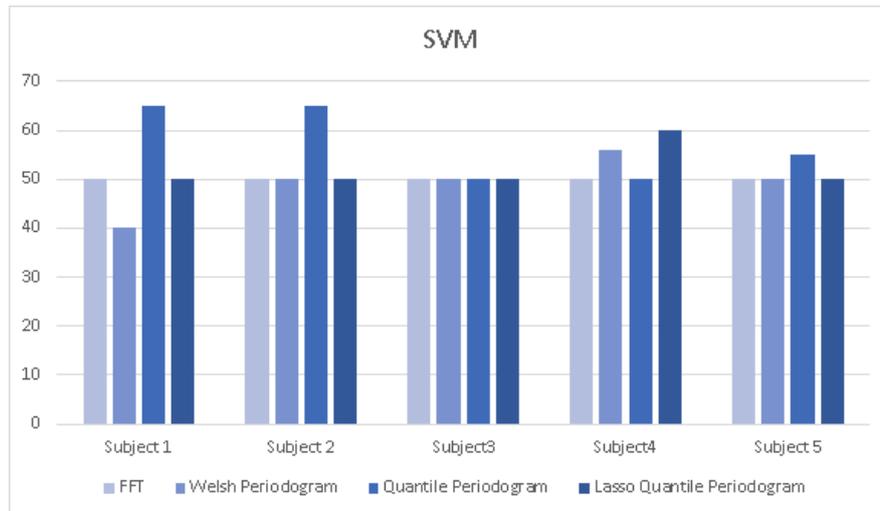


FIGURE 4.19: Taux de précision pour l’algorithme de Machine à vecteurs de support avec différents estimateurs spectraux pour tous les sujets. Performance de session à session.

à session de Knn, RF, LDA et SVM respectivement, avec tous les sujets et avec les estimateurs de spectre à l’étude. À cette fin, nous devons mentionner que les capacités des classifieurs sont également importantes dans ce cas. Il est clair à partir des résultats obtenus que le Knn, et les méthodes RF montrent de meilleurs résultats. En outre, les approches basées sur les quantiles et en particulier le quantile périodogramme du lasso permet d’améliorer les performances pour tous les sujets.

Les résultats des classifieurs pourraient s’expliquer par les exigences restrictives du LDA et du SVM, qui sont considérés comme des classifieurs linéaires. Comme mentionné dans la section 4, le classifieur LDA fonctionne sous les hypothèses de la normalité des fonctions de densités de probabilités conditionnelles. Alors que l’algorithme SVM a besoin d’un espace de caractéristiques linéairement séparable pour être efficace. Le KNN et le RF n’exigent aucune hypothèse spécifique sur la structure des données.

4.8 Conclusion

Dans cette application, on a proposé deux nouveaux estimateurs du spectre en utilisant la régression quantile et la régularisation de la norme L_1 . Les nouvelles approches sont utilisées comme des méthodes d’extraction de caractéristiques à partir des si-

gnaux EEG. On a montré que les nouvelles approches surpassent les techniques classiques telles que la FFT et le périodogramme de Welsh. Ce qui peut suggérer que le quantile périodogramme et le quantile périodogramme du LASSO peuvent mieux caractériser l'information contenu dans les signaux non-stationnaires, non-gaussiennes et non-linéaires. Des recherches dans le futures sont nécessaires pour améliorer le temps de calcul et le choix de paramètres.

Chapitre 5

Lasso quantile périodogramme pour la reconnaissance d'activités humaines

“Errors using inadequate data are much less than those using no data at all.”

– Charles Babbage

L'application de reconnaissance d'activité a été conçue pour permettre l'analyse automatique des données brutes d'accélération 3D afin de détecter des patterns spécifiques d'activités physiques humaines, qui peuvent être collectés lors de séances d'auto-exercice de rééducation. Ce travail a été effectué dans le cadre de projet Européen MEDOLUTION (Medical care evolution), afin de concevoir et de développer des applications intelligentes utiles pour l'évaluation des systèmes de soins de santé. En particulier, montrer la valeur ajoutée des données capturées par des capteurs, des appareils d'imagerie et d'autres dispositifs médicaux pour les décisions de diagnostic et de traitement. Le module proposé a été conçu pour permettre l'analyse automatique des capteurs ambiants afin de détecter des patterns spécifiques dans les activités humaines lors des exercices de rééducation. En outre, la surveillance des signes vitaux et des symptômes des patients signalés par des capteurs via la plate-forme Hapicare et son

application mobile.

5.1 Introduction

Dans la dernière décennie, une croissance constante a été constatée dans la population âgée [93]. Avec la préférence de rester dans leurs maisons [10], des problèmes comme les soins de santé ont apparu, puisque la santé se détériore avec le vieillissement. Aujourd'hui, assister les personnes âgées, construire des systèmes intelligents qui peuvent les aider à gérer leur vie quotidienne et accroître leur sécurité sont considérés comme des défis majeurs dans la recherche scientifique. Ainsi, des techniques basées sur les technologies des capteurs portables ont été proposées (les smartwatches, les smartphones et les capteurs portables) [108]. En outre, ces systèmes peuvent également être utilisés pour évaluer l'efficacité de différentes interventions visant à augmenter les niveaux d'activité physique et de motiver les individus à devenir plus actifs physiquement.

Bien que cette technologie offre la plate-forme idéale pour surveiller les patterns d'activité quotidienne, des algorithmes efficaces sont également nécessaires pour interpréter les données accélérométriques. Plusieurs approches de machine learning et d'intelligence artificielle ont été proposées pour identifier les différentes activités à partir des données de capteurs fixés sur le corps [96, 77]. Ces techniques fonctionnent généralement via deux étapes principales. L'extraction des caractéristiques à partir des données accélérométriques. Après, un classifieur est utilisé pour identifier l'activité correspondante à chaque fenêtre de temps. Une gamme de différentes approches ont été proposées dans la littérature pour obtenir les caractéristiques à partir des données accélérométriques [76]. Certains d'entre eux ont préféré de travailler avec les caractéristiques temporelles [60, 24, 4], et d'autres à travers l'analyse fréquentielle [9, 70, 79]. Plus récemment, l'analyse des ondelettes a été utilisée pour obtenir ce qu'on appelle les caractéristiques du domaine temps-fréquence [96].

La nature des signaux joue un rôle important pour déterminer quelle méthode peut on utiliser pour extraire des caractéristiques utiles. Comme la propriété de non-stationnarité

apparaît dans ce type de données, des méthodes classiques ne peuvent pas caractériser l'information contenue dans les signaux [76]. Dans ce travail, on propose deux nouvelles approches pour surmonter ce problème. Ces approches sont basées sur la régression quantile et la régularisation du LASSO. La performance des nouvelles approches est comparée aux estimateurs classiques telle que la FFT.

5.2 Contexte de l'assistance de la vie quotidienne

La disponibilité des systèmes automatiques capables de classifier des activités physiques quotidiennes effectués par des sujets humains est très attirante pour plusieurs applications du domaine de suivi de la santé et dans le développement des interfaces machine-humain. Ces activités physiques peuvent être des postures statiques comme s'asseoir, debout, allongé ou bien des mouvements dynamiques comme le cas étudié dans cette partie : les exercices de rééducation physique. Dans les dernières années, les chercheurs

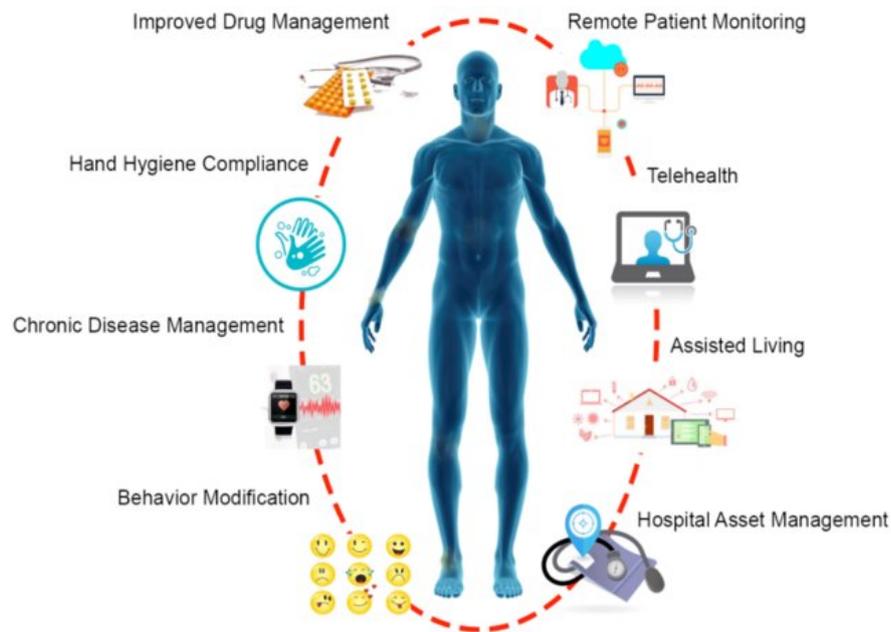


FIGURE 5.1: Illustration graphique pour un système d'assistance de la vie quotidienne [83]

ont développé diverses technologies d'assistance basées sur l'intelligence ambiante. L'intelligence ambiante est un nouveau paradigme des technologies de l'information visant à renforcer les capacités des individus en développant des environnements numériques

sensibles, adaptatifs et qui répondent aux besoins humains [85, 1]. Les technologies d'assistance de vie peuvent être utilisées pour prévenir, guérir et améliorer le bien-être et l'état de santé des personnes âgées. Ces technologies peuvent notamment être utiles dans les cas de rappels des médicaments [80], systèmes de détection de chute [2] et des systèmes de surveillances par vidéos [27] (voir figure (5.1) pour une illustration graphique).

Récemment, l'avancement dans plusieurs domaines technologiques ont aidé la vision d'assistance de la vie quotidienne pour devenir une réalité. Ces technologies incluent les maisons intelligentes, les robots d'assistance et les téléphones et les capteurs portables. Dans cette étude, un système d'assistance pour surveiller des programmes de rééducation physique en utilisant des capteurs portables est considéré.

5.3 Capteurs portables et acquisition de données

L'une des choses les plus importante à considérer pour construire un système automatique de la reconnaissance d'activités humaines est de choisir les capteurs. Ces capteurs doivent être petits et avec un poids léger afin d'être attachés aux corps humain, sans nuire au confort de l'utilisateur, et pour lui permettre d'exécuter les exercices sous des conditions libres autant que possible.

Les capteurs inertiels ont été utilisés principalement pour la navigation aérienne, pour les bateaux, les véhicules et pour les robots. Avec le développement des technologies micro-mécaniques, des capteurs inertiels de taille petite, de poids léger et à bas prix ont été construit [8]. Le nombre et l'emplacement des capteurs sur le corps humain ont un grand impact sur la stabilité et l'efficacité des algorithmes de reconnaissance d'activités [61].

Plusieurs études ont été faites dans la littérature pour déterminer les emplacements pertinents des capteurs aussi bien que le nombre minimum tout en assurant un taux maximal de reconnaissance. Ces capteurs peuvent être placés sur différentes parties du

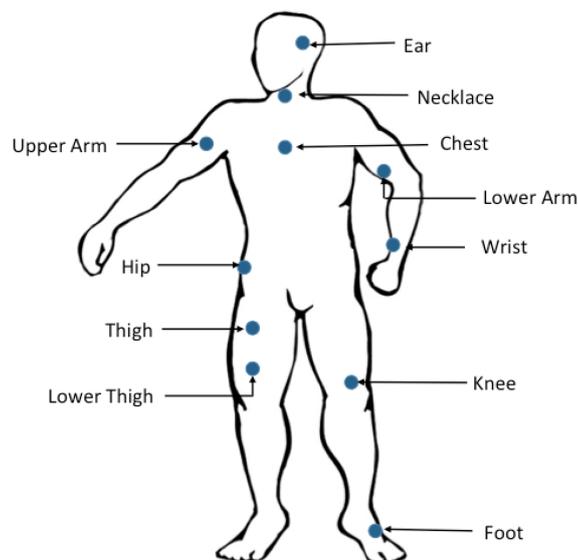


FIGURE 5.2: Illustration graphique pour l'emplacement de capteurs [8]

corps humain où les mouvements sont étudiés comme on peut le voir sur la figure 5.2. Les emplacements de capteurs les plus souvent utilisés sont le sternum [66], le bas du dos [14] et la taille [41]. Les meilleurs résultats ont été obtenus pour une configuration avec trois capteurs situés sur la poitrine, la cuisse et la cheville. Ces résultats ont démontré que la reconnaissance de l'activité humaine pourrait être améliorée de manière significative en combinant des accéléromètres situés à la fois dans les parties supérieure et inférieure du corps [33].

5.4 Description du jeu de données

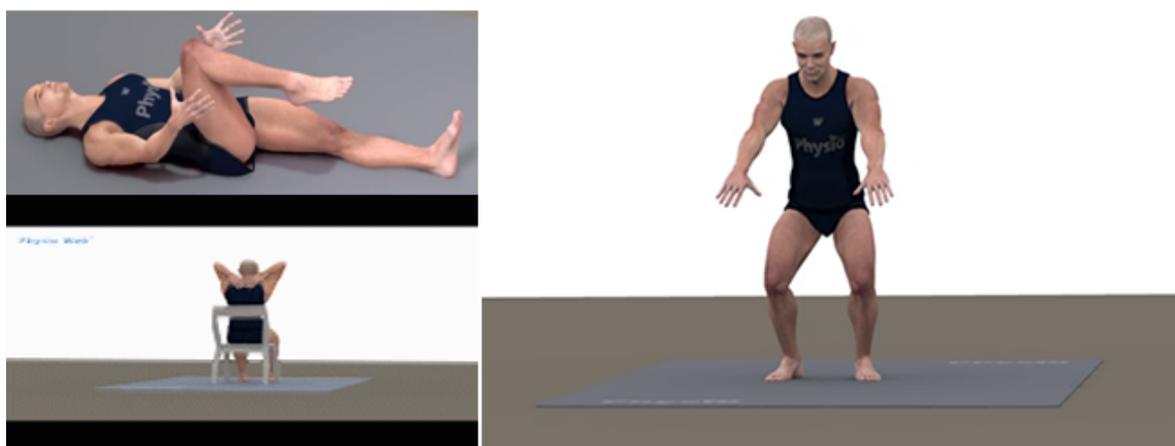


FIGURE 5.3: Positions des exercices pour le protocole de rééducation

L'ensemble de données utilisé dans cette étude a été collecté dans le cadre d'une étude sur les auto-exercices de rééducation physique pour les patients atteints de la maladie de Parkinson. Le but est de construire un système intelligent capable de reconnaître et de mesurer la qualité des exercices.

Un ensemble de 4 traqueurs de mouvements (2 téléphones et 2 smartphones) ont été déployé comme des capteurs portables pour enregistrer les mouvements des sujets. Ces capteurs collectent les données avec une fréquence d'échantillonnage de 60Hz. Ils sont placés comme montré dans la figure 5.4. Cette configuration a été motivé par le travail dans [84].



FIGURE 5.4: Emplacement des appareils

Les séances d'auto-exercice sont composées de séquences de mouvements répétitifs séparés par des pauses de repos. Dans l'expérimentation, Les données ont été collectées auprès de 13 sujets volontaires sains. Basé sur le scénario de jeu de données, chaque sujet avait la tâche de faire 9 exercices répétés dans trois positions différentes (debout, assis et allongé). Chaque exercice a dû être répété 5 fois pour être sûr d'obtenir un

échantillon représentatif et cohérent. Finalement, la taille de ce jeu de données est de $13 \times 9 \times 5$.

5.5 Analyse des signaux accélérométriques

Les signaux accélérométriques dans ce jeu de données sont de type à onde carré (regardez la figure (5.5)). Ce type de signaux est connu dans le domaine de traitement de signal comme train d'impulsions (pulse train) ou l'onde d'impulsion (pulse wave). Les ondes carrés sont des signaux constitués de transitions instantanées entre deux niveaux, ce qui crée une discontinuité dans le signal. Pour passer les points de discontinuité l'approximation de séries de Fourier créent des ondulations sur les deux cotés des sauts, ce phénomène est appelé le phénomène de Gibbs [72]. La transformée de Fourier pour une simple onde carrée est la fonction Sinc ($\sin(x)/x$). Une représentation de l'approximation et la transformée de Fourier pour un signal à ondes carrés est donnée dans la figure 5.6.

Les questions qui se posent, que peuvent le quantile périodogramme et le quantile périodogramme du LASSO apporter comme information sur ce type de signaux ? C'est quoi la différence entre les estimateurs spectraux classiques et les estimateurs quantiles ?

Pour répondre à ces questions, différentes expériences sont présentés. Figure 5.7 représente la médiane et l'intervalle de confiance des différents estimateurs spectraux sur toutes les répétitions de l'exercice 1, l'axe x du capteur de la main droite est utilisé. On peut remarquer dans un premier temps, que tous les estimateurs présentent une forme générale similaire. Ce qui est à noter, est que les estimateurs quantiles montrent une croissance de puissance entre la fréquence 8Hz et 15Hz. Cette information n'apparaît pas dans les estimateurs classiques.

La figure 5.8 illustre la médiane du quantile périodogrammes du LASSO effectué sur les données de l'exercice 1 et 4. l'axe x du capteur de la main droite a été utilisé. Différents valeurs du quantile d'intérêt ont été utilisées ($\tau = [0.25, 0.5, 0.75, 0.95]$). Il

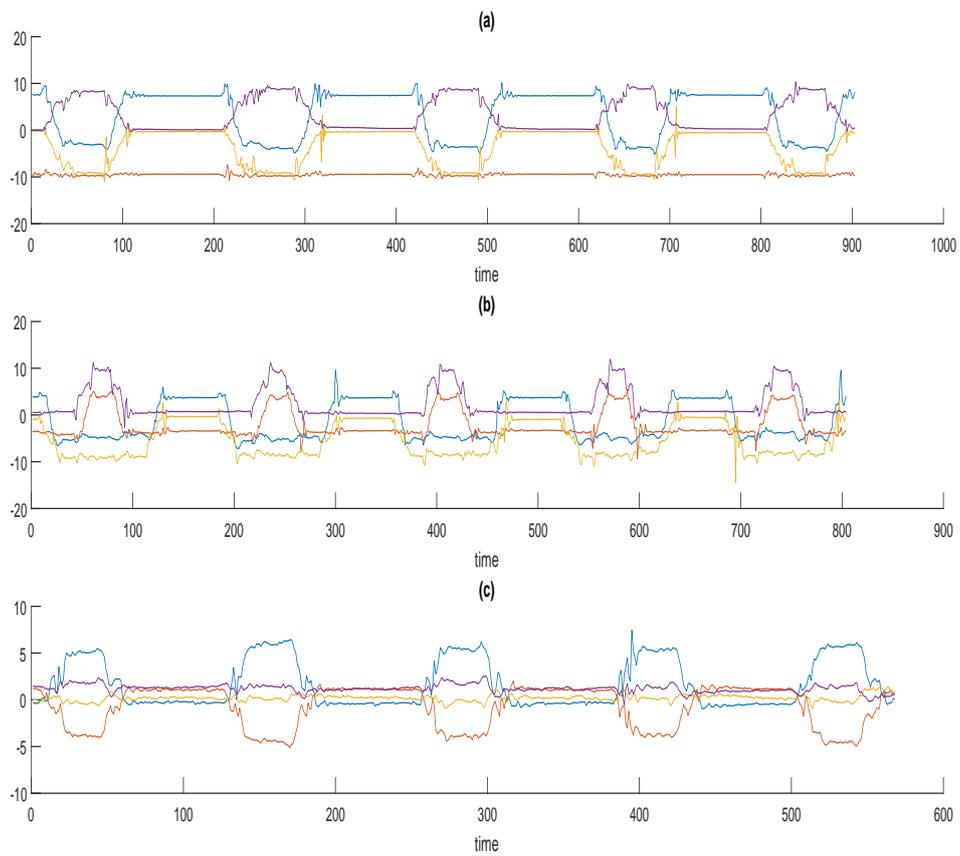


FIGURE 5.5: 5 répétition pour les activités 1,2 et 7. Les capteurs d'accélération ont été utilisés en choisissant les données de l'axe x. (a) représente les données de l'exercice 1, (b) les données de l'exercices 2 et (c) les données de l'exercice 7

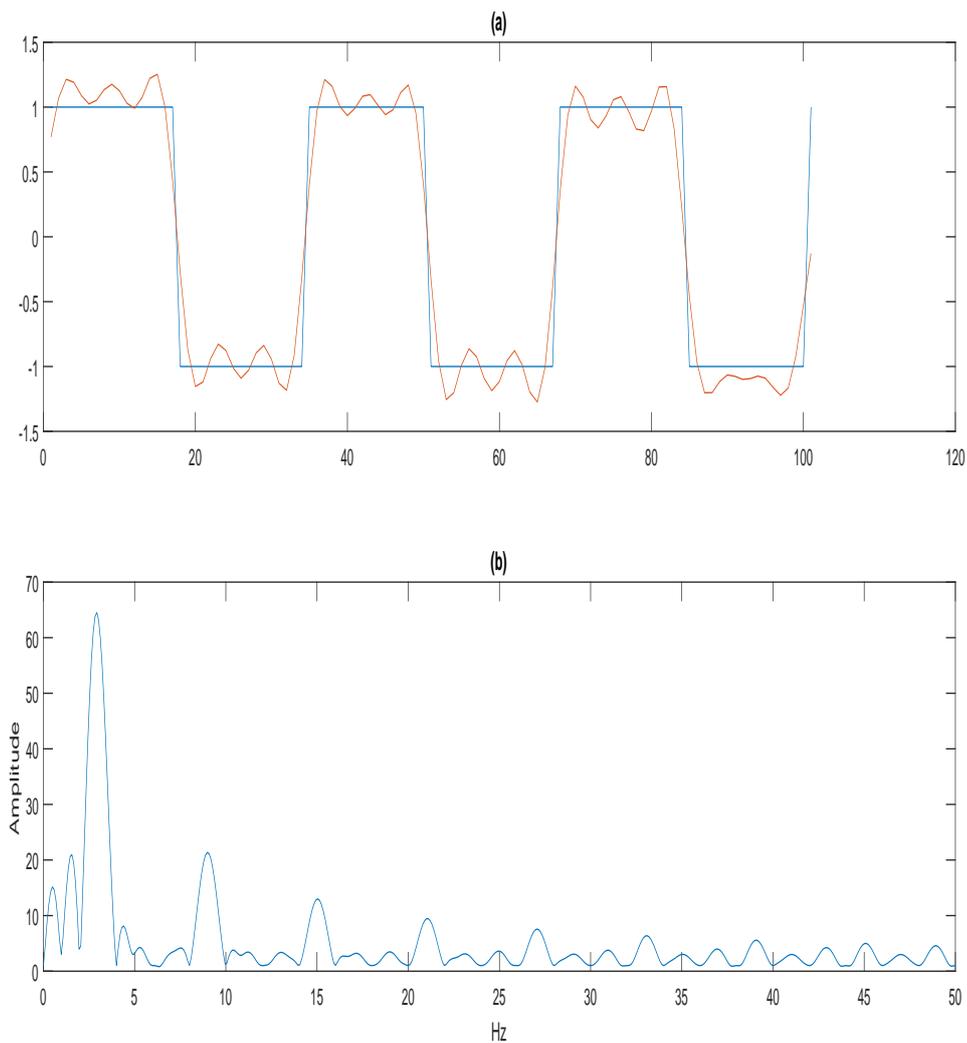


FIGURE 5.6: Le phénomène de Gibbs. (a) représente l'approximation de Fourier, (b) est la transformée de Fourier du signal à ondes carrés.

est à noter qu'il y a une différence de comportement de l'estimateur en changeant le quantile. Cela renforce l'idée selon laquelle les informations les plus discriminantes sont associées à des parties spécifiques de la distribution. La différence est observée comme un changement d'amplitude sur des bandes de fréquences spécifiques. Par exemple dans la représentation de l'exercice 4, il y a une croissance dans la bande de fréquence 5Hz-10Hz pour les quantiles (0.5, 0.75 et 0.95) alors qu'il est à plat pour le quantile 0.25. Ainsi, un choix approprié du quantile d'intérêt est nécessaire.

Figure 5.9 représente la médiane des différents estimateurs spectraux considérés appliqués sur les données des exercices 1, 4 et 7. L'axe x du capteur de la main droite est utilisé pour toutes les expériences. Il est à noter que dans les estimateurs du quantile, la différence entre les exercices est claire. Pour les estimateurs classiques, les représentations sont presque surimposées. Cela suggère que les estimateurs du quantile peuvent capturer une information supplémentaire que les estimateurs classiques.

5.6 Processus d'apprentissage automatique

Il y a plusieurs façons de faire de la reconnaissance d'activité dans la littérature. Cependant, les étapes les plus courantes sont : la visualisation, le pré-traitement, l'extraction de caractéristiques, la réduction de dimensionnalité et la classification. La figure (5.10) résume le processus d'apprentissage.

La visualisation et l'analyse

L'étape de visualisation est très importante d'un point de vue statistique. Des outils statistiques pourraient être utilisés pour communiquer des données et des informations, pour aussi les résumer et éviter toute distorsion des informations utilisables. La visualisation et l'analyse du jeu de données ont été discutés dans la partie analyse des signaux accélérométriques.

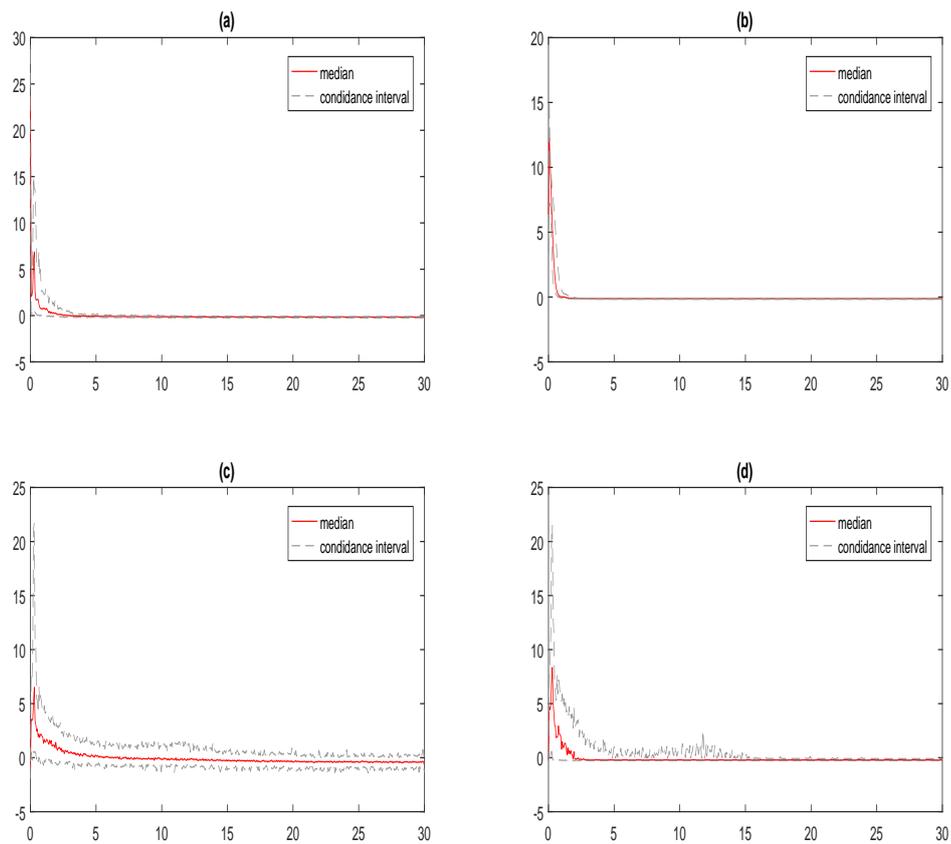


FIGURE 5.7: La médiane et l'intervalle de confiance des différents estimateurs spectraux sur toutes les répétitions de l'exercice 1, l'axe x du capteur de la main droite est utilisé. (a) représente les réalisations du FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO

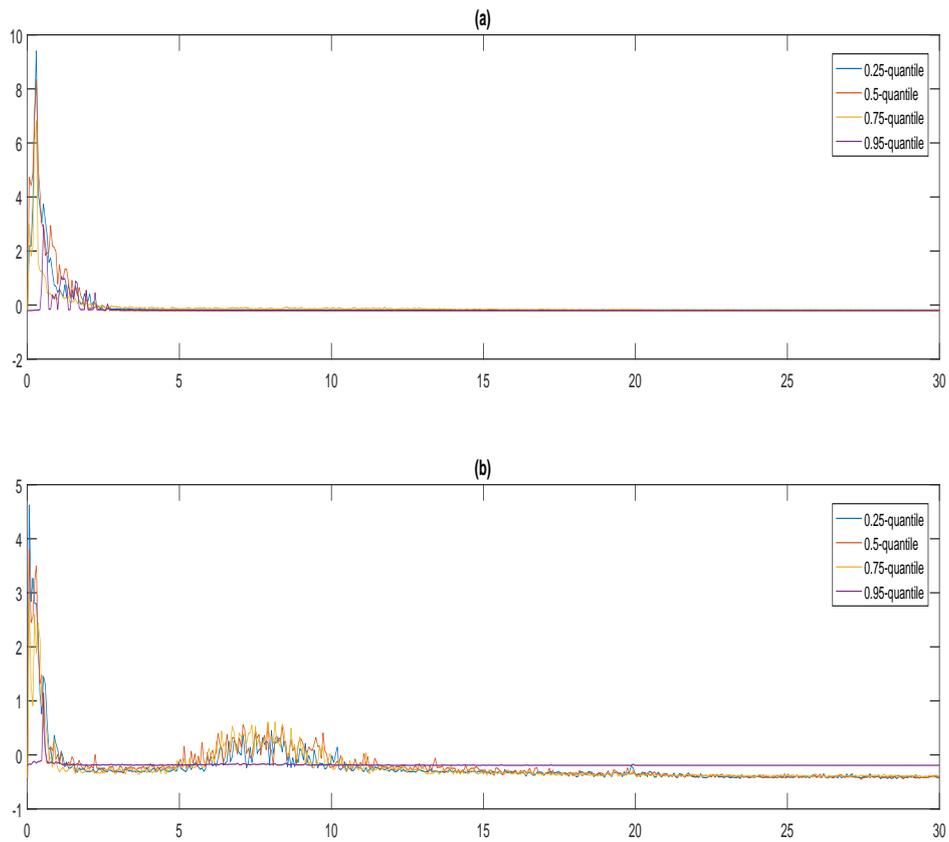


FIGURE 5.8: La médiane du Quantile périodogrammes du LASSO effectué sur les données de l'exercice 1 et 4. l'axe x du capteur de la main droite a été utilisé. (a) représente la réalisation de l'exercice 1 et (b) la réalisation de l'exercice 4.

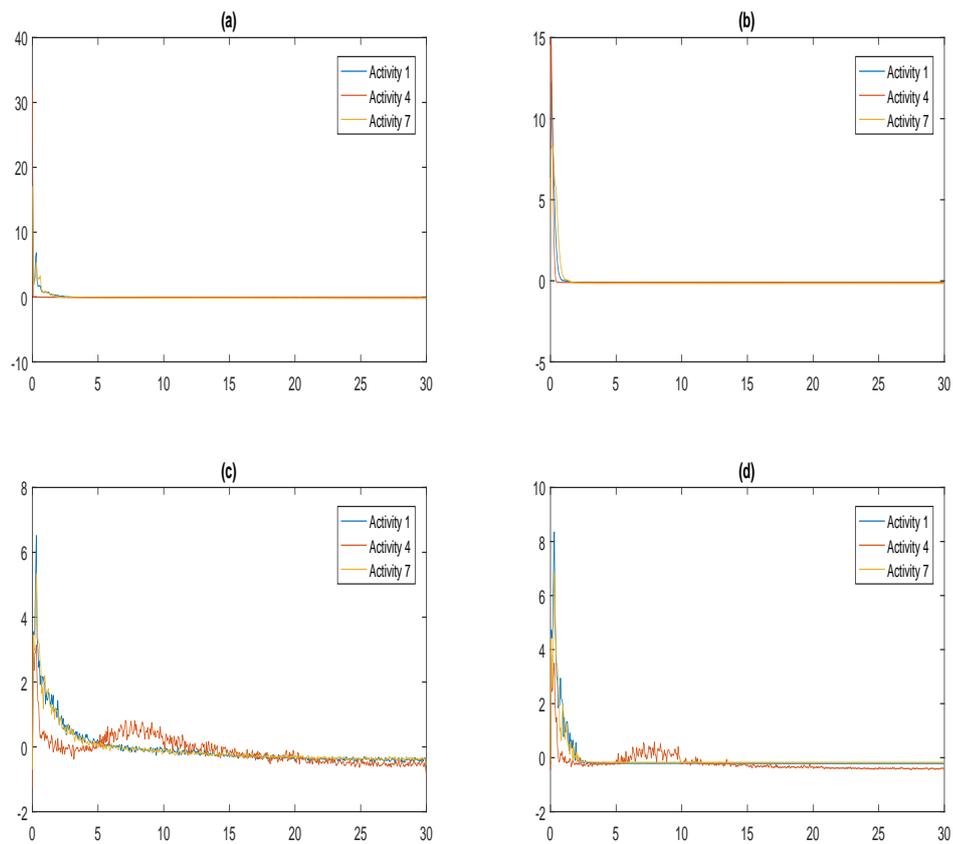


FIGURE 5.9: La médiane des différents estimateurs spectraux considérés appliqués sur les données des exercices 1,4 et 7. L'axe x du capteur de la main droite est utilisé pour toutes les expériences. (a) représente les réalisation du FFT, (b) le périodogramme de Welch, (c) le Quantile périodogramme et (d) le Quantile périodogramme du LASSO.

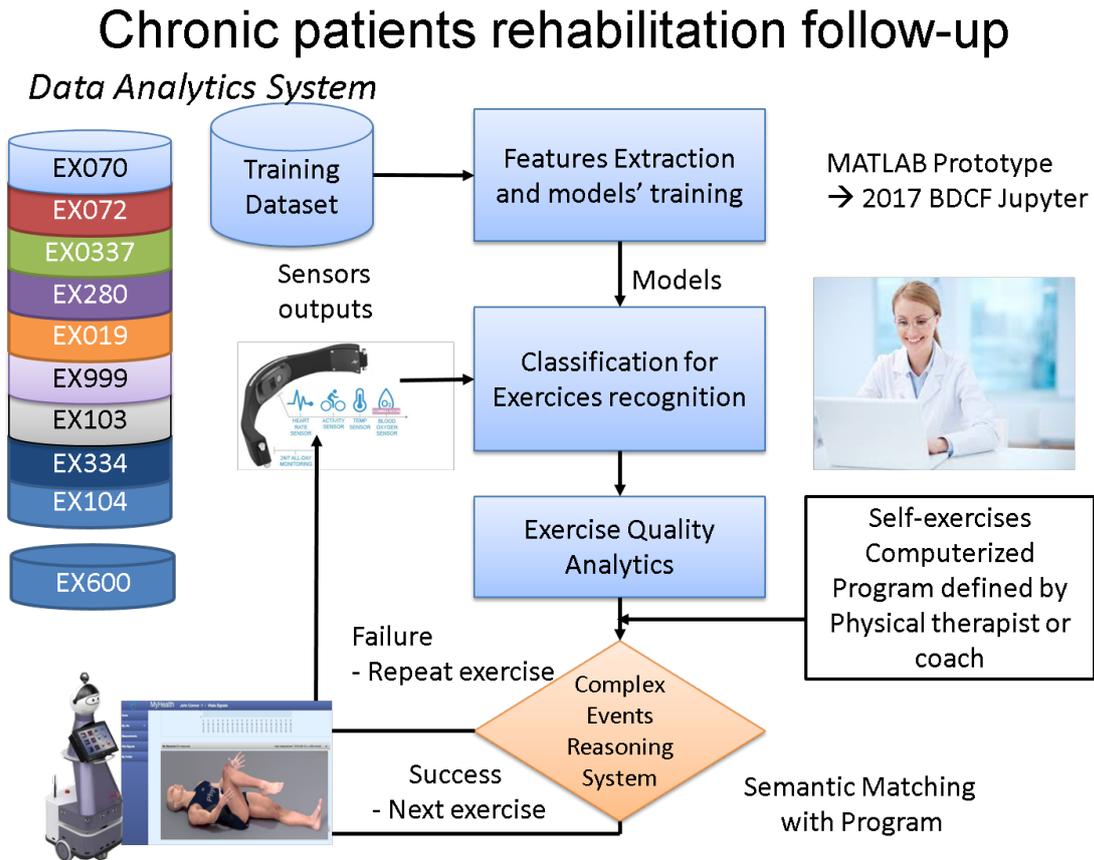


FIGURE 5.10: Processus de reconnaissance d'activités

Le pré-traitement

Dans le cas des capteurs inertiels, les techniques de pré-traitement sont couramment utilisées dans la littérature pour éliminer le bruit et les valeurs aberrantes qui peuvent perturber l'analyse. Par conséquent, des filtres non linéaires, passe-bas, médian et gaussien pourraient être utilisés. Dans cette étude, nous avons utilisé une normalisation des signaux [98].

L'extraction de caractéristiques

Les techniques d'extraction de caractéristiques peuvent être décrites comme l'abstraction des informations utiles qui pourraient être utilisées à la place des données brutes. Ce qui signifie la transformation d'un ensemble de données brutes de grande dimension en un ensemble de caractéristiques à une représentation réduite. Les caractéristiques

sont ensuite utilisées comme des entrées pour les algorithmes de classification.

Dans la littérature, les caractéristiques du domaine temporel [5], du domaine fréquentiel [67] et du domaine temps-fréquence [9] sont utilisées comme des techniques d'extraction de caractéristiques pour les capteurs inertiels. Dans le domaine temporel, les outils statistiques tels que la moyenne, la variance, l'écart-type, l'aplatissement, l'asymétrie, le pic à pic, le facteur et le rang sont largement utilisés. D'autre part, le composant DC dans le spectre FFT, la densité spectrale de puissance et le spectre d'entropie sont utilisés comme caractéristiques du domaine fréquentiel.

Enfin, la somme des coefficients d'ondelettes, la somme au carré des coefficients d'ondelettes et l'énergie des coefficients d'ondelettes sont utilisées comme caractéristiques du domaine temps-fréquence.

Comme on a mentionné dans le chapitre 2 et 3, L'analyse spectrale classique telle que la FFT est connue pour exiger des conditions restrictives. Des données gaussiennes et non-bruitées sont nécessaires, sinon elles ne peuvent pas donner de bonnes performances. Pour surmonter ce problème, on va appliquer le quantile périodogramme et le Lasso quantile périodogramme définis par (3.6) et (3.11) respectivement.

Le vecteur de caractéristiques pour chaque exercice est défini comme suit :

$$F = \{F_1, F_2, \dots, F_n\} \quad (5.1)$$

où, n est le nombre de capteurs et

$$F_k = \{\mu_k, \sigma_k^2, std_k, kur_k, ske_k, p2p_k, rang_k, DC_k, en_k\} \quad (5.2)$$

μ_k représente la moyenne du signal, σ_k^2 est la variance, std_k est l'écart type, kur_k est le coefficient d'aplatissement, ske_k est le coefficient d'asymétrie, $p2p_k$ est le pic à pic, $rang_k$ représente le rang, DC_k est le composant DC et en_k est l'énergie de la représentation spectrale.

La matrice de caractéristiques est alors de taille $n \times N$ où N est le nombre d'exercices.

La réduction de la dimension

La réduction de la dimension est un processus qui réduit le nombre de variables à l'étude [35]. Pour différencier entre les classes, les algorithmes de classification ont besoin d'un espace de caractéristiques représentatif. Des caractéristiques redondantes peuvent détériorer la performance des classifieurs. La réduction de dimensionnalité est une étape très importante lors de l'application des algorithmes de machine learning, comme elle réduit le temps de calcul et la complexité, tout en améliorant le taux de classification. Il existe plusieurs types de réduction de dimension pour plus de détails voir [56]. Dans ce travail, une application de l'analyse en composantes principales (ACP) est proposée.

La classification

Les algorithmes de classification statistique visent à créer un modèle qui prédit si une nouvelle observation appartient à une catégorie ou une autre. Il crée une fonction de décision en utilisant l'espace caractéristique pour produire une prédiction de nouvelles données [40]. Il ya deux directions principales des techniques de machine learning, les approches supervisées et non-supervisées [98, 40]. Les approches supervisées qui exigent des données d'activité entièrement étiquetées (labeled), comme k-plus proches voisins, l'arbre de décision, Naive Bayes et les réseaux de neurones. Les méthodes non-supervisées qui permettent de déduire automatiquement les étiquettes (labels) des données, comme les modèles de mélange Gaussien et les modèles de Markov cachés [81].

Ce travail compare la performance de 4 algorithmes de classification. Les algorithmes de k-plus proches voisins, l'arbre de décision (regarder le chapitre 4 pour plus de détails sur le Knn et l'arbre de décision), Naive Bayes et les réseaux de neurones sont considérés. La mesure du score de précision a été utilisée comme un quantificateur de la performance des résultats prédits.

Une technique de classification supervisée bien connue est le classificateur Naive Bayes, populaire en raison de sa simplicité et de sa facilité de mise en œuvre. Dans cette approche, les caractéristiques d'entrée sont supposées être indépendantes, tandis que la fonction de vraisemblance conditionnelle de chaque activité, peut être exprimée comme le produit de densités de probabilités [8].

Les modèles des réseaux de neurones artificielles sont des algorithmes de machine learning inspiré par les réseaux de neurones biologiques qui constituent le cerveau. Ils peuvent servir comme des méthodes de classification aussi bien que des méthodes de régression [98]. Un modèle de réseaux de neurones artificiel est basé sur un ensemble d'unités ou de nœuds connectés, appelés neurones artificiels. Chaque connexion peut transmettre une sortie d'un neurone artificiel à un autre. Cette sortie est calculé en utilisant une fonction non-linéaire des entrées. Les connexions entre neurones artificiels sont appelées "les poids". Pour plus de détails sur les réseaux de neurones artificiels voir [98]. Dans cette étude, le modèle de Perceptron multicouche (multilayer perceptron MLP) est considéré avec des paramètres par défaut, typiquement, avec une taille de couche cachée égal à 10.

5.7 Résultats et discussion

Le but de cette application est de voir comment les nouvelles approches spectrales peuvent améliorer les résultats de classification avec différents algorithmes.

Dans ce travail, les caractéristiques sont extraites des signaux accélérométriques à travers les estimateurs spectrales FFT, Quantile périodogramme et LASSO quantile périodogramme d'un côté, et de l'autre côté avec les fonctions statistiques temporelles présentées dans l'équation 5.2. Après, un classifieur est utilisé pour déterminer si la fenêtre du temps appartient à une activité ou une autre.

Pour l'évaluation des résultats de classification, un schéma de 10-fold cross validation est utilisé. Le résultat ensuite est obtenu à l'aide de la fonction score.

Algorithme	Knn	Arbre de décision	Naive bayes	Réseau de neurone
FFT – Toutes les caractéristiques	41% (+/-17)	89% (+/- 18)	89% (+/- 17)	22% (+/- 31)
FFT - Réduction de dimension	42% (+/-17)	86% (+/- 21)	92% (+/- 11)	13% (+/- 16)
QuatileP-Toutes les caractéristiques	26% (+/-6)	92% (+/- 14)	47% (+/-14)	12% (+/-7)
Qp - Réduction de dimension	75% (+/- 15)	92% (+/- 13)	95% (+/- 13)	80% (+/-41)
Lqp- Toutes les caractéristiques	39%(+/-10)	91% (+/-11)	87% (+/-18)	14%(+/-6)
LQP – Réduction de dimension	94% (+/- 9)	92% (+/-12)	94% (+/-15)	94% (+/-10)

TABLE 5.1: Les résultats de classification pour les données d'auto-exercices

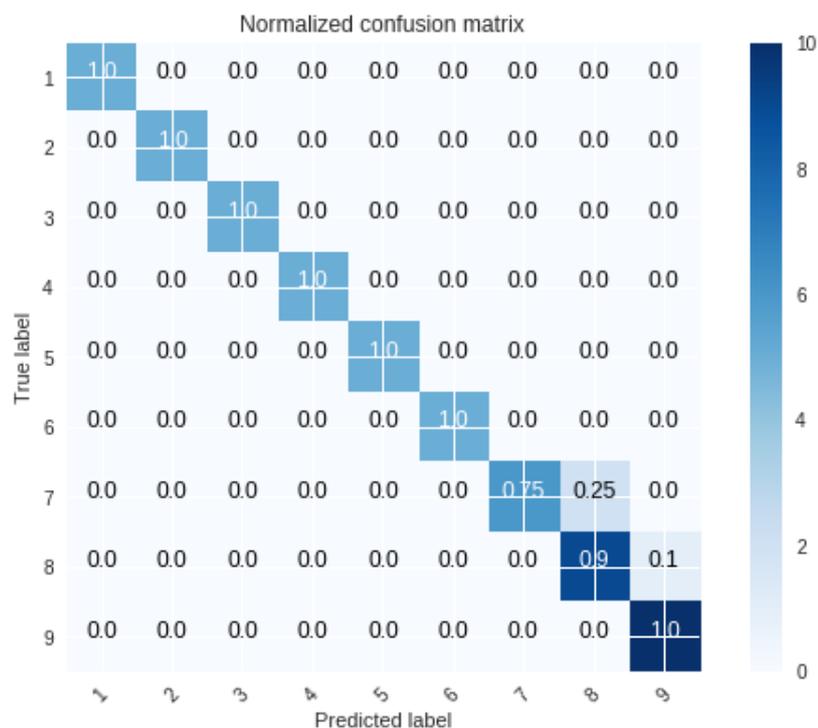


FIGURE 5.11: La matrice de confusion pour le modèle de réseau de neurone en utilisant les caractéristiques du Quantile périodogramme

La table (5.1) résume les taux de classification des 4 classifieurs considérés avec les trois approches spectrales. Deux cas d'espace de caractéristiques ont été considérés. Le cas avec toutes les caractéristiques, et dans le cas de la réduction de dimension en utilisant l'analyse des composantes principales (ACP). La table représente la moyenne des résultats de schéma 10- fold cross validation.

D'après les résultats, il est clair que les deux approches proposées surpassent les techniques classiques avec tous les algorithmes d'apprentissage automatique lorsque la réduction de dimension est appliquée. Pour le Knn, les résultats de classification ont été obtenus comme 42%, 75% et 94% pour la FFT, le quantile périodogramme et le quantile périodogramme du Lasso respectivement avec la réduction de dimension. Pour l'arbre de décision, les résultats suivants ont été obtenus : 86%, 92% et 92% pour la FFT, le quantile périodogramme et le quantile périodogramme du Lasso respectivement tout en appliquant l'ACP. Les résultats par rapport à l'algorithme de réseau de neurones artificiels sont : 13%, 80% et 94%. Comme on peut le voir, les résultats montrent pas seulement la supériorité du quantile et du quantile périodogrammes du lasso par rapport aux méthodes classiques, mais aussi la stabilité de ces nouvelles approches pour tous les algorithmes de classification, et plus particulièrement le quantile périodogramme du Lasso. Cela pourrait s'expliquer par la robustesse du quantile périodogramme du Lasso contre la distorsion non linéaire, les valeurs aberrantes et les signaux bruyants [64].

La matrice de confusion est une matrice qui mesure la qualité de classification pour chaque classe. Chaque ligne correspond à une classe réelle, chaque colonne à une classe estimée. Figure (5.11) représente la matrice de confusion pour le modèle de réseau de neurone en utilisant les caractéristiques du quantile périodogramme.

5.8 Conclusion

Dans cette application, on a proposé deux nouvelles approches du domaine fréquentiel en utilisant la régression quantile et la régularisation du LASSO. Les deux approches

sont utilisées comme des méthodes d'extraction de caractéristiques. On a montré que le quantile périodogramme et le quantile périodogramme du LASSO surpassent la FFT avec l'utilisation des méthodes de réduction de dimensionnalité. Des recherches futures sont nécessaires afin d'améliorer le temps de calcul et de choisir les paramètres d'intérêt.

Chapitre 6

Conclusion

Cette thèse introduit un nouvel estimateur spectrale nommé le quantile périodogramme pénalisé en utilisant des méthodes de régularisation L_1 avec la régression quantile avec une base harmonique. La motivation principale derrière cette proposition est la fuite spectrale (Leakage) observée dans l'estimation du quantile périodogramme, spécialement, dans le cas de présence de plusieurs pics de fréquences. Dans un premier temps, on a démontré que le ALQP a un comportement similaire au comportement du QP dans le sens qu'ils sont tous les deux proportionnels au spectre du quantile. La différence provient de la distribution asymptotique du ALQP, qui est χ^2 à 1 degré de liberté pour certaines fréquences. Les résultats des simulations montrent la supériorité du ALQP au sens de détection de périodicités cachées avec différents types de bruit, ainsi qu'avec plusieurs périodicités.

Deux applications avec des données réelles de l'approche proposée ont été discutées. La première application a été conduite dans un cadre d'étude des signaux EEG pour les systèmes BCI, dont la nouvelle approche a été utilisée comme une méthode d'extraction de caractéristiques. Ces caractéristiques sont introduites à des classifieurs comme le k-plus proches voisins, la forêt aléatoire, ..etc, pour déterminer si le signal appartient à une classe d'imagerie motrice de la main gauche, la main droite ou à la classe de repos. Pour l'évaluation des algorithmes, deux expérimentations ont été réalisées. La première expérimentation a pour objectif de mesurer la séparabilité des espaces de ca-

ractéristiques fournit par les différentes approches spectrales. Pour cela, un schéma de cross validation interne-externe a été appliqué (CV de 10 fois avec plusieurs exécutions). L'objectif de la seconde expérimentation est de mesurer la performance dite de session à session. La motivation derrière cela s'appuie sur les variations de la dynamique EEG entre les sessions à temps éloigné. Les résultats montrent que l'extraction de caractéristiques basée sur le quantile périodogramme pénalisé associée au classifieur de forêt aléatoire présente les meilleures performances. Le quantile périodogramme régularisé montre une amélioration significative des performances indépendamment du classifieur utilisé.

La deuxième application a été conçue pour pouvoir classifier des signaux accéléromètres des mouvements d'auto-rééducation. Un schéma de 10-fold cross validation a été utilisé pour l'évaluation des performances. Les résultats montrent que l'estimateur proposé est robuste avec n'importe quel classifieur introduit. De plus, il donne de meilleures performances.

En conclusion pour les futures travaux, on peut proposer des algorithmes qui peuvent réduire le temps de calcul. On peut aussi, étudier le cas d'optimisation de choix des paramètres τ et α . Quelqu'un peut examiner le biais et le MSE pour le nouvel estimateur proposé. Finalement, quelqu'un peut aussi généraliser le quantile périodogramme pénalisé pour le cas multivarié.

Bibliographie

- [1] Emillie AARTS et José Luis ENCARNAÇÃO. *True visions: The emergence of ambient intelligence*. 2006.
- [2] Hamid AGHAJAN et al. “Distributed vision-based accident management for assisted living”. In : *International conference on Smart homes and health telematics*. Springer. 2007, p. 196-205.
- [3] Metin AKAY. *Time Frequency and Wavelets in Biomedical Signal Processing*. IEEE press series in Biomedical Engineering, 1998.
- [4] Felicity R ALLEN et al. “Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models”. In : *Physiological measurement* 27.10 (2006), p. 935.
- [5] Kerem ALTUN, Billur BARSHAN et Orkun TUNÇEL. “Comparative study on classifying human activities with miniature inertial and magnetic sensors”. In : *Pattern Recognition* 43.10 (2010), p. 3605-3620.
- [6] Charles W ANDERSON, Erik A STOLZ et Sanyogita SHAMSUNDER. “Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks”. In : *IEEE Transactions on Biomedical Engineering* 45.3 (1998), p. 277-286.
- [7] Anton ANTONOV. *Quantile regression through linear programming*. 2013. URL : <https://mathematicaforprediction.wordpress.com/2013/12/16/quantile-regression-through-linear-programming/> (visité le 16/12/2013).

- [8] Ferhat ATTAL et al. "Physical human activity recognition using wearable sensors". In : *Sensors* 15.12 (2015), p. 31314-31338.
- [9] Ling BAO et Stephen S INTILLE. "Activity recognition from user-annotated acceleration data". In : *International Conference on Pervasive Computing*. Springer, 2004, p. 1-17.
- [10] Ada-Helen BAYER et Leon HARPER. *Fixing to stay: A national survey of housing and home modification issues*. AARP, 2000.
- [11] Benjamin BLANKERTZ, Gabriel CURIO et Klaus-Robert MÜLLER. "Classifying single trial EEG: Towards brain computer interfacing". In : *Advances in neural information processing systems*. 2002, p. 157-164.
- [12] Katarzyna BLINOWSKA et Piotr DURKA. "Electroencephalography (eeg)". In : *Wiley Encyclopedia of Biomedical Engineering* (2006).
- [13] Peter BLOOMFIELD. *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.
- [14] Alberto G BONOMI et al. "Detection of type, duration, and intensity of physical activity using an accelerometer". In : *Medicine & Science in Sports & Exercise* 41.9 (2009), p. 1770-1777.
- [15] Jaimie F BORISOFF et al. "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch". In : *IEEE Transactions on Biomedical Engineering* 51.6 (2004), p. 985-992.
- [16] Leo BREIMAN. "Random forests". In : *Machine learning* 45.1 (2001), p. 5-32.
- [17] Peter J BROCKWELL et Richard A DAVIS. *Time series: theory and methods*. Springer Science & Business Media, 2013.
- [18] Scott Shaobing CHEN et David L DONOHO. "Application of basis pursuit in spectrum estimation". In : *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. T. 3. IEEE, 1998, p. 1865-1868.

- [19] Xavier D’HAULTFÈUILLE et Pauline GIVORD. “La régression quantile en pratique”. In : *Journées de Méthodologie Statistique, Paris* (2012).
- [20] Xavier D’HAULTFÈUILLE et Pauline GIVORD. “La régression quantile en pratique”. In : *Economie et statistique* 471.1 (2014), p. 85-111.
- [21] Janis J DALY et Jonathan R WOLPAW. “Brain–computer interfaces in neurological rehabilitation”. In : *The Lancet Neurology* 7.11 (2008), p. 1032-1043.
- [22] Cristina DAVINO, Marilena FURNO et Domenico VISTOCCO. *Quantile regression: theory and applications*. John Wiley & Sons, 2013.
- [23] Holger DETTE et al. “Of copulas, quantiles, ranks and spectra: An L_1 -approach to spectral analysis”. In : *Bernoulli* 21.2 (2015), p. 781-831.
- [24] Jochen FAHRENBURG et al. “Assessment of posture and motion by multichannel piezoresistive accelerometer recordings”. In : *Psychophysiology* 34.5 (1997), p. 607-612.
- [25] Jianqing FAN et Runze LI. “Variable selection via nonconcave penalized likelihood and its oracle properties”. In : *Journal of the American statistical Association* 96.456 (2001), p. 1348-1360.
- [26] Mehrdad FATOURECHI et al. “A wavelet-based approach for the extraction of event related potentials from EEG”. In : *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP’04). IEEE International Conference on*. T. 2. IEEE. 2004, p. ii-737.
- [27] Sven FLECK et Wolfgang STRASSER. “Smart camera based monitoring system and its application to assisted living”. In : *Proceedings of the IEEE* 96.10 (2008), p. 1698-1714.
- [28] Martin FOX, Herman RUBIN et al. “Admissibility of quantile estimates of a single location parameter”. In : *The Annals of Mathematical Statistics* 35.3 (1964), p. 1019-1030.
- [29] Jerome H FRIEDMAN. “On bias, variance, 0/1—loss, and the curse-of-dimensionality”. In : *Data mining and knowledge discovery* 1.1 (1997), p. 55-77.

- [30] Wayne A FULLER. *Introduction to statistical time series*. T. 428. John Wiley & Sons, 2009.
- [31] Francis GALTON. “Kinship and correlation”. In : *The North American Review* 150.401 (1890), p. 419-431.
- [32] Carl Friedrich GAUSS. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser, 1809.
- [33] Hristijan GJORESKI, Mitja LUŠTREK et Matjaž GAMS. “Accelerometer placement for posture recognition and fall detection”. In : *2011 Seventh International Conference on Intelligent Environments*. IEEE. 2011, p. 47-54.
- [34] Damodar N GUJARATI. *Basic econometrics*. Tata McGraw-Hill Education, 2009.
- [35] Isabelle GUYON et André ELISSEEFF. “An introduction to variable and feature selection”. In : *Journal of machine learning research* 3.Mar (2003), p. 1157-1182.
- [36] Jacques HADAMARD. *Lectures on Cauchy’s problem in linear partial differential equations*. Courier Corporation, 2003.
- [37] Andreas HAGEMANN. “Robust spectral analysis”. In : (2011).
- [38] Andrew C HARVEY. “On comparing regression models in levels and first differences”. In : *International Economic Review* (1980), p. 707-720.
- [39] Pawel HERMAN et al. “Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification”. In : *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16.4 (2008), p. 317-326.
- [40] Gareth JAMES et al. *An introduction to statistical learning*. T. 112. Springer, 2013.
- [41] Dean M KARANTONIS et al. “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring”. In : *IEEE transactions on information technology in biomedicine* 10.1 (2006), p. 156-167.

- [42] Taichi KATO et Makoto UEMURA. “Period analysis using the least absolute shrinkage and selection operator (Lasso)”. In : *Publications of the Astronomical Society of Japan* 64.6 (2012), p. 122.
- [43] Keith KNIGHT et al. “Limiting distributions for L_1 regression estimators under general conditions”. In : *The Annals of Statistics* 26.2 (1998), p. 755-770.
- [44] Roger KOENKER. *Quantile regression*. 38. Cambridge university press, 2005.
- [45] Roger KOENKER et Gilbert BASSETT JR. “Regression quantiles”. In : *Econometrica: journal of the Econometric Society* (1978), p. 33-50.
- [46] Camille LANDAIS. “Les hauts revenus en France (1998-2006): une explosion des inégalités”. In : (2007).
- [47] Junbum LEE et Suhasini Subba RAO. “The quantile spectral density and comparison based tests for nonlinear time series”. In : *arXiv preprint arXiv:1112.2759* (2011).
- [48] Adrien Marie LEGENDRE. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot, 1805.
- [49] Steven LEMM, Klaus-Robert MÜLLER et Gabriel CURIO. “A generalized framework for quantifying the dynamics of EEG event-related desynchronization”. In : *PLoS computational biology* 5.8 (2009), e1000453.
- [50] Ta-Hsin LI. “Laplace periodogram for time series analysis”. In : *IBM research report* (2008), RC24473.
- [51] Ta-Hsin LI. “Laplace periodogram for time series analysis”. In : *Journal of the American Statistical Association* 103.482 (2008), p. 757-768.
- [52] Ta-Hsin LI. “Quantile periodograms”. In : *Journal of the American Statistical Association* 107.498 (2012), p. 765-776.
- [53] Youjuan LI et Ji ZHU. “L 1-norm quantile regression”. In : *Journal of Computational and Graphical Statistics* 17.1 (2008), p. 163-185.
- [54] Yaeji LIM et Hee-Seok OH. “Composite Quantile Periodogram for Spectral Analysis”. In : *Journal of Time Series Analysis* 37.2 (2016), p. 195-221.

- [55] Oliver LINTON et Yoon-Jae WHANG. “The quantilogram: With an application to evaluating directional predictability”. In : *Journal of Econometrics* 141.1 (2007), p. 250-282.
- [56] Huan LIU et Lei YU. “Toward integrating feature selection algorithms for classification and clustering”. In : *IEEE Transactions on knowledge and data engineering* 17.4 (2005), p. 491-502.
- [57] Fabien LOTTE et al. “A review of classification algorithms for EEG-based brain-computer interfaces”. In : *Journal of neural engineering* 4.2 (2007), R1.
- [58] Juliano MACHADO, Alexandre BALBINOT et Adalberto SCHUCK. “A study of the Naive Bayes classifier for analyzing imaginary movement EEG signals using the Periodogram as spectral estimator”. In : *Biosignals and Biorobotics Conference (BRC), 2013 ISSNIP*. IEEE. 2013, p. 1-4.
- [59] Y. MATANGA. “Analysis of control attainment in endogenous electroencephalogram based brain computer interface”. Mém. de mast. Tshwane University of Technology, Pretoria, South Africa, 2017.
- [60] Uwe MAURER et al. “Location and activity recognition using eWatch: A wearable sensor platform”. In : *Ambient Intelligence in Everyday Life*. Springer, 2006, p. 86-102.
- [61] Eric T MCADAMS et al. “Biomedical sensors for ambient assisted living”. In : *Advances in Biomedical Sensing, Measurements, Instrumentation and Systems*. Springer, 2010, p. 240-262.
- [62] Dennis J MCFARLAND et Jonathan R WOLPAW. “Brain-computer interface operation of robotic and prosthetic devices”. In : *Computer* 41.10 (2008).
- [63] Tarek MEDKOUR, Andrew T WALDEN et Adrian BURGESS. “Graphical modeling for brain connectivity via partial coherence”. In : *Journal of neuroscience methods* 180.2 (2009), p. 374-383.

- [64] Aymen MEZIANI et al. “A Lasso quantile periodogram based feature extraction for EEG-based motor imagery”. In : *Journal of Neuroscience Methods* 328 (2019), p. 108434.
- [65] Klaus-Robert MÜLLER et al. “Machine learning techniques for brain-computer interfaces”. In : (2004).
- [66] Bijan NAJAFI et al. “Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly”. In : *IEEE Transactions on biomedical Engineering* 50.6 (2003), p. 711-723.
- [67] Ben NHAM, Kanya SIANGLIULUE et Serena YEUNG. “Predicting mode of transport from iphone accelerometer data”. In : *Machine Learning Final Projects, Stanford University* (2008).
- [68] Luis Fernando NICOLAS-ALONSO et Jaime GOMEZ-GIL. “Brain computer interfaces, a review”. In : *Sensors* 12.2 (2012), p. 1211-1279.
- [69] Ernst NIEDERMEYER et FH Lopes da SILVA. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [70] Juha PARKKA et al. “Activity classification using realistic data from wearable sensors”. In : *IEEE Transactions on information technology in biomedicine* 10.1 (2006), p. 119-128.
- [71] Emanuel PARZEN. *Maximum entropy interpretation of autoregressive spectral densities*. Rapp. tech. TEXAS A et M UNIV COLLEGE STATION INST OF STATISTICS, 1982.
- [72] Donald B PERCIVAL et Andrew T WALDEN. *Spectral analysis for physical applications*. cambridge university press, 1993.
- [73] Gert PFURTSCHELLER et Christa NEUPER. “Motor imagery activates primary sensorimotor area in humans”. In : *Neuroscience letters* 239.2-3 (1997), p. 65-68.

- [74] G PFURTSCHELLER et al. "EEG-based discrimination between imagination of right and left hand movement". In : *Electroencephalography and clinical Neurophysiology* 103.6 (1997), p. 642-651.
- [75] D POPIVANOV et A MINEVA. "Testing procedures for non-stationarity and non-linearity in physiological signals". In : *Mathematical biosciences* 157.1-2 (1999), p. 303-320.
- [76] S PREECE et al. "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data". In : *IEEE Transactions on Biomedical Engineering* 56 (2009), p. 871-879.
- [77] Stephen J PREECE et al. "Activity identification using body-mounted sensors—a review of classification techniques". In : *Physiological measurement* 30.4 (2009), R1.
- [78] Maurice Bertram PRIESTLEY. "Spectral analysis and time series". In : (1981).
- [79] T PRILL et J FAHRENBERG. "Simultaneous assessment of posture and limb movements (eg, periodic leg movements) with calibrated multiple accelerometry". In : *Physiological Measurement* 27.10 (2006), N47.
- [80] Islam QUDAH, Peter LEIJDEKKERS et Valerie GAY. "Using mobile phones to improve medication compliance and awareness for cardiac patients". In : *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*. ACM. 2010, p. 36.
- [81] Lawrence R RABINER. "A tutorial on hidden Markov models and selected applications in speech recognition". In : *Proceedings of the IEEE* 77.2 (1989), p. 257-286.
- [82] Herbert RAMOSER, Johannes MULLER-GERKING et Gert PFURTSCHELLER. "Optimal spatial filtering of single trial EEG during imagined hand movement". In : *IEEE transactions on rehabilitation engineering* 8.4 (2000), p. 441-446.
- [83] Joel JPC RODRIGUES et al. "Enabling Technologies for the Internet of Health Things". In : *Ieee Access* 6 (2018), p. 13129-13141.

- [84] Daniel ROGGEN et al. "Collecting complex activity datasets in highly rich networked sensor environments". In : *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*. IEEE. 2010, p. 233-240.
- [85] Fariba SADRI. "Ambient intelligence: A survey". In : *ACM Computing Surveys (CSUR)* 43.4 (2011), p. 36.
- [86] Alois SCHLÖGL, Klaus LUGGER et Gert PFURTSCHELLER. "Using adaptive autoregressive parameters for a brain-computer-interface experiment". In : *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*. T. 4. Ieee. 1997, p. 1533-1535.
- [87] Pradeep SHENOY et al. "Towards adaptive classification for BCI". In : *Journal of neural engineering* 3.1 (2006), R13.
- [88] Julie SOLARD. "Les très hauts revenus: des différences de plus en plus marquées entre 2004 et 2007". In : *consommation* 99.84 (), p. 469.
- [89] David S STOFFER et Robert H SHUMWAY. "Time Series Analysis and Its Applications". In : *With R Examples* (2000).
- [90] Petre STOICA, Randolph L MOSES et al. *Spectral analysis of signals*. T. 1. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
- [91] Piotr SZACHEWICZ. "Classification of Motor Imagery for Brain-Computer Interfaces". Thèse de doct. Ph. D. thesis, Poznan University of Technology, 2013.
- [92] Robert TIBSHIRANI. "Regression shrinkage and selection via the lasso". In : *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), p. 267-288.
- [93] Nations UNITED. "World Population Prospects: The 2010 Revision, CD-ROM Edition." In : (2011).
- [94] WN VENABLES et BD RIPLEY. "Modern applied statistics with S. 4th Springer". In : *New York* (2002).

- [95] Hansheng WANG et Chenlei LENG. “Unified LASSO estimation by least squares approximation”. In : *Journal of the American Statistical Association* 102.479 (2007), p. 1039-1048.
- [96] Ning WANG et al. “Accelerometry based classification of walking patterns using time-frequency analysis”. In : *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE. 2007, p. 4899-4902.
- [97] Tao WANG, Jie DENG et Bin HE. “Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns”. In : *Clinical Neurophysiology* 115.12 (2004), p. 2744-2753.
- [98] Andrew R WEBB. *Statistical pattern recognition*. John Wiley & Sons, 2003.
- [99] Peter WELCH. “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms”. In : *IEEE Transactions on audio and electroacoustics* 15.2 (1967), p. 70-73.
- [100] Peter WHITTLE. “The simultaneous estimation of a time series harmonic components and covariance structure”. In : *Trabajos de estadística* 3.1-2 (1952), p. 43-57.
- [101] G WISE, Apostolos TRAGANITIS et J THOMAS. “The effect of a memoryless nonlinearity on the spectrum of a random process”. In : *IEEE Transactions on Information Theory* 23.1 (1977), p. 84-89.
- [102] Jonathan R WOLPAW et al. “Brain-computer interface technology: a review of the first international meeting”. In : *IEEE transactions on rehabilitation engineering* 8.2 (2000), p. 164-173.
- [103] Yichao WU et Yufeng LIU. “Variable selection in quantile regression”. In : *Statistica Sinica* (2009), p. 801-817.
- [104] Nobuyuki YAMAWAKI et al. “An enhanced time-frequency-spatial approach for motor imagery classification”. In : *IEEE transactions on neural systems and rehabilitation engineering* 14.2 (2006), p. 250-254.

- [105] G Udny YULE. “On the theory of correlation”. In : *Journal of the Royal Statistical Society* 60.4 (1897), p. 812-854.
- [106] Yu ZHANG et al. “Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface”. In : *Journal of neuroscience methods* 255 (2015), p. 85-91.
- [107] Shang-Ming ZHOU, John Q GAN et Francisco SEPULVEDA. “Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface”. In : *Information Sciences* 178.6 (2008), p. 1629-1640.
- [108] Chun ZHU et Weihua SHENG. “Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living”. In : *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 41.3 (2011), p. 569-573.
- [109] Hui ZOU. “The adaptive lasso and its oracle properties”. In : *Journal of the American statistical association* 101.476 (2006), p. 1418-1429.